
**Raciocínio Baseado em Casos para apoiar a
formação de grupos baseada nos traços de
personalidade do modelo *Big Five***

Taís Borges Ferreira



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Taís Borges Ferreira

**Raciocínio Baseado em Casos para apoiar a
formação de grupos baseada nos traços de
personalidade do modelo *Big Five***

Dissertação de mestrado apresentada ao
Programa de Pós-graduação da Faculdade
de Computação da Universidade Federal de
Uberlândia como parte dos requisitos para a
obtenção do título de Mestre em Ciência da
Computação.

Área de concentração: Inteligência Artificial

Orientador: Márcia Aparecida Fernandes

Uberlândia
2018

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

F383r
2018 Ferreira, Taís Borges, 1988-
Raciocínio baseado em casos para apoiar a formação de grupos baseada nos traços de personalidade do modelo Big Five [recurso eletrônico] / Taís Borges Ferreira. - 2018.

Orientadora: Márcia Aparecida Fernandes.
Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: <http://dx.doi.org/10.14393/ufu.di.2018.1214>
Inclui bibliografia.
Inclui ilustrações.

1. Computação. 2. Inteligência artificial. 3. Processamento de linguagem natural (Computação). 4. Ensino a distância. 5. Ensino auxiliado por computador. I. Fernandes, Márcia Aparecida (Orient.) II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDU: 681.3

Maria Salete de Freitas Pinheiro - CRB6/1262

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada “**Raciocínio Baseado em Casos para apoiar a formação de grupos baseada nos traços de personalidade do modelo *Big Five***” por **Taís Borges Ferreira** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, ___ de _____ de ____

Orientadora: _____

Profa. Dra. Márcia Aparecida Fernandes
Universidade Federal de Uberlândia

Banca Examinadora:

Prof. Dr. Renan Gonçalves Cattelan
Universidade Federal de Uberlândia

Prof. Dr. Eliseo Berni Reategui
Universidade Federal do Rio Grande do Sul

Agradecimentos

À minha família, especialmente à minha mãe, Conceição, pelo amor e compreensão.

À minha orientadora, Márcia Aparecida Fernandes, por seus ensinamentos, conselhos e tempo dedicado à orientação do meu trabalho. Também pela paciência e compreensão.

Ao colega Fabiano Silverio, grande apoio no início da pesquisa passando seu conhecimento sobre aprendizagem colaborativa e formação de grupos, o que me ajudou imensamente a definir o tema da pesquisa.

Aos colegas que cederam algumas de suas turmas e incentivarem seus alunos a responder ao inventário de personalidade usado neste trabalho: Cleon Xavier Pereira Júnior, Sérgio Piter Nogueira e Luciane de Fátima Silva.

Ao amigo Régis Michel dos Santos Sousa, cuja ajuda, fornecendo espaço para discutir as tarefas colaborativas e as métricas de avaliação, envolvendo suas turmas da FAGEN na coleta de dados e nos experimentos de formação de grupos, foi essencial para coletar dados relevantes para desenvolver meu trabalho.

Ao aluno de iniciação científica Angelo Neto, pelo tempo que trabalhamos juntos.

Aos amigos e companheiros de conversa, almoço e da hora do lanchinho da tarde: Sara Melo, Jean Roberto, Claudiney, Amanda Gaspar, Alex Dias, Giovani, Marcelo Silva, Acrísio, Caio Augusto, Vinícius Maciel, Lucas Miranda, Rodrigo Elias, Hiran Nonato, Juan Nogales, Rafael Dias e Fabíola Souza. Foram importante fonte de apoio, discussão de ideias e fizeram os dias no laboratório mais confortáveis.

Aos secretários da FACOM, Erisvaldo Araújo Fialho e Sônia Maria Miguel, pela disponibilidade e ajuda.

Aos professores da FACOM pelos ensinamentos.

A todos que nestes dois anos e meio de mestrado foram fonte de apoio para o desenvolvimento do meu trabalho.

Por fim, meu agradecimento à CAPES pelo apoio Financeiro.

*“Everything in life is important,
important things are simple,
simple things are never easy.”
(Murphy’s Laws)*

Resumo

A aprendizagem colaborativa envolve a construção de conhecimento por meio da interação entre os membros de um grupo para atingir uma meta compartilhada. Para que essa interação ocorra, é necessário assegurar que os grupos formados para trabalhar em conjunto não sejam um impedimento para a colaboração. O agrupamento de forma descuidada pode ser o gatilho para situações indesejáveis, como o isolamento de algum aluno dentro de um grupo. Por esse motivo, é essencial considerar uma estratégia de formação de grupos para promover interações produtivas e garantir que tal estratégia possa ser usada para dar apoio à colaboração. Dentre os critérios que podem ser considerados na formação de grupos, os traços de personalidade dos alunos indicam tendências de comportamento que podem influenciar seu desempenho. Por exemplo, em um grupo com muitos alunos com alto grau de Extroversão, eles poderiam se distrair com a interação e negligenciar a meta comum. Assim, conhecendo tais características é possível definir uma estratégia de formação de grupos de estudantes para aumentar a chance de interações produtivas para a colaboração. Com o objetivo de fornecer um meio de detectar traços de personalidade de forma transparente e tornar essa informação disponível para apoiar estratégias de formação de grupos para colaboração, este trabalho trata da avaliação de textos escritos por alunos, que podem, por exemplo, ser capturados em chats e fóruns em ambientes virtuais de aprendizagem. Apesar de não ter sido possível encontrar algoritmos que detectam os perfis com alta acurácia, os resultados dos experimentos apontam para a possibilidade de melhorar a qualidade da predição. Em paralelo à detecção dos traços de personalidade nos textos, testes com grupos de alunos foram conduzidos com o objetivo de formar uma base para Raciocínio Baseado em Casos. Conforme observado nas turmas em que foram conduzidos experimentos de formação de grupos a partir de consultas na base, os grupos formados apresentaram boa interação. Mais ainda, de acordo com o relato do professor, os grupos formados no experimento foram percebidos como melhores do que aqueles formados pelos próprios alunos. Quanto à base de casos, os experimentos conduzidos apontam para a melhora na qualidade e abrangência dos casos conforme a base é atualizada. Como essa atualização é parte do processo de funcionamento do RBC,

haverá melhora contínua da qualidade das sugestões de agrupamento e das informações que possam ser extraídas dela. Como a base de casos pode crescer muito, tornando as consultas mais pesadas e demoradas, também foram conduzidos experimentos no sentido de extrair regras de formação de grupos da base. Os resultados sugerem que a extração das regras é uma alternativa viável para formação de grupos, em caso de aumento do volume de casos.

Palavras-chave: Aprendizagem Colaborativa, Raciocínio Baseado em Casos, *Big Five*, Processamento de Linguagem Natural.

Abstract

Collaborative learning comprehends the developing of knowledge through interactions among group members to achieve a shared goal. For this interaction to happen, one needs to ensure that the students grouped together will not intervene negatively in the collaboration process. Grouping carelessly can trigger undesirable scenarios, such as a student isolated in a group. For this reason, it is essential to consider a group formation strategy to promote productive interactions and to ensure that such a strategy can be used to support collaboration. Among the criteria that can be considered in group formation, the personality traits of the students indicate behavioral tendencies that may influence their performance. For example, in a group with many students with a high degree of Extroversion, they could get distracted by interaction and neglect the common goal. Thus, by knowing these characteristics one can think of developing a group formation strategy to group students and increase the occurrence of productive interactions for collaboration. In order to provide a way to detect personality traits in a transparent way and to make this information available to support group formation strategies for collaboration, this work evaluates texts written by students which can be captured in chats and forums in virtual learning environments, for example. Although the tested algorithms do not reached high accuracy in the experiments, the results pointed a chance to improve the prediction accuracy. Along with the detection of personality traits in the text, experiments with groups of students were conducted with the objective of building the case base for Case Based Reasoning. As observed in the classes in which group formation experiments were conducted, by querying the case base, the new groups formed had good interaction. Moreover, according to the teacher's report, the groups formed were perceived as better than those formed by the students themselves. Regarding the case base, the experiments conducted indicate an improvement in the quality and comprehensiveness of the cases as the base was updated. Since this update is part of the Case Based Reasoning operation process, the base will be continuously improved, as well as the quality of grouping suggestions and the information that can be extracted from it. Once the case base can grow bigger, turning the queries time-consuming, experiments aiming to extract grouping rules from

the case base were also conducted. The results obtained suggest that using grouping rules is a feasible alternative to form groups, if the number of cases increases too much.

Keywords: Collaborative Learning, Case-Based Reasoning, Big Five, Natural Language Processing.

Lista de ilustrações

Figura 1 – Funcionamento do Raciocínio Baseado em Casos.	34
Figura 2 – Funcionamento e Fluxo de dados	44
Figura 3 – Estrutura do caso.	48
Figura 4 – Cálculo da similaridade entre os grupos de caso e uma consulta. (a) Calcula a similaridade entre todos os pares . (b) Seleciona a maior similaridade e exclui as linhas com indivíduos envolvidos no cálculo da similaridade selecionada. (c) Seleciona o próximo maior disponível. . .	50
Figura 5 – Similaridade entre os casos da Turma 1.	75
Figura 6 – Similaridade entre casos das Turmas 1 e 2.	76
Figura 7 – Similaridade entre casos das Turmas 1, 2, 3 e 4.	76
Figura 8 – Similaridade após inserção dos grupos sugeridos pelo RBC	77
Figura 9 – Perfis possíveis presentes em algum caso da base.	79
Figura 10 – Porcentagem de perfis com pelo menos 70% de similaridade com algum caso da base.	79
Figura 11 – Integração da aplicação com o Moodle.	102
Figura 12 – Formação de grupos dentro da aplicação.	103
Figura 13 – Reordenação manual dos grupos formados.	104
Figura 14 – Turmas as quais o usuário está relacionado e tem permissão de visualizar.	105
Figura 15 – Interface do curso com a lista de alunos matriculados e as opções de ação disponíveis para professor.	105
Figura 16 – Informações dos grupos formados no curso (visão do professor).	106
Figura 17 – Seleção do tamanho dos grupos para gerar novo agrupamento na turma.	106
Figura 18 – Regras associadas ao grupo 52 do novo agrupamento. O grupo está marcado com um sinal de alerta indicando a associação das caracterís- ticas do grupo com uma regra que indica possibilidade de desempenho ruim.	107
Figura 19 – Reordenação manual do grupos movendo um aluno.	107

Lista de tabelas

Tabela 1 – Categorias de palavras do dicionário utilizadas.	46
Tabela 2 – Inventário de 44 itens em Português do Brasil.	54
Tabela 3 – Número de alunos em cada uma das faixas de escore.	57
Tabela 4 – Resultados obtidos com os algoritmos que devolvem saídas contínuas. .	58
Tabela 5 – Resultados obtidos com os algoritmos de classificação para 3 classes de escores.	59
Tabela 6 – Dados dos grupos da Turma 1.	64
Tabela 7 – Dados dos grupos da Turma 2.	65
Tabela 8 – Dados dos grupos da Turma 3.	66
Tabela 9 – Dados dos grupos da Turma 4.	67
Tabela 10 – Sugestões de agrupamentos resultantes de consultas na BC.	70
Tabela 11 – Grupos da primeira atividade.	72
Tabela 12 – Grupos da segunda atividade.	73
Tabela 13 – Grupos da terceira atividade.	74
Tabela 14 – Similaridade entre caso e consulta com a troca das funções.	81
Tabela 15 – Valores do resumo dos traços de personalidade.	82
Tabela 16 – Regras extraídas da base atualizada com os 38 casos.	83

Lista de siglas

AICLS *Adaptive and Intelligent Systems for Collaborative Learning Support*

AVA Ambiente Virtual de Aprendizagem

BC Base de Casos

EAD Ensino à Distância

IA Inteligência Artificial

LIWC *Linguistic Inquiry and Word Count*

LWL *Locally Weighted Learning*

MLP *Multilayer Perceptron*

MSO Medida de Similaridade entre Objetos

PLN Processamento de Linguagem Natural

RBC Raciocínio Baseado em Casos

RBF *Radial Basis Function*

SG Similaridade entre os Grupos

SM Similaridade entre as Métricas de desempenho dos grupos

SMO *Sequential Minimal Optimization*

Sumário

1	INTRODUÇÃO	21
1.1	Motivação	22
1.2	Objetivos	24
1.3	Hipóteses	24
1.4	Contribuições Esperadas	24
1.5	Organização da Dissertação	25
2	REVISÃO DA LITERATURA	27
2.1	Fundamentação Teórica	27
2.1.1	Aprendizagem Colaborativa e Formação de grupos	27
2.1.2	<i>Big Five</i>	29
2.1.3	Extração de Informações do Texto	31
2.1.4	Raciocínio Baseado em Casos	32
2.2	Trabalhos Relacionados	37
2.2.1	Formação de Grupos para Colaboração	37
2.2.2	Deteção de Características do Aluno no Texto	41
3	FORMAÇÃO DE GRUPOS COM O MODELO <i>BIG FIVE</i>	43
3.1	Funcionamento e Fluxo de dados	43
3.2	Módulo de Classificação de Traços	44
3.3	Módulo de Agrupamento	47
3.3.1	Estrutura de um Caso	47
3.3.2	Cálculo da Similaridade entre Caso e Consulta	49
3.3.3	Etapas do RBC	51
4	EXPERIMENTO 1: DETECÇÃO DE TRAÇOS	53
4.1	Questionário e Coleta de Textos	53
4.2	Predição dos Traços de Personalidade	57

4.3	Discussão dos Resultados	59
5	EXPERIMENTO 2: FORMAÇÃO DE GRUPOS	63
5.1	Construção da Base de Casos	63
5.2	Formação de Grupos com RBC	68
5.2.1	Sugestão de Agrupamento	69
5.2.2	Grupos Formados para Atividade Colaborativa	71
5.3	Abrangência da Base de Casos	74
5.3.1	Casos x Casos	75
5.3.2	Consultas com Todos os Perfis de Indivíduo Possíveis	78
5.3.3	Efeitos da Mudança de Cálculo de Similaridade no Resultado da Consulta	80
5.4	Regras de Formação de Grupos	81
5.4.1	Mineração de Regras Base de Casos	82
5.4.2	Aplicação na Formação de Grupos	84
5.5	Discussão dos Resultados	84
6	CONCLUSÃO	87
6.1	Principais Contribuições	88
6.2	Trabalhos Futuros	89
6.3	Contribuições em Produção Bibliográfica	90
	REFERÊNCIAS	93

APÊNDICES 99

APÊNDICE A	–	APLICAÇÃO WEB PARA FORMAÇÃO DE GRUPOS	101
------------	---	---	-----

CAPÍTULO 1

Introdução

As teorias relacionadas à aprendizagem têm adotado a visão de que a aprendizagem não ocorre somente por meio instrucional e que o aluno é o principal agente responsável pelo próprio aprendizado, consumindo, criando e compartilhando conhecimento por meio de várias fontes distintas. Por exemplo, Vygotsky (VYGOTSKY, 1978) propôs a zona de desenvolvimento proximal como característica fundamental para a aprendizagem. Segundo essa proposta, o aprendizado se relaciona a vários processos internos de desenvolvimento que somente são despertados por meio da interação entre pessoas. Assim, o aprendizado pode ser considerado um processo social que ocorre de forma mais efetiva por meio de interações interpessoais em um contexto colaborativo.

No Ensino à Distância (EAD), além do contato entre alunos e desses com o professor ser menos frequente do que no ensino presencial, esse contato muitas vezes ocorre por meio das ferramentas de comunicação de algum Ambiente Virtual de Aprendizagem (AVA). A ausência de interação física em AVAs é considerada um grande desafio, de acordo com (KOP, 2011) e (EVERSON, 2011). Em EAD, o aluno encontra-se sozinho no processo de aprendizagem e o professor muitas vezes não tem ferramentas para conhecer e guiar o aprendizado dos alunos nessas plataformas. Nesse contexto, a aprendizagem colaborativa torna-se uma abordagem importante no estímulo da interação e melhoria da motivação do aluno acerca do próprio processo de aprendizagem.

A aprendizagem colaborativa envolve a construção de conhecimento por meio de interação entre os membros de um grupo para atingir uma meta compartilhada (HRON; FRIEDRICH, 2003). Além disso, um aluno trabalhando em colaboração com outros para realizar uma atividade colaborativa é exposto a pontos de vista diferentes, questionando o entendimento inicial e motivando o aprendizado. Assim, a aprendizagem por meio de atividade colaborativa também vai de encontro aos três atributos ligados à aprendizagem efetiva, apontados por (ALAVI, 1994), que são: aprendizagem ativa e construção do conhecimento, trabalho de equipe e aprendizagem usando resolução de problemas.

Uma forma de apoiar a colaboração é preocupar-se com a seleção de indivíduos para compor os grupos. A formação de um grupo é considerada crucial para desencadear

interações produtivas entre os pares (MAGNISALIS; DEMETRIADIS; KARAKOSTAS, 2011). Uma determinada combinação de alunos pode tanto contribuir para aumentar a probabilidade de colaboração como ser motivo para que esta não ocorra. Em ambientes presenciais, a formação de grupos normalmente ocorre sem interferência do professor, já que os alunos se agrupam de acordo com suas afinidades e conhecimento prévio sobre as características dos colegas.

Em AVAs, esse tipo de estratégia de formação de grupos é mais difícil. Daí a importância de fornecer, em tais ambientes, ferramentas que auxiliem o professor a conhecer melhor os alunos e também a formação de grupos para colaboração efetiva. Nesse sentido, este trabalho apoia-se no modelo *Big Five* (JOHN; SRIVASTAVA, 1999) como forma de distinguir características de personalidade que levam as pessoas a trabalharem bem em grupos e, assim, usar tais informações na formação de grupos para colaboração.

Como há possibilidade de identificação desses traços em textos escritos, o trabalho desenvolvido também busca empregar os textos escritos por alunos na identificação automática dessas características em AVAs. Dessa forma, as características do aluno podem ser usadas na formação de grupos, sem a necessidade de envolvimento direto na coleta de informações sobre os traços de personalidade.

1.1 Motivação

Cada aluno tem experiências únicas, visões de mundo e diferentes informações adquiridas no contexto em que vive. A aprendizagem colaborativa eficaz deve valorizar tal diversidade, uma vez que ela pode enriquecer as experiências de aprendizagem dos alunos (BRINDLEY; WALTI; BLASCHKE, 2009). Além disso, há uma correlação significativa entre as interações sociais e a aprendizagem (XIE; MILLER; ALLISON, 2013).

Quando boas interações são criadas e mantidas, há maior probabilidade de ocorrer também o compartilhamento de conhecimento (BOZIOSELOS, 2017). Assim, além de serem capazes de aprender conteúdo proposto por meio da atividade colaborativa, os alunos também aprendem através da sua rede de interações (KENT; LASLO; RAFAELI, 2016). Dentre os vários fatores que podem influenciar o processo de colaboração, a composição do grupo é considerada crucial para desencadear interações produtivas entre os membros do grupo (MAGNISALIS; DEMETRIADIS; KARAKOSTAS, 2011).

Como as características dos membros do grupo influenciam na maneira que as pessoas trabalham juntas para alcançar um objetivo, a formação do grupo é um aspecto essencial para permitir a aquisição de conhecimento através de atividades colaborativas (MANSKE et al., 2015). Uma das formas de conhecer as características dos alunos para aplicar na formação de grupos, é avaliar seus traços de personalidade. Dentre as diferentes propostas na literatura para medir aspectos de personalidade, a construção psicométrica mais usada é o *Five Factor Model* ou *Big Five*.

O modelo *Big Five* (GOLDBERG, 1981; DIGMAN, 1990; JOHN; SRIVASTAVA, 1999) tem sido largamente utilizado na formação de grupos, pois permite distinguir elementos da personalidade que são compatíveis com relações sociais e, portanto, com o agrupamento de pessoas. Por exemplo, alguém com alto grau de Neuroticismo tem maior chance de evitar a aproximação de outras pessoas (ROBERTS et al., 2008). No entanto, uma pessoa com tal característica pode se sentir confortável para interagir se estiver em contato com pessoas que são mais sociáveis e amigáveis (BOZIO NELOS, 2017).

Pouco se sabe sobre a influência de características não cognitivas dos alunos, como os traços de personalidade no desempenho acadêmico (ALTANOPOULOU; TSELIOS, 2015). Por outro lado, cada traço de personalidade possui características próprias que dão origem a determinados comportamentos (REIS et al., 2015b) e, lançando mão de alguma estratégia para conhecer a forma como afetam o aprendizado, podem ajudar na identificação de aspectos importantes para promover colaboração.

Ainda que não se conheça todos os efeitos dos traços de personalidade no desempenho do grupo, o uso de uma abordagem como o Raciocínio Baseado em Casos (RBC), que busca uma solução tentando adaptar soluções anteriores, pode contribuir para encontrar boas combinações de alunos para colaboração com base nos traços de personalidade. Os grupos que fizeram atividade colaborativa, sua combinação de traços dos indivíduos e o desempenho, armazenados na base do RBC, vão servir de insumo para propor um novo agrupamento com base naqueles grupos que obtiveram bons resultados.

Mesmo a personalidade sendo reconhecida como fator importante na recomendação de ações personalizadas, o esforço de capturá-la de forma explícita é indesejado e pode dificultar seu uso para recomendações no mundo real (WU; CHEN, 2015). Tentando enfrentar esse desafio, vários trabalhos foram desenvolvidos para extrair conhecimento sobre o aluno, como por exemplo, (TAUSCZIK; PENNEBAKER, 2010) que apontam para a relação entre as palavras que uma pessoa escolhe e sua personalidade.

A partir de seus resultados, (MACHADO et al., 2015) e (PAIM; CAMATI; ENEMBRECK, 2016) indicam a possibilidade de identificar os traços de personalidade do *Big Five* por meio da avaliação de textos escritos em Português do Brasil e sugerem que há espaço para a melhoria da qualidade dos métodos de detecção da personalidade por meio de textos escritos. Ou seja, é possível desenvolver uma forma de detecção dos traços de personalidade sem o uso de questionários. Assim, a relevância dos aspectos de personalidade no desempenho de grupos de alunos e a possibilidade de identificá-los em textos escritos é o principal motivador da inclusão da avaliação dos textos dos alunos dentre os objetivos propostos na próxima seção.

1.2 Objetivos

O objetivo do trabalho foi explorar as características de cada um dos cinco traços de personalidade do *Big Five* (Abertura, Extroversão, Neuroticismo, Afabilidade e Conscienciosidade) para apoiar a sugestão e avaliação de formações de grupo favoráveis à aprendizagem colaborativa. Para atingir o objetivo proposto, os seguintes objetivos específicos foram definidos:

- ❑ Desenvolver um modelo de aprendizagem para os traços de personalidade por meio da análise de textos escritos.
- ❑ Encontrar combinações de alunos que possam ser usadas para sugerir bons grupos para colaboração.
- ❑ Encontrar regras de agrupamento baseadas nas combinações de perfis de personalidade dos grupos encontrados.
- ❑ Aprender novas combinações de alunos e, conseqüentemente, novas regras de agrupamento, evoluindo a base de regras a partir de informações capturadas dos grupos trabalhando em atividades colaborativas.

1.3 Hipóteses

O trabalho proposto visou validar duas hipóteses principais. Para que seja possível detectar os traços de personalidade nos textos escritos pelos alunos, deve existir uma relação entre tais textos e as características do aluno. Assim, a primeira refere-se à relação entre os traços de personalidade e textos escritos.

A segunda hipótese era de que o uso do RBC na formação de grupos, por meio das suas quatro etapas que levam à atualização e melhoria frequente da base de casos, poderia dar suporte à sugestão de boas combinações alunos para trabalhar em atividades colaborativas, considerando seus traços de personalidade.

1.4 Contribuições Esperadas

Uma vez validadas as duas hipóteses, acreditou-se ser possível construir um modelo de detecção transparente e automática de traços de personalidade sem a necessidade, por exemplo, de responder questionários, e usar os traços de personalidade para sugerir bons grupos para trabalhar em tarefas colaborativas. Além disso, com a atualização e melhoria frequentes dos casos, à medida que novos casos surgissem, acreditou-se que o RBC seria capaz de retornar sugestões cada vez melhores. Dessa forma, atingir os objetivos propostos implicaria em:

- ❑ Ter um modelo que, se integrado ao AVA, poderia oferecer suporte ao professor/tutor em AVAs para conhecer melhor os alunos quando atuando de forma colaborativa.
- ❑ Viabilizar ações pedagógicas tais como a escolha de bons parceiros para colaboração em AVAs, contribuindo para a motivação dos alunos e aprendizagem colaborativa.
- ❑ Identificar características de um indivíduo, como os traços de personalidade, a partir da avaliação de textos coletados das ferramentas de comunicação.
- ❑ Ter uma ferramenta que pode ser integrada a sistemas de recomendação baseada no perfil pessoal, sem que haja necessidade de abordagem direta ao indivíduo sendo analisado.
- ❑ Definir uma estrutura de caso e estratégias de funcionamento do RBC, que fossem capazes de sugerir novos bons grupos. Ainda que os novos indivíduos a serem agrupados não correspondam exatamente àqueles representados na base, mesmo assim seria possível sugerir um bom agrupamento. Com a possibilidade de evolução da base pela inclusão de novos casos, as regras mineradas da base devem também ser melhoradas.
- ❑ Obter um conjunto de regras de formação de grupo de alunos, baseadas nos traços de personalidade, que recomendem grupos com interação e troca de conhecimento satisfatórias, durante a execução de atividade colaborativa. Tudo isso de forma transparente, sem que o aluno perceba como os traços de personalidade e comportamento são avaliados.

1.5 Organização da Dissertação

Esta dissertação está organizada em seis capítulos. No Capítulo 2 é apresentada a revisão de literatura abordando os conceitos teóricos que fundamentam a proposta e trabalhos relacionados. O Capítulo 3 apresenta a visão geral do trabalho proposto, incluindo o processamento do texto e os passos envolvidos no processo de formação de grupos e, também, a abordagem adotada para uso do RBC, tratando da estrutura do caso adotada para descrever o problema de avaliar um grupo como bom ou não e o funcionamento da estrutura do RBC como um todo.

Os Capítulos 4 e 5 apresentam os experimentos conduzidos para validar a proposta. No Capítulo 4, são apresentados os experimentos relativos à formação da base de treinamento para o modelo de detecção dos traços de personalidade, os experimentos com algoritmos de aprendizagem que poderiam ser empregados na detecção e os resultados encontrados. Os experimentos do Capítulo 5 buscaram investigar a validade do uso do RBC e, por isso, envolveram a construção de uma base inicial de casos para o RBC, a formação de grupos a partir de consultas à base de casos e mineração de regras na base. O Capítulo 6 encerra

a apresentação do trabalho, trazendo as considerações finais, contribuições e perspectivas futuras.

Revisão da Literatura

De acordo com os objetivos apresentados no Capítulo 1, técnicas de Inteligência Artificial (IA) foram necessárias bem como alguns fundamentos sobre teoria da personalidade para o desenvolvimento da proposta do Capítulo 3. Por isso, esta revisão contém tanto os aspectos teóricos e técnicas (Seção 2.1) em que se baseia esta proposta, quanto os trabalhos relacionados (Seção 2.2) que envolvem a aplicação do modelo *Big Five*, processamento de linguagem natural e Raciocínio Baseado em Casos.

2.1 Fundamentação Teórica

Este trabalho se fundamenta no uso dos traços de personalidade e formação de grupos (também chamado de agrupamento) para aprendizagem colaborativa. O que se entende por aprendizagem colaborativa e como a formação de grupos se relaciona com sua efetividade no apoio ao processo de aprendizagem são apresentados na Seção 2.1.1. A Seção 2.1.2 descreve a teoria sobre traços de personalidade avaliados pelos inventários de personalidade do *Big Five*, que foram empregados na formação de grupos.

O desenvolvimento da proposta ainda se apoia em técnicas de IA para detecção das características dos alunos através de texto escrito. Assim, a Seção 2.1.3 traz os aspectos relacionados ao processamento de linguagem natural (texto). Na Seção 2.1.4 são descritos os conceitos de Raciocínio Baseado em Casos e suas etapas na solução de um problema.

2.1.1 Aprendizagem Colaborativa e Formação de grupos

Modelos cognitivos de aprendizagem consideram que a aprendizagem é um processo ativo, construtivo e orientado à meta, assim, a aprendizagem é mais efetiva ao envolver os estudantes na construção do conhecimento por meio da aquisição, análise, e manipulação da informação (ALAVI, 1994). Algumas teorias, como por exemplo (VYGOTSKY, 1978), ainda enfatizam o aprendizado como um processo social que ocorre de forma mais efetiva por meio de interações interpessoais.

Segundo a proposta de (VYGOTSKY, 1978), o aprendizado desperta uma variedade de processos internos de desenvolvimento que só são capazes de operar por meio da interação com as pessoas no ambiente. Na criança, observa-se que ao adquirir a capacidade de uso de linguagem, essa passa a ser um importante meio de contato social com outras pessoas e por meio da qual consegue ferramentas auxiliares na solução de tarefas difíceis. O caminho que liga o aprendiz ao objeto da aprendizagem passa por outra pessoa por meio da linguagem.

Conforme observado por (ALAVI, 1994), a aprendizagem efetiva está ligada a três fatores: 1) aprendizagem ativa e construção do conhecimento, 2) cooperação e trabalho em equipe, e 3) aprendizagem usando resolução de problemas. O trabalho com outras pessoas expõe o indivíduo a pontos de vista diferentes, questionando o entendimento inicial e motivando-o. Assim, o aprendizado é visto como resultado de situações que envolvam resolução de problemas desafiadores nos quais o modelo mental dos estudantes é testado, expandido e refinado até encontrar solução efetiva e confiável.

A aprendizagem colaborativa efetiva é baseada nessas três premissas sobre o processo de aprendizagem efetiva. Ela envolve processos sociais por meio dos quais um pequeno grupo de estudantes trabalham juntos para completar uma atividade desenvolvida para promover o aprendizado (ALAVI, 1994). Aprendizagem colaborativa pode ser definida como a construção conjunta de significado, por meio da interação com outros e pode ser caracterizada por uma união para atingir uma meta compartilhada (HRON; FRIEDRICH, 2003).

Em (LAMPRECHT; SIEMON; ROBRA-BISSANTZ, 2016), a cooperação é definida como dois parceiros, pelo menos, interagindo em uma relação baseada no acordo mútuo sobre as respectivas contribuições e os resultados que devem ser entregues para atingir a meta definida em conjunto. Nesse contexto, a cooperação não requer uma estratégia compartilhada para usar de forma eficiente os recursos e capacidades disponíveis. Cada um adota uma estratégia individual para resolver sua parte do problema, ainda que sua estratégia possa afetar os resultados dos demais membros do grupo.

Esse comportamento pode produzir mais resultados mas, sem uma estratégia compartilhada em que todos colaborem entre si, todos podem adquirir menos conhecimento. Cada um faz o que pode, porém perde a chance de aprender por meio da discussão com os demais. Por outro lado, para que haja colaboração, é necessário que os indivíduos em um mesmo grupo cooperem entre si. Assim, além das interações e troca de conhecimento, a aprendizagem colaborativa só será efetiva se todos executarem uma parte da solução.

Para aumentar as oportunidades de aprendizado colaborativo efetivo usando tecnologias de computador, o grupo deve estar engajado na solução da atividade e produzir algum resultado, como por exemplo um relatório textual ou apresentações verbais integrando as idéias discutidas pelo grupo (STAHL, 2006). Assim, considera-se também como fator importante para verificar a efetividade de um grupo, a capacidade de cumprir o prazo

proposto e entregar uma solução resultante das ideias discutidas em grupo.

Conforme definido por (MAGNISALIS; DEMETRIADIS; KARAKOSTAS, 2011), sistemas que utilizam tecnologias e métodos para implementar uma estratégia específica de suporte aos aprendizes (objetivo pedagógico), utilizando suporte adaptativo ou baseado em inteligência (alvo de intervenção) e representação (modelo) de aspectos chaves do ambiente de aprendizagem, são chamados de Sistemas Adaptativos e Inteligentes para Suporte à Aprendizagem Colaborativa, em inglês *Adaptive and Intelligent Systems for Collaborative Learning Support* (AICLS). Nesses sistemas, o alvo de intervenção se refere àquilo que será adaptado ou alvo do suporte baseado em inteligência.

O alvo de intervenção em um AICLS pode ser na preparação da atividade ou no suporte à execução da atividade, por meio de apoio de domínio ou da intervenção na interação entre pares. Nesse contexto a formação de grupos entra como suporte à preparação da atividade colaborativa (MAGNISALIS; DEMETRIADIS; KARAKOSTAS, 2011). A forma como o grupo é montado para a atividade colaborativa pode aumentar a probabilidade de ocorrência de interação produtiva ou aumentar os conflitos sociais que poderiam tornar a colaboração impraticável (KWON; LIUB; JOHNSONC, 2014).

A formação do grupo influencia o trabalho conjunto para alcançar um objetivo, sendo esse um dos aspectos mais importantes para abrir espaço para os efeitos positivos da colaboração na aprendizagem (MANSKE et al., 2015). Embora a colaboração tenha efeito positivo na aprendizagem, há vários fatores a cerca da construção de um grupo que influenciam o processo de colaboração e devem ser levados em conta no agrupamento (LAI, 2011). Por exemplo, as características dos alunos, como traços de personalidade, as características da atividade, como o tempo de resolução, e a quantidade de membros em um grupo.

2.1.2 *Big Five*

Dentre as várias propostas na literatura para medir aspectos de personalidade, a construção psicométrica mais usada é o *Big Five* (ALTANOPOULOU; TSELIOS, 2015). O *Big Five*, também conhecido como *Five Factor Model*, funciona como uma organização hierárquica de traços de personalidade em dimensões básicas que resumem um grande número de características de personalidade distintas e específicas (MCCRAE; JOHN, 1992). O termo *Big Five*, geralmente atribuído a (GOLDBERG, 1981), foi criado para expressar essa abrangência na representação da personalidade.

De acordo com (JOHN; SRIVASTAVA, 1999), esse modelo surgiu como resultado de vários estudos sobre os termos de linguagem que são usados para falar de si próprio e sobre outras pessoas. Tais estudos convergiram para cinco traços de personalidade que representam a personalidade em nível de abstração mais abrangente. Apesar de haver alguma diferença na definição dos termos que nomeiam cada dimensão na literatura relacionada, trabalhos recentes como (MAPLES et al.,) usam com maior frequência para

definir os cinco traços as seguintes denominações: Openness to experience, Conscientiousness, Extraversion, Agreeableness e Neuroticism. Há uma variação no nome dado as cinco dimensões em Português, o que pode ser notado em (PAIM; CAMATI; ENEMBRECK, 2016) e (ANDRADE, 2008). Neste trabalho, os cinco traços serão referenciados como Abertura, Conscienciosidade, Extroversão, Afabilidade e Neuroticismo.

Pessoas com grau mais alto de Abertura são curiosas, questionadoras, interessadas em novas ideias e em artes. Conscienciosidade representa pessoas que são responsáveis e bem organizadas. Extroversão caracteriza pessoas que no geral demonstram emoções positivas, gostam de estar com outras pessoas e conversar bastante. Afabilidade é relacionada a ser cooperativo, amável, sociável e agradável. Neuroticismo é geralmente relacionado àqueles indivíduos que expressam alto grau de emoções negativas e que são geralmente hostis, desconfiados e emocionalmente instáveis.

Pesquisadores que desenvolveram trabalhos envolvendo *Big Five* (DIGMAN, 1990; JOHN; SRIVASTAVA, 1999; MCCRAE; JOHN, 1992) afirmam que os cinco fatores, sozinhos ou combinados, podem ser encontrados em praticamente todos os instrumentos de avaliação de personalidade. Ainda que o modelo seja muito amplo para explicar todas as possíveis diferenças de personalidade, foi suficientemente estudado e testado em um grande número de culturas e é considerado uma ferramenta confiável para medir traços de personalidade.

Os traços de personalidade do *Big Five* são bipolares. Assim, as características citadas são mais frequentes em pessoas com escore alto naquela dimensão, ou traço de personalidade. Escores baixos significam que o indivíduo tende a expressar as características opostas. Há ainda indivíduos cujos escores encontram-se em um valor médio, entre os dois pólos de cada traço. Nesse caso, haverá presença das características de ambos os pólos, caracterizando um perfil de transição ou intermediário.

As características definidas por cada um dos traços fornecem informações importantes para construção de grupos para colaboração. Por exemplo, indivíduos com alto grau de Abertura são vistos como benéficos para os grupos pois, dada a característica de interesse por conhecimentos novos, é um tipo de parceiro de grupo que pode ser mais criativo e trazer novas opções de solução de problemas. Indivíduos com baixa Abertura gostam de rotina e processos bem definidos, são bons para cumprir tarefas bem definidas. Apesar de não muito favoráveis a ideias novas, talvez pouco criativos, são vistos como bons parceiros de trabalho.

Pessoas altamente conscienciosas, podem contribuir com o grupo com a sua capacidade de planejamento e foco na meta. São características que ajudam a garantir que o grupo obtenha sucesso no cumprimento do prazo estabelecido para a solução da tarefa colaborativa. Apesar de mais desorganizados e menos preocupados com objetivos e metas, pessoas com baixo grau de Conscienciosidade apresentam também melhor capacidade de improviso e criatividade para encontrar soluções quando não há um plano bem definido.

Desde que a interação é importante para a colaboração, o aluno extrovertido é visto com um tipo de integrante interessante por gostar mais de interagir com outras pessoas. Normalmente, são indivíduos que iniciam conversas e que funcionam como elemento de ligação em um grupo. Por outro lado, aqueles com baixo escore em Extroversão tendem a ser mais focados e reflexivos. Apesar de preferirem trabalhar sozinhos, quando se sentem à vontade para interagir, são indivíduos que podem contribuir com a análise do problema a ser resolvido em grupo.

As pessoas com alto escore em Afabilidade apresentam todas as características tidas como desejáveis ao trabalho em grupo. Já aquelas com baixa Afabilidade têm tendência a cooperar menos. Por outro lado, caso o laço social seja percebido como positivo para atingir os próprios objetivos, também irão interagir e criar laços sociais.

O traço de Neuroticismo, apesar de remeter a atributos negativos, não significa que alto grau de Neuroticismo indique algum problema psicológico ou social. Alto grau de Neuroticismo implica em uma pessoa mais preocupada com o que os outros pensam e menos tolerante ao estresse, aumentando a chance desses indivíduos serem mais propensos a demonstrar emoções negativas. Por outro lado, essa característica também faz com que cuidem mais das relações sociais para evitar conflitos ou situações que os deixem agitados. Apesar de desconfiados, quando se trata de abordar outras pessoas para formar novos laços, são também mais propensos a manter e cuidar dos laços sociais já formados. Assim, podem ser bons parceiros de grupo quando encontram ambiente favorável. Quando o grau de Neuroticismo é baixo, há mais tolerância ao estresse.

Todas as características associadas aos traços de personalidade podem ser mais ou menos apropriadas em um grupo, dependendo dos traços de seus membros. Por esse motivo, o *Big Five* é também uma boa ferramenta para ajudar a compreender o comportamento dos alunos em grupo. Conhecendo essas características é possível ajudá-los a encontrar parceiros adequados para atividade em grupo e tirar proveito da aprendizagem colaborativa. A proposta apresentada no Capítulo 3 foi desenvolvida buscando combinar tais características na formação de grupos de alunos para trabalhar em atividades colaborativas.

2.1.3 Extração de Informações do Texto

Como o texto é um dado de natureza não estruturada precisa passar por alguma forma de tratamento para, então, ser empregado como fonte de informação na detecção de características pessoais do seu autor. Uma das formas de obter características a partir do texto é a utilização de léxicos que associam palavras extraídas do texto às categorias. Essa associação permite sua utilização, por exemplo, para análise de sentimentos, extração de informação ou categorização textual, problemas tipicamente relacionados à área de Processamento de Linguagem Natural (PLN) (CHOWDHURY, 2003).

Existem vários dicionários e léxicos disponíveis para a classificação de palavras em categorias, como emoções positivas e negativas, que é uma das abordagens utilizadas para análise de sentimento (LIU, 2010). Dentre eles, estão o léxico *WordnetAffectBr* e o dicionário interno da ferramenta *Linguistic Inquiry and Word Count* (LIWC). Ambos trazem listas de categorias de palavras validadas para o Português Brasileiro e que permitem a classificação de palavras extraídas de textos dentro de suas categorias.

O *WordnetAffectBr* foi desenvolvido a partir do trabalho de (PASQUALOTTI; VEIRA, 2008), integrando as áreas de Computação Afetiva e PLN no desenvolvimento de uma base lexical de palavras de emoções para a Língua Portuguesa. Apesar de permitir a classificação de palavras em categorias de emoção, não possui outras categorias que poderiam ser empregadas, por exemplo, na identificação de traços de personalidade. Nessa tarefa, as palavras relacionadas à emoção ajudam, mas ainda é necessário outras características relacionadas ao comportamento de um indivíduo para permitir uma melhor classificação.

O dicionário interno do LIWC (TAUSCZIK; PENNEBAKER, 2010) é composto de categorias de palavras que expressam tanto a estrutura gramatical a qual a palavra pertence quanto processos psicológicos e cognitivos, incluindo também as emoções. Ao fornecer outras categorias que permitem classificar as palavras que expressam processos cognitivos e psicológicos, pode ser empregado na classificação dos traços de personalidade do autor do texto. A ferramenta e o dicionário foram, originalmente, desenvolvidos para analisar textos em inglês. No entanto, na versão LIWC 2007 foi incorporada uma versão do dicionário em Português Brasileiro (FILHO; ALUÍSIO; PARDO, 2013).

Outro método utilizado para estruturação de texto é o uso de modelos de palavras como saco de palavras (*bag-of-words*) ou *n-grams*. Wen et al. (WEN; YANG; ROSE, 2014) enfatizam que o uso de análises baseadas em características linguísticas definidas na literatura, por exemplo, categorias de palavras, é mais efetivo na prática e de mais fácil generalização para outros domínios do que modelos de palavras como *bag-of-words*.

2.1.4 Raciocínio Baseado em Casos

O Raciocínio Baseado em Casos (RBC) usa experiências anteriores para encontrar uma solução para um novo problema (KOLODNER, 1992). O processo de aprendizagem no RBC ocorre de duas maneiras: a primeira é a adaptação de soluções antigas de forma a responder o novo problema, não sendo necessário, por exemplo, executar um processo de treinamento para lidar com ele; a segunda forma é a melhoria da capacidade de derivar novas respostas à medida que a base é atualizada (POOLE; MACKWORTH, 2010).

A representação do problema e sua respectiva solução na base de conhecimento do RBC é chamada de caso. O RBC usa os casos armazenados no cálculo da nova solução sem que haja necessidade de transformar a base de treinamento em uma estrutura secundária como árvores de decisão ou redes neurais (POOLE; MACKWORTH, 2010). Para tanto, a

implementação do RBC precisa de uma base inicial de soluções relevantes para o contexto representado.

De forma resumida, o RBC atua por meio de quatro tarefas básicas: recuperação, reúso, revisão e retenção de casos. Nesse processo, existem quatro elementos: um problema, uma consulta, os casos da Base de Casos (BC) e uma solução. Em primeiro lugar, o problema é convertido em uma consulta que aproxima a estrutura do caso. O processo do RBC continua dessa consulta até obter uma solução. Para encontrar a solução de um novo problema, usa-se algum mecanismo para indexar as soluções e um critério para calcular a proximidade com os casos da base.

O critério de comparação usado na varredura da base para encontrar uma solução para a consulta (recuperação), comparando as características do problema apresentados com os casos da base, é a similaridade entre os casos. Uma estratégia que pode ser usada para encontrar os casos mais próximos é o algoritmo k-vizinhos mais próximos. Nesse caso, o cálculo da proximidade dos casos pode ser feita com a distância euclidiana, por exemplo (POOLE; MACKWORTH, 2010).

A resposta a uma consulta será a solução daqueles casos que são mais semelhantes ao problema descrito na consulta. Assim, ao recuperar os casos mais próximos para resolver uma nova situação, inicia-se o processo de interpretação por meio do qual a qualidade daquela solução no novo contexto e a necessidade de adaptação são avaliadas (reúso). Respostas interessantes serão produto das estratégias empregadas na adaptação (KOLODNER, 1992).

A fim de avaliar e corrigir as experiências já conhecidas, é necessário que a implementação do RBC também considere alguma forma de avaliação das suas soluções (revisão). Neste trabalho, as informações empregadas na revisão de uma solução são o desempenho e características dos grupos, avaliados durante alguma atividade colaborativa. Uma vez revisada, a solução pode então ser adicionada à BC atual ou não (retenção). A Figura 1 mostra essas etapas do RBC, que são descritas a seguir. Esse comportamento será aproveitado na sugestão de grupos e na melhoria das regras de agrupamento com base no desempenho dos grupos em atividade colaborativa.

2.1.4.1 Medidas de Similaridade

As medidas de similaridade são as funções empregadas para calcular o quão próximos estão os casos dos dados fornecidos em uma consulta. Os valores retornados pelas medidas de similaridade são entre 0% e 100%, onde 0% indica que o caso e a consulta são completamente diferentes e 100% que são completamente iguais. Como a representação dos casos dentro de um dado contexto pode assumir tanto uma representação simples, como um vetor de variáveis, quanto uma estrutura formada por mais de um objeto (BERGMANN; STAHL, 1998), o cálculo da proximidade ou similaridade entre os casos e o novo problema depende dessa representação.

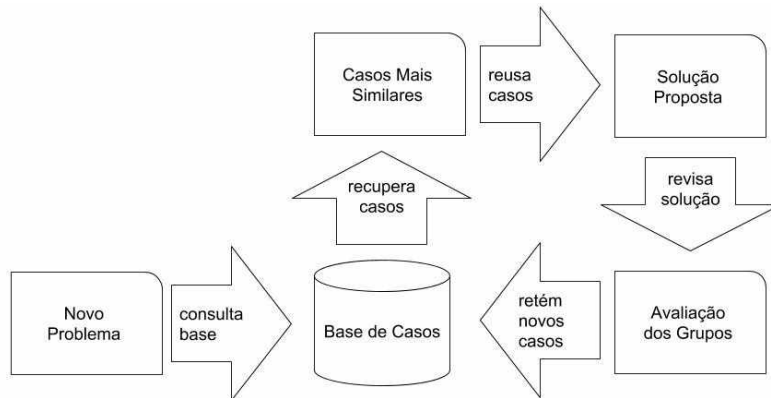


Figura 1 – Funcionamento do Raciocínio Baseado em Casos.

Para lidar com os possíveis tipos de dados presentes nos casos, Stahl (STAHL, 2003) dividiu o cálculo da similaridade entre caso e consulta, em duas etapas: o cálculo da similaridade local e o cálculo da similaridade global. A similaridade global é a função que agrega as similaridades locais calculadas e pode ser, por exemplo, uma soma ponderada dos resultados. A similaridade local diz respeito ao cálculo da distância entre os pares de elementos que compõe o caso e a consulta, a fim de definir o quão próximos são os pares.

A similaridade local depende do tipo de dado do elemento do caso. Por exemplo, se um dos elementos do caso é uma variável binária, o valor da similaridade poderia ser atribuído por uma função que retorna 100% de similaridade, casos sejam iguais, ou 0%, se fossem diferentes. Para valores contínuos, por exemplo, a similaridade pode ser calculada usando diferenças entre o valor no caso e na consulta como entrada para uma função *sigmoid*, por exemplo.

Uma consulta que não é exatamente igual aos casos da base terá uma solução parecida com a consulta, embora não tenha 100% de similaridade. Nessa situação, pode ocorrer diferença entre medidas de similaridade. Desde que o emprego de uma medida pode ser mais vantajosa que outra, este trabalho considerou a possibilidade de troca das funções que calculam a similaridade no momento da consulta. Uma medida comum empregada no cálculo da similaridade é a distância entre os objetos de comparação.

Duas funções estão disponíveis para o cálculo da distância (Distância Linear e Distância Logarítmica) e quatro opções de funções que usam as distâncias calculadas para devolver a similaridade (*Threshold*, Linear, Exponencial e *Sigmoid*), conforme descrições na Subseção 2.1.4.2. As funções que agregam os valores das similaridades entre atributos em um único valor de similaridade para o objeto se encontram na Subseção 2.1.4.3.

2.1.4.2 Medida de Similaridade entre Atributos

O valor da similaridade entre os atributos de um objeto é definido a partir do cálculo da diferença entre o valor de tal atributo para o objeto no caso e na consulta. Essa função

de diferença será chamada, a partir deste ponto, de $dif(c_i, q_i)$, onde c_i é o atributo i do objeto no caso e q_i , o mesmo atributo i no objeto da consulta. Essa diferença pode ser calculada conforme Eq. 1 e Eq. 2.

A função linear pode ser considerada a forma mais natural de calcular a diferença entre dois valores, pois apenas subtrai um do outro. Na função logarítmica, a diferença ocorre entre os logaritmos (ln) dos valores, a fim de capturar alguma nuance dos dados dando mais força à diferença quando os valores dos atributos são maiores. Se algum valor é igual a zero, $ln(0)$ é considerado 0.

$$\text{Função linear:} \quad dif(c_i, q_i) = |c_i - q_i| \quad (1)$$

$$\text{Função logarítmica:} \quad dif(c_i, q_i) = |ln(c_i) - ln(q_i)| \quad (2)$$

Em ambos os casos, a função módulo é aplicada para devolver o valor absoluto da diferença. Como os valores dos atributos para os alunos são valores numéricos que não necessariamente pertencem a $[0, 1]$, os valores de c_i e q_i foram normalizados para esse intervalo. Dessa forma, todas as funções de similaridade serão capazes de retornar um valor entre 0 e 1.

A diferença calculada através de uma das funções apresentadas acima é empregada no cálculo do valor da similaridade. O valor da similaridade retornada está no intervalo $[0, 1]$, correspondendo a valores entre 0% e 100%. A similaridade pode ser calculada usando as funções *Threshold* (Eq. 3), *Linear* (Eq. 4), *Exponencial* (Eq. 5) ou *Sigmoid* (Eq. 6).

$$\text{Função Threshold:} \quad sim(c_i, q_i) = \begin{cases} 1, & \text{se } dif(q_i, c_i) \leq t \\ 0, & \text{caso contrário} \end{cases}, \quad (3)$$

onde t representa o limite (*threshold*) de $dif(q_i, c_i)$ abaixo do qual a similaridade será 1 (100%). O valor de t foi definido como 0. Assim se $dif(q_i, c_i)=0$, a similaridade será 100% e se $dif(q_i, c_i)>0$, será 0%.

$$\text{Função Linear:} \quad sim(c_i, q_i) = \frac{\max - dif(q_i, c_i)}{\max - \min} \quad (4)$$

$$\text{Função Exponencial:} \quad sim(c_i, q_i) = e^{dif(c_i, q_i) * \alpha}, \quad (5)$$

onde $\alpha = -1$ e $dif(q_i, c_i) \in [0, 1]$. Como a função de similaridade precisa retornar um valor entre 0 (0%) e 1 (100%), devido às características da função, se $\alpha = 1$, por exemplo, o resultado seria algo entre 1 e 2.7. Caso $dif(q_i, c_i)$ seja igual a 0, 0.5 ou 1, com $\alpha = -1$, a função exponencial retorna 1, 0.61 e 0.37, que são valores no intervalo desejado.

Nas outras funções, quando a distância é 1, a similaridade devolvida é 0. No entanto, é importante observar que, usando a função exponencial, o valor de similaridade devolvido será 0.37. Assim, haverá um aumento do valor da similaridade entre caso e consulta muito

diferentes e da chance de trazer uma solução que, na verdade, não é similar à consulta. Por outro lado, a possibilidade de encontrar uma solução é maior quando a consulta é completamente diferente dos casos na base. Pode ser interessante ter uma solução mesmo quando não há casos similares para responder apropriadamente à consulta.

$$\text{Função Sigmoid:} \quad \text{sim}(c_i, q_i) = \frac{1}{1 + e^{\frac{\text{dif}(c_i, q_i) - \theta}{\alpha}}} , \quad (6)$$

onde $\alpha = 0.01$ e $\theta = 0.5$, sendo θ o valor do ponto central indicando mudança de direção da curva. Como a diferença é normalizada em $[0, 1]$, o valor central foi definido como 0.5. Se a amplitude da curva for muito grande, os valores de similaridade estariam fora do intervalo. Com $\alpha = 0.01$, para um $\theta = 0.5$, estará no intervalo $[0, 1]$. Por outro lado, a similaridade tende a ser 0 ou 1. Valores entre 0 e 1 somente serão devolvidos quando o valor da diferença for muito próximo do valor central. Por exemplo, se a distância entre dois atributos é 0.7 a similaridade será 0.

Outra situação que precisa ser considerada no cálculo da similaridade entre dois atributos é a inexistência de um valor atribuído ao atributo, seja no caso ou na consulta. Nesses casos, a comparação entre os atributos segue basicamente duas estratégias. A primeira, apresentada por (RICCI; AVESANI, 1995), considera $\text{sim}(c_i, q_i) = 0.5$ sempre que c_i ou q_i tiverem valor desconhecido. A segunda estratégia (SURMA; VANHOOF, 1995) atribui distância 0 quando ambos os valores são desconhecidos ($\text{sim}(c_i, q_i) = 1$) e distância 1, quando um deles é conhecido ($\text{sim}(c_i, q_i) = 0$).

2.1.4.3 Medida de Similaridade entre Objetos (MSO)

Todos os componentes de um caso são objetos representados por um vetor de atributos. O valor de similaridade entre dois objetos é dada pela agregação das similaridades entre atributos em um único valor. Para realizar tal agregação, duas opções de funções de agregação foram implementadas, Minkowski (Eq. 7) e *Simple Matching* (Eq. 8).

$$\text{Minkowski:} \quad \left(\frac{\sum_{i=1}^n \text{sim}(c_i, q_i)^{\text{grau}}}{n} \right)^{\frac{1}{\text{grau}}} , \quad (7)$$

onde n é o número de atributos. Na Eq. 7 os valores numéricos são agregados usando operação de exponenciação e radiciação conforme o *grau* especificado. Por exemplo, na Eq. 7, quando *grau* = 1, tem-se a distância de Manhattan e a similaridade corresponde à média das similaridades entre os pares de atributos do objeto. Se *grau* = 2, o valor obtido é agregado conforme Distância Euclidiana.

$$\text{Simple Matching:} \quad \frac{\#igual}{n} , \quad (8)$$

onde n é o número de atributos e $\#igual$ é número de atributos iguais. De acordo com a função *Simple Matching*, similaridade é a taxa de atributos do objeto que tinham valores iguais. Esse tipo de função funciona melhor para atributos de valor binário, onde só

há a possibilidade de ser igual ou diferente. Para valores contínuos, é mais interessante considerar a similaridade como sendo a proximidade dos valores, mesmo se não forem exatamente iguais.

2.2 Trabalhos Relacionados

Os objetivos deste trabalho envolvem a formação de grupos para colaboração e detecção automática das características dos alunos a partir de textos escritos. Os trabalhos relacionados a estes objetivos são apresentados, respectivamente, nas Subseções 2.2.1 e 2.2.2.

2.2.1 Formação de Grupos para Colaboração

A efetividade do grupo depende da coordenação de esforços dos seus membros para trabalhar e aprender juntos, ou seja, depende das interações sociais que estão desenvolvendo (DILLENBOURG, 1999). Nesse contexto, a composição do grupo se torna crucial para favorecer interações produtivas entre os pares (MAGNISALIS; DEMETRIADIS; KARAKOSTAS, 2011), já que o agrupamento descuidado pode desencadear situações indesejáveis, como o isolamento de algum membro do grupo. Assim, é necessário considerar uma estratégia de formação de grupo baseada em critérios que favoreçam a colaboração.

Alguns trabalhos, como (WANG; LIN; SUN, 2007) e (SILVA; MOTTA; SANTORO, 2010), reforçam essa ideia apontando os grupos formados automaticamente, usando algum critério de agrupamento, como melhores do que aqueles formados sem critério. Em estudo conduzido com alunos de Engenharia de Computação, (CARRO; SANCHEZ-HORREO, 2017) perceberam influência de características relacionadas à personalidade do aluno, inteligência e estilos de aprendizagem no seu desempenho. Seus resultados apontam critérios de agrupamento específicos, como os citados, podem levar os alunos a experiências de colaboração bem-sucedidas.

A capacidade social associada a fatores pedagógicos e tecnológicos vão influenciar a aprendizagem do aluno, conforme observado em (XING; KIM; GOGGIN, 2015). Eles concluíram que a habilidade social é o fator mais relevante para a interação social e o aprendizado dos alunos. Então, como traços de personalidade descrevem características que podem ser associadas à habilidade social do aluno, eles vêm sendo empregados para dar suporte a formação de bons grupos para aprendizagem colaborativa.

Considerando a hipótese de que determinado traço de personalidade poderia facilitar ou atrapalhar a aprendizagem, o estudo conduzido por (PACHECO; SISTO, 2003) avaliou crianças de escolas públicas em relação a quatro traço de personalidade. De acordo com os resultados, há relação entre traços de personalidade e aprendizagem por conflito sócio-cognitivo, resultante da interação entre indivíduos e espera-se, por exemplo, que crianças

com maior Extroversão se exponham mais a conflitos sócio-cognitivos, por estarem mais abertas à interação e apresentarem maior grau aprendizagem.

Em (ROBERTS et al., 2008), os resultados sobre a influência de Extroversão e Neuroticismo na capacidade de socialização indicam que a Extroversão está relacionada com alta intensidade de atividade na rede social de um indivíduo, como solicitar ajuda aos outros e buscar novas as conexões. Já o Neuroticismo está relacionado com baixa intensidade de atividade na rede social. Como extrovertidos são mais eficientes na construção de novas conexões ao mudar de contexto social, conseguem se integrar mais facilmente a novos grupos e, por isso, podem atuar como aqueles que estabelecem a ligação com os demais membros de um grupo.

(SPOELSTRA et al., 2015) propõem o agrupamento de indivíduos conscienciosos, além de usar suas preferências e conhecimentos, para formar grupos produtivos e de aprendizagem. Como os conscienciosos tendem à organização e são orientados à meta, grupos formados com esse traço devem ser mais produtivos e eficientes quanto à realização da tarefa. Além disso, a dedicação presente nesse traço permite ganhos maiores de aprendizagem em grupos com alunos mais conscienciosos. Também sugerem grupos compostos por alunos com diferentes graus de Conscienciosidade como bons para gerar soluções criativas, conduzindo a grupos produtivos.

Apesar de encontrar relação positiva entre Conscienciosidade e o ganho de aprendizagem de alunos em atividade colaborativa, os resultados de (ALTANOPOULOU; TSELIOS, 2015) sugerem que aqueles com alto grau de Conscienciosidade devem ser distribuídos entre os grupos para que influenciem a habilidade de cumprir o prazo proposto. Por outro lado, de acordo com (ALTANOPOULOU; TSELIOS, 2015), grupos com apenas indivíduos altamente extrovertidos podem influenciar negativamente o ganho de aprendizado, já que extrovertidos tendem a se distrair com a interação social.

Mesmo estudos conduzidos em áreas não relacionadas à Educação, observa-se que os traços de personalidade têm papel importante nas interações entre os indivíduos e sobre a colaboração. Grant et al. (GRANT, 2013), por exemplo, estudaram os efeitos da Extroversão sobre os resultados de venda, percebendo que alto grau desse traço faz o indivíduo permanecer focado na própria perspectiva e desconsiderar necessidades, interesses e valores do outro. Esse comportamento também pode ser entendido pelo interlocutor como invasivo ou agressivo, dificultando a comunicação.

O estudo de (AKHTAR et al., 2015) identificou principalmente a relação positiva entre engajamento e inteligência emocional. Para os traços de personalidade medidos com um inventário de personalidade *Big Five*, notou Abertura, Extroversão e Conscienciosidade como preditores para o engajamento. Entre os traços de personalidade, a Abertura é apontada como o segundo preditor mais forte para o engajamento no trabalho e atribuem esse resultado à capacidade de indivíduos com alto grau de Abertura lidarem de forma eficiente com o ambiente.

O estudo de (BOZIONELOS, 2017) investigou a relação entre troca de recursos de trabalho na rede de conexões no trabalho e a personalidade. Os resultados apontaram pontuações médias nos traços de Neuroticismo, Abertura e Afabilidade como tendo menor vantagem no compartilhamento de recursos de rede comparados àqueles com escores extremos nesses traços. Para Conscienciosidade o resultado foi o oposto. Essa inversão é atribuída à tendência à irresponsabilidade, quando há baixa Conscienciosidade, e à negligência das relações sociais por excesso de preocupação com as próprias atividades no outro extremo do traço.

(BOZIONELOS, 2017) destaca que, apesar de indivíduos com alto Neuroticismo serem menos propensos a se relacionar devido a hostilidade, desconfiança e evitamento de socialização, são indivíduos muito preocupados com normas e convenções sociais, facilitando a manutenção dos laços já adquiridos. Ou seja, se o outro parceiro de grupo estabelece comunicação e aproximação, há muita chance de estabelecerem boa interação a partir daí para trabalharem juntos e compartilhar recursos.

De acordo com os resultados, apesar do perfil teoricamente ruim para formação de laços sociais, tanto alto quanto baixo grau de Neuroticismo permitem sucesso no compartilhamento de recurso de rede, ao contrário daqueles com escores intermediários. O mesmo se observou para os níveis baixos de Abertura e Afabilidade. Aqueles com mais desvantagem em relação aos recursos de rede foram os que obtiveram escores intermediários nos traços.

Também é interessante evitar combinações de alunos em que alguém possa se sentir intimidado e deixar de contribuir. Por exemplo, uma característica de alta Extroversão é o excesso de foco na própria necessidade de falar, fazendo com que alguém com baixa Extroversão ou alto Neuroticismo evite participar das discussões do grupo. Em relação à Conscienciosidade, também seriam mais adequados grupos com indivíduos que apresentem diferentes níveis nesse traço para que colaborem entre si.

Quanto à distribuição das características dos alunos no grupo, o estudo de (SANTOS et al., 2003) mostrou que grupos heterogêneos funcionam melhor do que aqueles onde há alunos semelhantes. Eles perceberam que grupos homogêneos demoram para colaborar de maneira efetiva. O estudo de caso de (RUTHERFOORD, 2006) também concluiu que grupos heterogêneos quanto à personalidade são melhores.

Em um grupo homogêneo quanto aos traços de personalidade, todos os membros terão as mesmas habilidades e fraquezas, não havendo membro para contrabalancear essas características. Se o grupo é heterogêneo, há características distintas e maior variabilidade de forças e fraquezas, tornando o grupo capaz de gerenciar os problemas.

Os resultados obtidos nos estudos citados sugerem maneiras de combinar indivíduos considerando uma característica específica:

- Grupos cujos membros tem diferentes níveis de Conscienciosidade podem ser efetivos e criativos.

- ❑ Grupos onde todos os membros tem alto grau de Abertura também formam grupos efetivos, com indivíduos engajados no trabalho.
- ❑ Grupos com todos os indivíduos extrovertidos devem ser evitados, porque alunos extrovertidos são mais propensos a negligenciar o objetivo comum devido à interação social.
- ❑ Grupos com todos os membros com alto grau de Conscienciosidade devem ser evitados pois, nesse caso, há a possibilidade dos alunos focarem nas suas próprias atividades e negligenciarem a interação social.

Nenhum dos estudos anteriores foram conclusivos no sentido de apontar objetivamente estratégias de agrupamento, a partir dos traços de personalidade. De acordo com o estudo de (REIS et al., 2015a), conduzido para mapear o uso de estados afetivos na formação de grupos em AVAs com aprendizagem colaborativa, apesar dos vários trabalhos encontrados, a maioria daqueles que usaram traços de personalidade ainda são estudos iniciais e dão abertura para explorar o assunto.

Os estudos anteriores também não trazem resultados sobre a influência que a combinação de características de em um mesmo indivíduo pode ter sobre o grau de influência de um determinado traço no grupo. Por exemplo, grupos formados apenas por indivíduos extrovertidos poderiam atingir bons resultados, se todos os membros do grupo fossem também conscienciosos.

Neste trabalho o uso de Raciocínio Baseado em Casos (RBC) para dar suporte à sugestão de novos grupos é apontado como uma boa alternativa para entender como os alunos se organizam para solucionar uma atividade colaborativa e como os traços de personalidade afetam o desempenho do grupo como um todo (FERREIRA; FERNANDES, 2017). O processo de aprendizagem no RBC, atualizando os casos da base e ajustando os casos disponíveis para propor uma solução, pode auxiliar a formação grupos com boas combinações de traços de personalidade.

Sobre implicação para a Educação, (KOLODNER, 1997) menciona o RBC como um modelo que pode fornecer, por exemplo, sugestões sobre como um aluno será capaz de ter experiências de aprendizado mais ricas, dando a chance de aplicar o aprendizado. Essa declaração reforça a ideia de usar o RBC para explorar dados relacionados a casos de grupos anteriores, a fim de extrair informações que ajudem os alunos a aproveitar a colaboração.

Embora o RBC tenha sido aplicado com sucesso em muitas outras áreas, há apenas algumas aplicações do RBC em contextos educacionais e no agrupamento para colaboração (COCEA; MAGOULAS, 2012). Com o objetivo de explorar essa lacuna, no trabalho de (COCEA; MAGOULAS, 2012), o comportamento do aluno é modelado pelo RBC, que é uma fonte de informação para alimentar uma abordagem de agrupamento de alunos. Da mesma forma, nesta proposta, o RBC modela e atualiza informações de alunos e grupos.

2.2.2 Detecção de Características do Aluno no Texto

Apesar da personalidade ser reconhecida como importante na construção de sistemas de recomendação personalizados, o esforço de capturar tal informação explicitamente é indesejado e pode atrapalhar o uso de recomendações baseadas na personalidade em situações reais (WU; CHEN, 2015). Vários trabalhos foram desenvolvidos com o objetivo de enfrentar o desafio de extrair informações sobre os alunos em AVAs.

(WEN; YANG; ROSE, 2014), por exemplo, aplicou a análise linguística às postagens dos fóruns de alunos para encontrar a relação entre o uso da língua e a motivação do aluno, bem como o engajamento cognitivo no contexto dos cursos *online* massivos abertos. (KUMAR; ROSE, 2014) incorporou PLN em um agente conversacional, a fim de ajudar os alunos em uma atividade colaborativa e apoiar conversas produtivas.

Como a linguagem é usada para expressar pensamentos e emoções, carrega parte das características do interlocutor e, portanto, além das informações sobre motivação dos aprendizes é possível extrair de textos escritos os traços de personalidade (TAUSCZIK; PENNEBAKER, 2010). Em (WATSON; CLARK, 1997) notou-se que, apesar de um extrovertido também demonstrar emoções negativas, as emoções positivas são mais significativamente correlacionadas com Extroversão, pois geralmente refletem entusiasmo, energia e otimismo. Emoções negativas são correlacionadas com Neuroticismo, caracterizado por sentimentos como raiva, tristeza e ansiedade.

Tais características estão relacionadas às categorias do LIWC, como emoções positivas e negativas. Além das categorias de palavras para emoção positiva e negativa, existem categorias relacionadas ao processo social e preocupações pessoais, por exemplo. Em relação às preocupações pessoais, tem-se as subcategorias trabalho e realização que são relacionadas à Conscienciosidade. Alguns estudos foram realizados com o objetivo de encontrar a relação entre o texto escrito e os traços de personalidade do *Big Five*.

Usando o LIWC, (MACHADO et al., 2015) apresentam um estudo inicial sobre o uso de texto em Português para extrair informações sobre a personalidade do autor do texto, de acordo com características psicológicas descritas pelo *Big Five* e das facetas do NEO-IPIP 120. O objetivo do estudo foi criar um léxico que viabilizasse a identificação em textos escritos em Português da polaridade das palavras, do sentimento associado e das características de personalidade. Para avaliar o conjunto de palavras que apontam algum traço de personalidade, os autores extraíram textos de *chats* de alunos de um curso de pós-graduação.

Em (KWANTES et al., 2016) é apresentado um estudo para avaliar a possibilidade de detectar a personalidade em texto escrito através da análise do conteúdo semântico das palavras. Os textos foram adquiridos solicitando aos participantes para descreverem as reações ou sentimentos em cinco cenários, que eram capazes de estimular a criação de narrativas relevantes para cada traço de personalidade do *Big Five*. Usando o modelo conhecido como *Latent Semantic Analysis* (LSA) para analisar o texto, compararam os

escores obtidos nos cinco traços utilizando um questionário com os resultados do LSA. Os resultados obtidos apontaram evidências dos traços de personalidade nos textos analisados.

(PAIM; CAMATI; ENEMBRECK, 2016) desenvolveram um método de inferência da personalidade por meio de textos em Português adquiridos na rede social *Facebook*, usando o LIWC para extrair características linguísticas do texto. O objetivo do estudo foi contribuir para a detecção de traços de personalidade em Português do Brasil. Para validar a proposta, os indivíduos envolvidos nos experimentos responderam ao inventário de personalidade NEO-IPIP 120. Como resultado, apontaram a correlação moderada entre os traços de personalidade e os textos extraídos da rede social. Como (MACHADO et al., 2015), também indicaram possibilidade de contribuição para melhorar a qualidade dos métodos de inferência da personalidade por meio de textos escritos em Português do Brasil.

Em relação aos algoritmos usados na construção de modelos de detecção automática de características de personalidade, o mais comum é o uso de algoritmos indutores de regressão. Com esse objetivo, (WU; CHEN, 2015), através das métricas de comportamento capturados em uma plataforma de consumo de vídeos, usou os modelos *Gaussian Process*, *Pace Regression* e *M5Rules*, indicando o *Gaussian Process* como melhor algoritmo naquele contexto.

(PAIM; CAMATI; ENEMBRECK, 2016) construíram um modelo de predição capaz de reconhecer os traços de personalidade, usando as características psicométricas extraídas com o LIWC. Experimentos com os modelos regressão linear, SMO, M5P e LWL foram realizados, tendo os resultados apontado SMO e LWL como os melhores, considerando o uso do LIWC e de métodos de seleção de atributos. Os resultados encontrados em (FERREIRA; FERNANDES, 2017), também usando as categorias do LIWC, indicaram os algoritmos M5P, SMO e LWL como os mais eficazes para predizer o escore nos traços de personalidade.

Diferente dos trabalhos apresentados anteriormente, que tentaram prever o escore em cada um dos traços, os experimentos de (BUIAR; PIMENTEL; OLIVEIRA, 2017) foram no sentido de classificar as amostras de textos nas classes que representavam os dois pólos de cada traço. Exceto para o traço de Abertura, os valores de acurácia obtidos ficaram entre 52% e 60%. Apesar da taxa de acerto maior do que aquela esperada da escolha aleatória, são valores baixos. Por outro lado, abrem espaço para melhoria do processo de classificação.

Formação de Grupos com o Modelo *Big Five*

Esta proposta consiste no uso dos traços de personalidade dos alunos, deduzidos a partir de textos extraídos da sua interação com ambientes virtuais de aprendizagem (AVAs), para agrupá-los de forma a favorecer a aprendizagem colaborativa. Para viabilizar sua implementação, dois módulos principais foram projetados: Classificação de Traços e Agrupamento. Ambos integrados de forma a permitir atualização e reúso das informações sobre estudantes e grupos na formação de novos grupos para colaboração.

As seções deste capítulo estão divididas de forma a apresentar os componentes da proposta e seu funcionamento, de acordo com a estratégia adotada de coleta de dados e uso de IA para dar suporte ao desenvolvimento dos dois principais módulos. Na Seção 3.1, é dada uma visão geral do fluxo de dados e a integração entre esses módulos. As Seções 3.2 e 3.3 descrevem, respectivamente, cada um dos dois módulos.

3.1 Funcionamento e Fluxo de dados

Neste trabalho, considera-se a identificação dos traços de personalidade a partir de textos escritos por estudantes. A viabilidade de obtenção desses textos pode ser constatada no fato de que a maioria dos AVAs possuem ferramentas de comunicação, tais como fóruns e chats, utilizadas em atividades propostas aos estudantes. No Moodle, por exemplo, o conteúdo das postagens nos fóruns e mensagens trocadas entre os usuários, bem como o autor de cada texto, são gravadas em um banco de dados relacional. A partir de uma consulta nesse banco, é possível obter amostras de textos escritos pelos alunos.

Considerando as informações que podem ser obtidas a partir do banco de dados de um AVA, o fluxo de identificação de grupos, conforme exibido na Figura 2, inicia-se com a extração e armazenamento de dados em um banco de dados local para posterior uso pelo módulo de classificação de traços e pelo módulo de agrupamento. No primeiro módulo, palavras dos textos são utilizadas para a classificação dos traços, sendo esse resultado

armazenado em um banco de dados. A partir da classificação, o segundo módulo, através de Raciocínio Baseado em Casos, sugere formações de grupos.

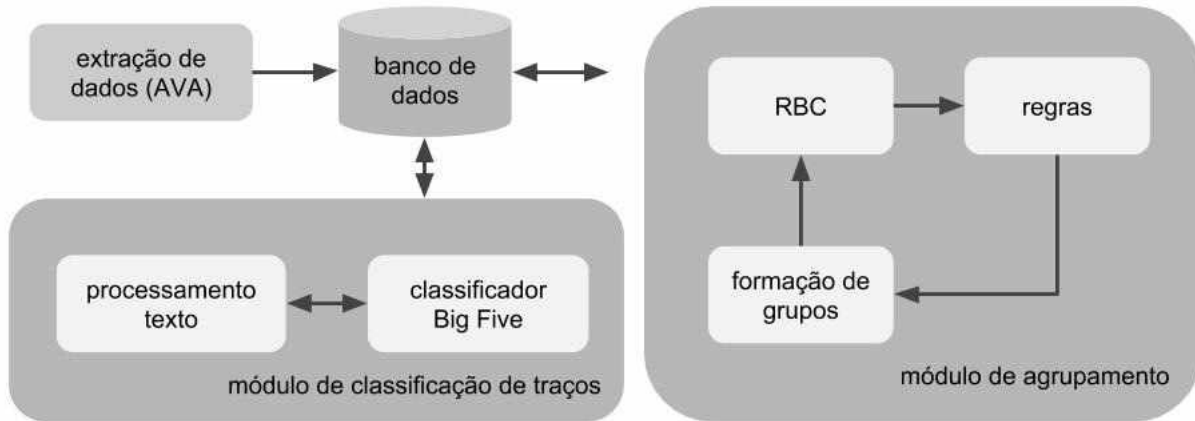


Figura 2 – Funcionamento e Fluxo de dados

Além dos textos, o banco de dados também armazena outras informações sobre os grupos, tais como desempenho, fornecido pelo professor e/ou capturado das plataformas usadas para realizar alguma atividade colaborativa. Os traços de personalidade dos membros do grupo e a nota do grupo em uma atividade colaborativa são inseridos na base de casos do RBC. Os casos são usados na extração de regras para serem aplicadas na formação de novos grupos.

Assim que um novo grupo formado faz uma atividade colaborativa e sua nota se torna disponível, tais informações são armazenadas no banco de dados para realimentar o módulo de agrupamento. Como o fluxo apresentado funciona de maneira dinâmica e cíclica, as interações entre alunos capturadas por meio das ferramentas de comunicação do AVA, devem gerar mais trocas de mensagens. Consequentemente, mais textos serão coletados, permitindo ajustes na classificação de cada um dos traços de personalidade, bem como as revisão e atualização da base usada para sugestão de grupos.

3.2 Módulo de Classificação de Traços

A identificação dos traços de personalidade dos estudantes a partir dos textos coletados é realizada em duas etapas. A primeira é o tratamento do texto. Os textos capturados precisam de tratamento para extração de métricas que indiquem características do texto e que podem ser usadas pelos algoritmos de aprendizagem de máquina na segunda etapa, a detecção dos traços de personalidade. Assim, algumas partes de texto, como acentuação e pontuação, são excluídas.

O texto resultante é convertido em uma lista de palavras. Esse processamento é feito usando a biblioteca NLTK do *Python* (BIRD; KLEIN; LOPER, 2009), desenvolvida para processamento de linguagem natural. Cada palavra da lista é associada a categorias de palavras que indicam alguma característica do texto. As categorias usadas são aquelas do dicionário interno do *Linguistic Inquiry and Word Count* (LIWC) (TAUSCZIK; PENNEBAKER, 2010) que são relacionadas a processos psicológicos e cognitivos, como raiva, tristeza e *insight*, cujo uso frequente em um texto escrito poderiam ser associadas aos traços de personalidade do *Big Five* (NETO; FERREIRA; FERNANDES, 2017).

A versão do dicionário escolhida para a categorização de palavras foi o *Brazilian Portuguese LIWC 2007 Dictionary* de (FILHO; ALUÍSIO; PARDO, 2013). O dicionário também tem categorias relacionadas a elementos de gramática, como advérbios, pronomes, verbos e etc. Tais categorias foram desconsideradas por não serem relevantes para a classificação dos traços. Entretanto, isso não implica em desconsiderar também as palavras, pois uma mesma palavra pode ser classificada em mais de uma categoria do dicionário. O verbo gritar, por exemplo, está na categoria verbo e também na categoria raiva, sendo contabilizado em ambas. As categorias de palavras consideradas neste trabalho são apresentadas na Tabela 1.

Após associar todas as palavras extraídas do texto às categorias do dicionário, é feita a contagem de palavras que aparecem em cada categoria. A partir dessa contagem, a frequência de cada categoria presente no texto é calculada. Assim, após o processamento do texto, obtém-se uma lista de categorias e respectivas frequências, que constituem a entrada para o algoritmo de aprendizagem de máquina.

Todas as categorias de palavras utilizadas neste trabalho podem estar presentes no discurso de qualquer pessoa, independente de suas características. A categoria raiva, por exemplo, tende a aparecer com alguma frequência quando a amostra de texto se torna maior. Por outro lado, dados os comportamentos previstos de cada membro de um grupo, de acordo com os escores de cada traço, espera-se que a frequência de cada categoria seja mais expressiva para todo o grupo.

Palavras que expressam raiva, por exemplo, podem ser mais frequentes no discurso de um indivíduo com alto grau de Extroversão, pois tende a ser mais expressivo que aqueles com baixo grau de Extroversão. Ainda considerando categorias que expressam raiva, nota-se que o número de palavras agressivas tende a ser menor para indivíduos com maior escore para o traço de Afabilidade, já que este tende a ser amável.

Assim, maior frequência em uma certa categoria poderia ser um fator para determinar o traço de personalidade. A frequência é um valor normalizado no intervalo $[0, 1]$, onde 0 indica ausência de palavras da categoria e 1 significa que todas as palavras do texto foram classificadas naquela categoria. Para cada um dos traços, a saída será a classificação em faixas de escore alto, médio ou baixo. A escolha do algoritmo para classificação dos traços é descrita no Capítulo 5. A classificação nos cinco traços do *Big Five* pode então ser usada

Tabela 1 – Categorias de palavras do dicionário utilizadas.

Grupo	Categoria	Exemplo	Descrição
Processos psicológicos	<i>social</i>	absorto, acasala, induzido	Processo social
	<i>friend</i>	vizinho, aliado, amigo	Amigos
	<i>family</i>	avô, conjugal, filhos	Família
	<i>humans</i>	abandono, governo	Humanos
	<i>affect</i>	abafado, comoção, errado	Processos afetivos
	<i>negemo</i>	desconfortável, ensurdecido	Emoção negativa
	<i>anx</i>	aflito, desejo, abalado	Ansiedade
	<i>anger</i>	aborrecer, agressivo	Raiva
	<i>sad</i>	solenidade, acabado, triste	Tristeza
Processos cognitivos	<i>posemo</i>	libertação, aberto, feliz	Emoção positiva
	<i>cogmech</i>	abrandar, confiável, ensinei	Processo cognitivo
	<i>cause</i>	contaminado, aparente	Causação
	<i>insight</i>	assumido, consciente	Compreensão intuitiva
	<i>certain</i>	afirmativo, atual, reveja	Firme convicção
	<i>discrep</i>	deveria, confundida, errado	Discrepância, desacordo
	<i>tentat</i>	talvez, provável, acho	Não é certo; experimental
	<i>inhib</i>	abrandar, parar, bloquear	Inibição
	<i>incl</i>	incluindo, e, com, mais	Inclusão, abrangência
Processos perceptivos	<i>excl</i>	exceto, mas, excluído	Exclusão, restrição
	<i>percept</i>	observar, ouvir, sentir	Processo percepção
	<i>hear</i>	grito, silêncio, falar	Ouvir
	<i>feel</i>	sentir, tocar, aguçado	Sentir
Processos biológicos	<i>see</i>	alvo, belo, avistar, olhar	Ver
	<i>bio</i>	abdômen, pulsar, flor	Processo biológico
	<i>body</i>	mão, junta, excreção	Corpo
	<i>ingest</i>	água, comer, prato, pizza	Ingestão
	<i>health</i>	berçário, câncer, comprimido	Saúde
Relatividade	<i>sexual</i>	beijo, lascivo, conchinha	Sexual
	<i>relativ</i>	área, abertura, saída, fim	Relacionado; dependente
	<i>space</i>	dentro, embaixo, alargador	Espaço
	<i>time</i>	abrupto, tempo, até, fim	Tempo
Preocupações pessoais	<i>motion</i>	chegar, carro, ir, acenar	Movimento
	<i>achieve</i>	ganhar, adiantar, receber	Realização, meta
	<i>home</i>	acamado, família, almofada	Casa, lar
	<i>money</i>	abastado, auditor, moeda	Dinheiro
	<i>leisure</i>	academia, atlético, filme	Lazer
	<i>death</i>	assassino, cadáver, autópsia	Morte
	<i>relig</i>	alcorão, alma, deus, igreja	Religião
Fala/Discurso	<i>work</i>	advogado, trabalho, cópia	Trabalho
	<i>filler</i>	blá, seila, tipo, sei	Enchimento
	<i>nonfl</i>	aiai, ué, ah, ata	Nonfluências
Processo linguístico	<i>assent</i>	aham, kkkk, legal	Concordância, aprovação
	<i>number</i>	segundo, cem, infinito	Números
	<i>swear</i>	porra, merda, foda	Palavrões

para formação de grupos.

3.3 Módulo de Agrupamento

Conforme a Figura 2, o módulo de formação de grupos é constituído por um componente de Raciocínio Baseado em Casos (RBC) e outro de extração de regras de agrupamento (regras) que são empregados na formação de novos grupos. Esse módulo se utiliza de informações sobre grupos de alunos que trabalharam anteriormente em atividades colaborativas. Tais informações são armazenadas na base de dados associando os traços de personalidade de cada aluno do grupo às métricas de desempenho do grupo resultantes da sua avaliação. De posse das métricas de grupo e perfil dos alunos da base de dados, tal informação é processada no RBC e passa a fazer parte da sua BC.

O Raciocínio Baseado em Casos auxilia na identificação da qualidade dos grupos sendo formados, podendo, por exemplo, identificar grupos que precisam ser desfeitos para evitar um desempenho ruim. A operação cíclica do RBC contribui para a evolução da BC, devido à adição e atualização de casos. Portanto, espera-se que a solução encontrada por meio de consulta na BC seja mais precisa à medida que o ciclo de RBC é executado.

Se por um lado a atualização da base e inclusão de novos casos faz com que ela se torne mais abrangente e precisa, por outro, a base de casos vai se tornando também cada vez maior. Esse crescimento pode dificultar a busca por casos apropriados para formar novos grupos. Por esse motivo, uma vez em posse da base de casos atualizada, inicia-se dentro do módulo de agrupamento a extração de regras para a formação de novos grupos.

A mineração das regras na base de casos é feita com o algoritmo *Predictive Apriori* (SCHEFFER, 2001). O resultado da mineração de regras na base de dados é um conjunto de informações mais reduzido porém contendo as informações mais relevantes relacionadas à formação de grupos. Dessa forma, de posse da lista de regras extraídas pelo *Apriori*, torna-se mais simples também a identificação de grupos que têm maior chance de sucesso ou fracasso de acordo com suas características. A partir da lista de regras os grupos são formados no componente de formação de grupos.

Após agrupar todos os alunos de uma turma usando as regras que formam bons grupos, as regras que apontam para desempenho ruim podem ser usadas para avaliar se há necessidade de desfazer ou modificar algum grupo. Por fim, os grupos formados voltam para a base de dados e podem, após serem avaliados, voltar para a BC no RBC. As estratégias empregadas na implementação do RBC dentro da proposta são apresentadas nas Subseções 3.3.1, 3.3.2 e 3.3.3.

3.3.1 Estrutura de um Caso

Um caso é formado por dois objetos, um representa o grupo e o outro, as métricas de grupo. O grupo é representado por seus membros, então cada membro de um grupo

é representado por um vetor de características individuais relativas aos cinco traços de personalidade: Abertura (O), Conscienciosidade (C), Extroversão (E), Afabilidade (A) e Neuroticismo (N). As métricas de grupo são compostas pelo tamanho do grupo, o prazo para entrega da atividade e o desempenho. O prazo é representado por uma string que indica o tempo máximo para entrega da atividade proposta. Foram considerados os valores C e W. O primeiro indica que a tarefa deveria ser concluída durante a aula e o segundo significa que foi dado um prazo de uma semana para conclusão da tarefa. A estrutura do caso na BC pode ser vista na Figura 3.

O desempenho é uma lista de atributos que permitem classificar os grupos como bons ou ruins. Para tal classificação, consideram-se as seguintes métricas:

- ❑ EC: Indica se todos os membros do grupo participam da solução e contribuem, de forma significativa, para a solução da atividade.
- ❑ TC: Indica se o grupo conclui e entrega a atividade.
- ❑ G: A nota do grupo atribuída pelo professor.
- ❑ I: É uma medida do nível de comunicação entre os membros do grupo. Assim, indica se os membros do grupo interagem entre si utilizando alguma ferramenta de troca de mensagens, em AVAs, ou face-a-face, em aula presencial e avaliado pelo professor. Quanto menor a interação do grupo, maior a chance de haver algum membro que não troca informação com os demais.

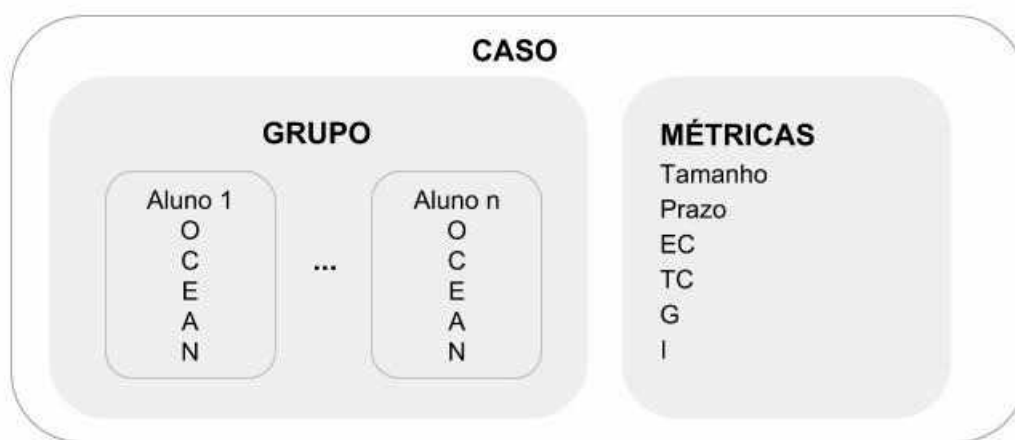


Figura 3 – Estrutura do caso.

3.3.2 Cálculo da Similaridade entre Caso e Consulta

Como apresentado anteriormente, um caso neste trabalho é composto por mais de um objeto e seus atributos, e não apenas por uma lista de atributos, como se poderia esperar em uma aplicação convencional de RBC. Mais ainda, a quantidade de membros do grupo e a lista de objetos do tipo aluno podem variar conforme o tamanho definido para os grupos.

Considerando tais particularidades, o processo de cálculo da similaridade entre caso e consulta foi desenhado com alguns passos extras, a fim de considerar os dois objetos que compõem o caso. Assim, calcula-se a Similaridade entre as Métricas de desempenho dos grupos (SM) e a Similaridade entre os Grupos (SG). Entretanto, para se obter SG , primeiro, é necessário selecionar qual indivíduo da consulta deve ser comparado com qual indivíduo do caso. A seleção desses pares de indivíduos é descrita a seguir, dadas as seguintes variáveis:

- $AL(O, C, E, A, N)$: membro de um grupo representado por uma lista de traços de personalidade
- $G_c = (AL_1, \dots, AL_n)$ a lista dos n membros do grupo, onde c indica caso na base
- $G_q = (AL_1, \dots, AL_m)$, a lista de m membros do grupo, onde q indica a consulta
- MAX_{SIM} a lista de similaridades entre os pares de alunos que retornam o maior valor de similaridade
- $MSO(O_i, O_j)$: similaridade entre dois objetos.

Algoritmo: Seleção de pares de indivíduos

1. Calcule $MSO(AL_i, AL_j)$, para cada AL_i em G_c e AL_j em G_q
2. Atribua a MAX_{SIM} o maior valor de $MSO(AL_i, AL_j)$ encontrado
3. Exclua todos os valores de MSO que envolvam AL_i e AL_j
4. Se ainda resta algum valor de MSO não usado ou excluído, retorne ao passo 2

Como os tamanhos dos grupos G_c e G_q podem ser diferentes, pode ocorrer de algum membro não compor par para calcular a similaridade. Nesse caso, foram descartados no passo 3. Após a execução dos quatro passos descritos acima, o valor final de SG é, então, computado calculando a média dos valores armazenados em MAX_{SIM} .

Um exemplo da execução desses passos é mostrado na Figura 4, onde ambos os grupos têm tamanho dois. Após calcular as similaridades entre todos os pares de alunos, seleciona os maiores valores de similaridade, resultando em $MAX_{SIM} = \{0.9, 0.8\}$ e, por fim,

	aluno 1 (caso)	aluno 2 (caso)
aluno 1 (consulta)	0.6	0.8
aluno 2 (consulta)	0.9	0.8

(a)

	aluno 1 (caso)	aluno 2 (caso)
aluno 1 (consulta)	0.6	0.8
aluno 2 (consulta)	0.9	0.8

(b)

	aluno 1 (caso)	aluno 2 (caso)
aluno 1 (consulta)	0.6	0.8
aluno 2 (consulta)	0.9	0.8

(c)

Figura 4 – Cálculo da similaridade entre os grupos de caso e uma consulta. (a) Calcula a similaridade entre todos os pares. (b) Seleciona a maior similaridade e exclui as linhas com indivíduos envolvidos no cálculo da similaridade selecionada. (c) Seleciona o próximo maior disponível.

retorna $SG = (0.9 + 0.8)/2 = 0.85$. Para esse exemplo, a similaridade entre os grupos é de 85%.

As métricas de desempenho de um grupo podem ser representadas por um vetor de atributos numéricos, dado por $MDG(EC, TC, G, I)$, permitindo usar diretamente a MSO para calcular o valor de SM , a similaridade entre as métricas de desempenho do caso e da consulta. Assim, o valor de SM é dado por $MSO(MDG_c, MDG_q)$.

Se existirem valores para as métricas de grupo na consulta, com os valores de SG e SM , a similaridade entre o caso e a consulta, S , é dada por $S = (SG + SM)/2$. Caso contrário, se as métricas não forem fornecidas na construção da consulta, a similaridade entre o caso e consulta será dado por $S = SG$. A implementação do RBC neste trabalho permite, ainda, a troca das funções empregadas no cálculo da similaridade entre dois objetos, conforme aquelas descritas no Capítulo 2.

3.3.3 Etapas do RBC

De forma resumida, o processo de Raciocínio Baseado em Casos pode ser visto como um ciclo envolvendo as tarefas de recuperação, reuso, revisão e retenção dos casos. Assim, na tarefa de recuperação ocorre a busca por casos similares; no reuso, as soluções encontradas são adaptadas para os novos casos; na revisão, a solução encontrada é avaliada quanto à sua eficácia; e, na etapa de retenção, toma-se a decisão de incluir ou não um novo caso na base de casos já existentes.

Nas abordagens envolvendo RBC, em geral, há a definição de quais elementos do caso são a descrição do problema e quais elementos são a solução para o problema descrito. No entanto, como esta proposta foi desenvolvida buscando usar o conhecimento descrito nos casos tanto para formar grupos quanto para avaliá-los como bons ou não, o que se define como descrição do problema e solução, vai depender do cenário para o qual se busca uma solução.

Quando as consultas são feitas na base com o objetivo de formar grupos, a solução é a definição das características dos indivíduos que devem compor os grupos e a descrição do problema, as métricas de grupo. Ao avaliar um grupo já formado, para o qual as características dos membros são conhecidas, estas compõem a descrição do problema e as métricas de grupos relacionadas ao seu desempenho compõem a solução. A forma como o RBC funciona é tratada nesta subseção.

3.3.3.1 Recuperação

Para recuperação de casos devem ser fornecidos dados dos elementos da estrutura de casos e, opcionalmente, limites mínimos para cada métrica de grupo. Tais limites funcionam como filtros para as consultas na BC. Por exemplo, se $G = 0,6$ e $I = 0,8$ forem dados como limites na consulta, apenas os casos em que $G \geq 0,6$ e $I \geq 0,8$ serão recuperados.

3.3.3.2 Reuso

Uma consulta recupera todos os casos similares observando os limites fornecidos na consulta. A similaridade (S) entre cada um dos casos recuperados e a consulta é calculada. Os casos mais similares são selecionados e aplicados na solução do problema. Por exemplo, se a consulta foi construída fornecendo somente os traços de personalidade dos estudantes, sem qualquer informação sobre o desempenho do grupo, os casos similares recuperados serão aqueles em que a formação do grupo do caso mais se aproxima à formação do grupo da consulta.

Os casos mais similares retornados podem ser aplicados para estimar a possível performance do grupo da consulta. Essa abordagem permite prever se um grupo tem mais

possibilidade de funcionar bem ou de falhar, antes mesmo que os estudantes sejam, de fato, agrupados por meio da avaliação das suas características fornecidas na consulta.

3.3.3.3 Revisão

A revisão da solução consiste no cálculo do desempenho do grupo usando as métricas *EC*: todos contribuem para a solução, *G*: nota do grupo, *I*: nível de interação, usando ferramentas de comunicação ou face-a-face e *TC*: conclusão da atividade proposta. Um grupo é classificado como BOM ou RUIM de acordo com o desempenho do grupo (DP) dado por $DP = 0.4 * EC + 0.1 * G + 0.3 * I + 0.2 * TC$, a média ponderada das métricas de grupo. Assim, se $DP > 0.5$, o grupo é dito BOM, e RUIM, caso contrário.

Os pesos mais altos foram aplicados a *EC* e *I* porque são as métricas que indicam o quanto os membros do grupo estão interagindo para resolver a tarefa proposta. Essas variáveis também podem apontar se a colaboração está ou não acontecendo, mesmo que o grupo receba uma nota alta, já que elas indicam, por exemplo, quando há algum aluno isolado e que não está trabalhando com o grupo. *TC* e *G* também são métricas que ajudam a medir a qualidade do grupo. Portanto, além das métricas de interação, para ser um bom grupo, elas precisam alcançar uma nota satisfatória e concluir a tarefa proposta.

3.3.3.4 Retenção

O funcionamento comum do RBC geralmente corta soluções que não são classificadas como boas soluções na etapa de revisão. Nesta proposta, soluções ruins são tão úteis quanto as boas. As soluções boas ajudam a formar novos grupos bons e as soluções ruins ajudam a identificar composições de grupo que podem não ter sucesso. Por esse motivo, grupos bons e ruins são inseridos na BC na tarefa de retenção.

A maneira como um caso revisado e inserido no BC depende da combinação de alunos que formam o grupo do caso. Se não houver outro caso com essa combinação de características de alunos na BC, o novo grupo e suas métricas de desempenho serão inseridos na base como um novo caso. Se for uma combinação que já é representada por um caso da BC, com o mesmo tamanho e os mesmos perfis dos membros que compõem o grupo, as métricas de desempenho do caso revisado serão usadas para atualizar os valores das métricas de caso da base.

Outros fatores além das características dos indivíduos podem influenciar o desempenho de um grupo. Assim, mesmo uma combinação de indivíduos que, na maioria das vezes produz bons resultados, pode não funcionar bem. Ao combinar as métricas de grupo resultantes de todas as vezes que uma dada combinação de indivíduos trabalhou junto, o valor das métricas de grupo do caso começará a se aproximar do desempenho mais comum ou esperado para aquele tipo de grupo.

Experimento 1: Detecção de Traços

Os traços de personalidade podem ser identificados através de questionário, conforme descrito no Capítulo 2. Entretanto, um dos objetivos deste trabalho foi utilizar textos escritos para essa identificação, a fim de viabilizar a concepção de uma ferramenta automática para detecção de traços e que pudesse ser integrada em AVAs. Por outro lado, para criar o modelo de aprendizagem, foi necessário utilizar o questionário para obter um conjunto de treinamento para aprendizagem supervisionada. Como o questionário continha apenas questões fechadas, foram introduzidas cinco questões abertas, a fim de obter texto escrito dos que responderam o questionário. Dessa forma, o questionário consistiu de duas partes.

Alunos de diferentes níveis de ensino responderam ao questionário. Vale ressaltar que a realização dos experimentos foi aprovada pelo Comitê de Ética da UFU. O número do Certificado de Apresentação para Apreciação Ética (CAAE) atribuído ao projeto foi 69342717.3.0000.5152. As seções seguintes descrevem a obtenção do conjunto de treinamento e os resultados obtidos pelos algoritmos de aprendizagem.

4.1 Questionário e Coleta de Textos

A coleta de textos escritos em Português e dados para cálculo dos escores dos participantes em cada um dos cinco traços de personalidade do *Big Five* foi feita por meio de um questionário. A primeira parte do questionário era composta de perguntas do inventário de 44 itens do *Big Five*, que avalia os cinco traços de personalidade de forma eficiente e flexível (JOHN; SRIVASTAVA, 1999). A versão do questionário usada neste trabalho foi traduzida para Português e sua validade foi testada no Brasil por (ANDRADE, 2008). As perguntas do questionário e o traço de personalidade correspondente podem ser vistos na Tabela 2.

As respostas dessa parte do questionário são dadas em escala *Likert* de 5 pontos a fim de calcular o escore em cada traço. Cada uma das perguntas da Tabela 2 deveria ser respondida com uma das seguintes opções: Concordo totalmente, concordo parcialmente,

Tabela 2 – Inventário de 44 itens em Português do Brasil.

Traço	Polo	Pergunta
Extroversão	+	É conversador, comunicativo.
Afabilidade	-	Às vezes é frio e distante.
Afabilidade	-	Tende a ser crítico com os outros.
Conscienciosidade	+	É minucioso, detalhista no trabalho.
Extroversão	+	É assertivo, não teme expressar o que sente.
Conscienciosidade	+	Insiste até concluir a tarefa ou o trabalho.
Neuroticismo	+	É depressivo, triste.
Afabilidade	+	Gosta de cooperar com os outros.
Abertura	+	É original, tem sempre novas idéias.
Neuroticismo	+	É temperamental, muda de humor facilmente.
Abertura	+	É inventivo, criativo.
Extroversão	-	É reservado.
Abertura	+	Valoriza o artístico, o estético.
Neuroticismo	-	É emocionalmente estável, não se altera facilmente.
Afabilidade	+	É prestativo e ajuda os outros.
Extroversão	-	É, às vezes, tímido, inibido.
Conscienciosidade	-	Pode ser um tanto descuidado.
Afabilidade	+	É amável, tem consideração pelos outros.
Conscienciosidade	-	Tende a ser preguiçoso.
Conscienciosidade	+	Faz as coisas com eficiência.
Neuroticismo	-	É relaxado, controla bem o estresse.
Conscienciosidade	-	É facilmente distraído.
Neuroticismo	-	Mantém-se calmo nas situações tensas.
Abertura	-	Prefere trabalho rotineiro.
Abertura	+	É curioso sobre muitas coisas diferentes.
Extroversão	+	É sociável, extrovertido.
Conscienciosidade	+	É geralmente confiável.
Afabilidade	-	É, às vezes, rude (grosseiro) com os outros.
Extroversão	+	É cheio de energia.
Afabilidade	-	Começa discussões, disputas com os outros.
Conscienciosidade	+	É um trabalhador de confiança.
Conscienciosidade	+	Faz planos e os segue a risca.
Abertura	+	Tem uma imaginação fértil.
Neuroticismo	+	Fica tenso com frequência.
Abertura	+	É engenhoso, alguém que gosta de analisar profundamente as coisas.
Neuroticismo	+	Fica nervoso facilmente.
Extroversão	+	Gera muito entusiasmo.
Conscienciosidade	-	Tende a ser desorganizado.
Abertura	+	Gosta de refletir, brincar com as ideias
Afabilidade	+	Tem capacidade de perdoar, perdoa fácil.
Neuroticismo	+	Preocupa-se muito com tudo.
Extroversão	-	Tende a ser quieto, calado.
Abertura	-	Tem poucos interesses artísticos.
Abertura	+	É sofisticado em artes, música ou literatura.

neutro, discordo parcialmente e discordo totalmente. Se a relação entre pergunta e traço é positiva (polo +), a pontuação atribuída será 4 para concordo totalmente, 3 para concordo parcialmente, 2 para neutro, 1 para discordo parcialmente e 0 para discordo totalmente.

Se a relação é negativa (polo -), concordar com a afirmação significa diminuir o valor do escore naquele traço. Assim, a pontuação atribuída é o inverso do polo positivo, ou seja 0 para concordo totalmente, 1 para concordo parcialmente e assim por diante. Usando as pontuações das respostas, o escore em cada traço é dado pela Eq. (9).

$$\text{Escore} = \frac{\sum p_i}{n}, \quad (9)$$

onde p_i é a pontuação da resposta ao i -ésimo item e n é o número de itens relativos ao traço.

A segunda parte do questionário é constituída de cinco cenários nos quais o aluno deveria descrever ou dar exemplos da sua possível resposta ou reação em cada um deles. Os participantes foram orientados a pensar em sua interação com seus amigos ou colegas de turma para responder às perguntas. Dessa forma, foi obtida uma base de textos com linguagem próxima àquela que os alunos utilizam em suas conversas nas ferramentas de comunicação em AVAs. As perguntas foram as seguintes:

1. O que você geralmente fala ao iniciar uma conversa com seus(suas) amigos(as)?
Deixe alguns exemplos do que você escreve para puxar assunto.
2. Seu(sua) amigo(a) marca um compromisso com você e não avisa nem aparece. Descreva como se sente e o que faz nessa situação. Deixe exemplos do que falaria ao conversar com ele(a) depois do que aconteceu.
3. O prazo de entrega do seu trabalho está bem próximo e você ainda tem detalhes para acertar. Um(a) colega te envia uma mensagem pedindo ajuda para concluir o dele(a). Escreva abaixo suas respostas mais comum nessa situação.
4. Imagine que você recebeu a tarefa de entrar em contato com algumas pessoas desconhecidas. Você vai ter que conversar com elas para coletar informações. Descreva como você se sentiria executando tal tarefa.
5. Tente se lembrar de quando fez algum trabalho ou tarefa que envolvia um assunto que você que nunca tinha visto e, por isso, precisou aprender mais para concluir. Escreva o que achou de se envolver nesse tipo de tarefa.

Alunos de pós-graduação, graduação e ensino médio técnico foram convidados a responder ao questionário. Alunos de pós-graduação que decidiram responder ao questionário estavam cursando pós-graduação em Engenharia de *Software* na Faculdade Pitágoras de Uberlândia ou eram alunos dos programas de mestrado e doutorado da Faculdade de Computação na Universidade Federal de Uberlândia (UFU).

Os alunos de graduação estavam matriculados nas disciplinas de Gestão de Processo do Centro Universitário do Triângulo (Unitri) e nas disciplinas de Análise Financeira, Finanças de Curto Prazo e Dados e Informações Empresariais, da Faculdade de Gestão e Negócios da UFU. Os alunos de ensino médio técnico, convidados a participar, estavam matriculados nas disciplinas de Engenharia de *Software* e Lógica de Programação do Instituto Federal Goiano - Câmpus Campos Belos.

A resposta ao questionário foi facultativa. Um total de 236 alunos, com idade entre 15 e 56 anos, responderam ao questionário, sendo 106 do sexo feminino e 130 do sexo masculino. Destes, 35 eram alunos do ensino médio técnico em Informática, 173 alunos em nível de graduação (concluída ou não) e 28 alunos em nível de pós-graduação (concluída ou não).

Para cada aluno, as respostas dadas a essas cinco perguntas foram agrupadas como um único texto para formar a base. Como a orientação para responder essas perguntas foi escrever da mesma forma que fazem no dia a dia, os textos capturados foram curtos. Os textos fornecidos tinham entre 14 e 570 palavras, com média de aproximadamente 82 palavras por texto.

O escore computado a partir das respostas ao questionário é uma porcentagem que indica a proximidade do indivíduo com o polo positivo de cada traço. Para o traço de Extroversão, por exemplo, quanto mais próximo de 100% for o escore, mais características de uma pessoa extrovertida o respondente deve apresentar. Considerando o mesmo traço, quanto mais próximo de 0%, deve apresentar mais características de uma pessoa introvertida.

Aqueles que atingem escores próximos a 50% e, portanto, tendem a apresentar características de ambos os extremos, não podem ser classificados em um dos polos de um determinado traço. No entanto, não há definições para escores alto (A), médio (M) e baixo (B). Assim as faixas de porcentagem para A, M e B foram definidas com base nas perguntas do questionário. Alunos classificados com escore A, para algum traço, foram aqueles cujas respostas resultaram em uma porcentagem maior que 62,5%. Isso significa que a maioria de suas respostas correspondiam a pontuações altas na escala adotada para o questionário. Na faixa B, foram classificados aqueles com porcentagem menor que 38,5%, ou seja, a maioria das suas respostas resultaram em pontuação baixa no traço. Finalmente, a faixa M corresponde ao intervalo entre os dois valores anteriores, isto é, escores maiores ou iguais a 38,5% e menores ou iguais a 62,2%.

A partir da classificação dos alunos nas três faixas de escore, foi possível observar que, em relação as suas características, a maioria atingiu escores médios ou altos (A ou M) nos traços de Afabilidade, Abertura e Conscienciosidade. Para os traços Neuroticismo e Extroversão, foram encontrados indivíduos nos dois extremos (A e B), contudo a maioria dos indivíduos obteve escore médio (M) nesses traços, como pode ser observado na Tabela 3.

Tabela 3 – Número de alunos em cada uma das faixas de escore.

Escore	A	M	B
Extroversão	73	123	40
Afabilidade	111	118	7
Abertura	116	114	6
Conscienciosidade	121	106	9
Neuroticismo	54	103	79

4.2 Predição dos Traços de Personalidade

Com os dados coletados do experimento anterior, formou-se a base de treinamento para os algoritmos de aprendizagem de máquina. Os textos coletados com o questionário passaram pelo tratamento para então determinar as frequências de cada categoria de palavras. Com essas frequências, os experimentos de classificação dos traços foram conduzidos, utilizando a ferramenta Weka (FRANK; HALL; WITTEN, 2016).

Como o questionário devolve o valor em cada traço como porcentagem (um valor entre 0 e 1), os primeiros algoritmos testados foram aqueles capazes de devolver um valor numérico como escore em cada traço. Dentre aqueles os algoritmos capazes de lidar com saída numérica no Weka, foram testados *Locally Weighted Learning* (LWL), *Sequential Minimal Optimization* (SMO), Regressão Linear (RL), *Radial Basis Function* (RBF), *Multilayer Perceptron* (MLP), M5P, *Fast Decision Tree Learner* (REPTree) e *Gaussian Process* (GP). Como o valor devolvido pelas funções são valores contínuos, para avaliá-los na tarefa de classificar o aluno nas faixas de escore alto, médio ou baixo, os valores de porcentagem devolvidos pelos algoritmos foram convertidos para os valores de A, M e B. A Tabela 4 apresenta os algoritmos (Alg) que mostraram os melhores resultados para cada um dos traços de personalidade (TP): Abertura (O), Conscienciosidade (C), Extroversão (E), Afabilidade (A) e Neuroticismo (N).

As métricas usadas na avaliação dos algoritmos foram a taxa de acertos (Precis), porcentagem dos valores total de amostras da classe que foram classificados corretamente (VR), e porcentagem dos valores previsto em uma classe que realmente pertenciam à classe (VP). Para um dado algoritmo, se o VP_B=0%, significa que o algoritmo não classificou nenhuma amostra como B, e portanto, a porcentagem de valores corretos previstos como B ficou zerada. Se VR_B=0%, o algoritmo pode ter classificado alguma amostra como B, mas nenhuma delas eram de fato B, ou seja, nenhuma amostra B real foi classificada corretamente como B.

Para cada traço, foram selecionados para compor a Tabela 4 os algoritmos com os quais foi possível atingir a maior precisão ou melhores VP e PR. Para Extroversão e Neuroticismo, só há uma entrada, pois os melhores algoritmos considerando todas as métricas foram também os de maior precisão. Assim, de acordo com os resultados, M5P e SMO são melhores para prever os valores de Extroversão e Neuroticismo, respectivamente.

Tabela 4 – Resultados obtidos com os algoritmos que devolvem saídas contínuas.

	Alg	TP	Precis	VR_A	VR_B	VR_M	VP_A	VP_M	VP_B
1	LWL	O	57,20%	94,74%	0,00%	9,28%	57,53%	52,94%	0,00%
2	RBF	O	51,27%	63,91%	16,67%	36,08%	57,82%	42,68%	14,29%
3	GP	C	62,71%	76,92%	0,00%	45,24%	68,75%	50,00%	0,00%
4	MLP	C	56,36%	68,53%	11,11%	40,48%	66,67%	42,50%	11,11%
5	M5P	E	52,54%	22,22%	5,00%	90,43%	60,00%	51,49%	50,00%
6	MLP	A	59,75%	100,00%	0,00%	2,22%	59,40%	100,00%	0,00%
7	M5P	A	57,63%	67,63%	14,29%	45,56%	63,95%	46,59%	100,00%
8	SMO	N	43,22%	18,75%	26,58%	74,19%	41,38%	42,07%	48,84%

Para os demais traços, os algoritmos de maior precisão (Entradas 1, 3 e 6 da Tabela 4) não foram capazes de se ajustarem para prever valores de porcentagens na faixa de escore B. Nesse caso, não foram considerados boas opções. Ainda que com precisão um pouco menor, os algoritmos 2, 4 e 7 foram melhores. Assim, para classificar traços a partir de textos escritos, as melhores opções, de acordo com os experimentos, foram RBF para Abertura, MLP para Conscienciosidade e M5P para Afabilidade.

Os valores das métricas apresentadas são resultado da média obtida das execuções usando o método de divisão da amostra *leave one out*. Com esse método, um exemplo é separado para teste e os demais para treinamento, e o processo se repete até que todos os exemplos da base tenham sido usados como teste. É um método de custo alto computacional, considerando que houve uma execução para cada aluno na base. Por outro lado, como a base era pequena (236 alunos) não foi tão custoso e gerou valores de taxa de erro e, por consequência de precisão, menos tendenciosos.

Os experimentos com os algoritmos de classificação também foram conduzidos usando a mesma estratégia de divisão da base. Para esse segundo experimento foram selecionados os algoritmos, da ferramenta Weka, capazes de lidar com entradas numéricas (as frequências nas categorias de palavras) e devolver uma das três classes de escore para cada traço de personalidade.

As funções do Weka que implementam algoritmos de classificação (Alg) usados foram: as redes neurais *RBFClassifier* (RBFCl), *RBFNetwork* (RBFNet) e MLP, as árvores de decisão REPTree, J48 e *Random Forest* (RF), e também os algoritmos SMO, LWL, *Naive Bayes* (NB) e Regressão Logística. As métricas de avaliação dos algoritmos são as mesmas do experimento anterior. Os resultados são apresentados na Tabela 5.

Para o traço de Extroversão (E), o algoritmo *Random Forest* obteve maior precisão e foi também melhor classificado, considerando as demais métricas. Para os demais traços, os algoritmos que atingiram maior precisão não foram capazes de se ajustarem para prever as três classes possíveis. Para os algoritmos das entradas 1, 3 e 6 da Tabela 5, os valores obtidos para VR_B e VP_B foram iguais a zero. Já o algoritmo da entrada 8 não conseguiu prever a classe A. Nesses casos, de acordo com os resultados, os melhores algoritmos

Tabela 5 – Resultados obtidos com os algoritmos de classificação para 3 classes de escores.

	Alg	TP	Precis	VR_A	VR_B	VR_M	VP_A	VP_M	VP_B
1	LWL	O	53,81%	75,19%	0,00%	27,84%	57,14%	44,26%	0,00%
2	MLP	O	45,34%	56,39%	16,67%	31,96%	52,08%	34,07%	100,00%
3	REPTree	C	59,75%	91,61%	0,00%	11,90%	61,50%	43,48%	0,00%
4	RBFNet	C	55,08%	71,33%	11,11%	32,14%	63,35%	38,03%	25,00%
5	RF	E	49,58%	33,33%	15,00%	73,04%	42,19%	54,55%	33,33%
6	RBFC1	A	59,75%	78,42%	0,00%	35,56%	63,37%	50,00%	0,00%
7	NB	A	54,66%	65,47%	14,29%	41,11%	61,90%	47,44%	9,09%
8	LWL	N	44,07%	0,00%	17,72%	96,77%	0,00%	42,25%	66,67%
9	J48	N	39,41%	20,31%	44,30%	48,39%	25,00%	42,86%	44,30%

seriam MLP para Abertura, RBFNet para Conscienciosidade, NB para Afabilidade e J48 para Neuroticismo.

Em ambos os experimentos, todos os algoritmos foram testados com a configuração padrão do Weka (sem alteração dos parâmetros) e variando os parâmetros disponíveis na ferramenta. Por exemplo, para MLP, o experimento foi repetido variando os valores da taxa de aprendizagem e quantidade de camadas escondidas da rede. Para o SMO e LWL, que permitem trocar os algoritmos usados internamente no cálculo dos pesos, por exemplo, foram executados trocando tais funções também.

Para a maioria dos traços, o uso de um algoritmo que classificava cada aluno em uma das faixas de escore, independente dos parâmetros escolhidos, gerou resultados piores do que algoritmos que retornavam a porcentagem no traço. Comparando com o primeiro experimento, os resultados para Extroversão e Neuroticismo foram ainda piores, considerando o valor da precisão atingida pelo melhor algoritmo selecionado. Para o uso proposto neste trabalho, os algoritmos capazes de retornarem a porcentagem como um valor contínuo parecem ser melhores opções, se comparados àqueles que retornaram as classes A, M e B.

4.3 Discussão dos Resultados

A partir das respostas ao questionário, observou-se que os alunos participantes tinham perfil médio ou alto para três dos traços de personalidade do modelo *Big Five*. O fato de não haver muitos indivíduos com escore baixo para Abertura e Conscienciosidade pode ser explicado pelo grau de formação acadêmica dos participantes. Dificilmente, alguém que não tenha interesse em obter novos conhecimentos, procuraria por mais formação.

Além disso, é natural que pessoas com mais tendência ao planejamento, característica do traço de Conscienciosidade, procurem por cursos das áreas que envolvem gestão de negócios e tecnologia da informação. Essas são áreas nas quais o profissional geralmente precisa, além da capacidade de organização e planejamento, também de interesse por novidades (característica do traço de Abertura).

O fato da amostra de participantes não ser balanceada, dificultou o ajuste dos algoritmos de aprendizagem para prever tais traços. Como não haviam muitos indivíduos com escore baixo para Afabilidade, Conscienciosidade e Abertura, se nenhum aluno for classificado como baixo, a taxa de acerto será alta já que a chance de errar será menor. O mesmo ocorreu com o traço de Neuroticismo, para o qual havia mais exemplos de alunos com escore médio nos dados coletados. Assim, se nenhum aluno for classificado com escore alto, a taxa de acerto também pode ser alta, ainda que o algoritmo não tenha se ajustado para prever escore alto.

Devido a essas características da base, a escolha do algoritmo mais adequado para prever cada traço levou em consideração, além a acurácia, outras métricas que ajudaram a avaliar a qualidade do algoritmo também foram consideradas. Dessa forma, em alguns casos, o algoritmo de maior acurácia foi descartado em favor de outro que se ajustou melhor aos dados, considerando todas as métricas avaliadas.

Para os cinco traços de personalidade, ao considerar as três faixas de porcentagem, A, M e B, nenhum resultado apresentando alta taxa de acerto foi encontrado. No entanto, para a maioria dos traços, os melhores algoritmos selecionados obtiveram acurácia maior que 50%. A única exceção foi o traço de Neuroticismo cuja maior acurácia encontrada foi abaixo de 50%.

Como as métricas extraídas do texto se referem à frequência de cada uma das categorias de palavras do dicionário adotado, há a possibilidade de interferência de um traço no outro. Por exemplo, se uma dada categoria é muito frequente para o traço de Extroversão, se todos os indivíduos extrovertidos da amostra forem também muito conscienciosos, essa categoria será também contada com alta frequência no traço de Conscienciosidade.

Se a amostra for suficientemente abrangente em relação às características dos alunos, espera-se que as frequências das categorias associadas aos traços mais significativos do indivíduo sejam maiores. Além disso, se o volume de texto capturado de cada aluno for suficientemente grande, as palavras mais características do vocabulário do aluno vão dar mais peso para uma ou outra categoria.

Além de capturar mais texto, para melhorar a capacidade dos algoritmos de prever cada um dos traços outra alternativa que deve ser considerado é o acréscimo de outras métricas que poderiam ajudar a identificar o perfil. Como o volume de atributos extraídos do texto já é muito alto, uma avaliação de quais atributos são realmente importante para classificar cada traço poderia melhorar a qualidade da predição. Por exemplo, incluir uma etapa de seleção de atributos antes de treinar os algoritmos.

Algumas dificuldades foram encontradas nos experimentos conduzidos para selecionar um algoritmo de aprendizagem de máquina para prever os traços por meio do texto. No entanto, como observado, essas dificuldades podem ser minimizadas, por exemplo, capturando maior volume de texto, capturando texto de alunos dos perfis menos frequentes na base atual, e aplicando alguma estratégia de seleção de atributos. Portanto, ainda há

alternativas a serem empregadas na melhora da qualidade da previsão permitindo, assim a construção do modelo de classificação automática dos traços de personalidade a partir de texto. Uma vez desenvolvido, tal modelo pode ser integrado a AVAs e disponibilizado para professores e tutores como ferramenta para conhecer melhor os alunos.

Experimento 2: Formação de Grupos

O Capítulo 4 demonstrou a viabilidade de classificar traços de personalidade a partir de textos escritos e um dos objetivos é o uso dessa classificação na formação de grupo. Assim, neste capítulo são apresentados e discutidos os experimentos conduzidos em relação à formação de grupos. Na Seção 5.1 são descritos os experimentos para a formação da base de casos, a avaliação da base para formação de grupos está na Seção 5.2, e o quanto a base pode ser considerada abrangente para formar grupos na Seção 5.3. A Seção 5.4 mostra os experimentos envolvendo o uso de regras de associação na formação e avaliação de grupos. A discussão dos resultados é apresentada na Seção 5.5.

5.1 Construção da Base de Casos

Os dados para a base inicial de casos do RBC foram coletados de grupos de alunos em atividades colaborativas, que frequentavam 4 turmas no segundo semestre de 2017. Apesar da participação na atividade ter sido obrigatória por fazerem parte da programação de cada disciplina, nem todos os alunos responderam ao questionário, já que era facultativo. Por isso, não foi possível determinar os traços de alguns alunos. Os grupos foram avaliados de acordo com as métricas (EC, TC, G e I) definidas no Capítulo 3.

A atividade avaliativa proposta pelo professor e a coleta dos dados foram diferentes entre as turmas, bem como os tamanhos dos grupos, mas contendo todos 5 alunos, no máximo. Em todas as turmas, a escolha dos membros de grupos foi uma decisão dos próprios alunos. Os dados coletados em cada turma são descritos nas tabelas que se seguem.

A primeira turma era de alunos do curso Técnico em Informática para Web do Instituto Federal Goiano - Câmpus Campos Belos. A atividade proposta pelo professor consistia de 6 questões para reforçar e testar os conhecimentos na disciplina de Engenharia de Software. Todas as questões deveriam ser respondidas no horário de aula e sem consulta, trocando informações apenas com os colegas de grupo. Com exceção de um grupo com 2 alunos, todos os demais eram formados por 3 alunos.

As atividades deveriam ser concluídas usando a ferramenta de escrita compartilhada do *Google Drive* (Documentos Google). Como a ferramenta armazena *logs* de alteração dos documentos, o histórico de revisão foi usado para verificar se todos participaram da atividade de forma significativa, seja respondendo às questões ou corrigindo e revisando a resposta do colega. O documento também foi usado para verificar quais grupos concluíram ou não a tarefa.

Não foi possível capturar as conversas dos alunos pois o Documentos Google não armazenam bate-papos que ocorrem quando todos estão visualizando o documento, simultaneamente. No entanto, de acordo com o relato do professor, todos os alunos estavam, preferencialmente, trocando informações por meio do bate-papo enquanto respondiam às questões propostas. Os dados coletados nessa turma são apresentados na Tabela 6.

Tabela 6 – Dados dos grupos da Turma 1.

Grupo	TC	I	G	EC	O	C	E	A	N
1	1	1	0,67	1	M	M	M	M	A
					M	A	M	A	B
					M	A	A	A	A
2	1	1	0,73	1	B	A	A	M	A
					A	A	A	M	A
					M	A	A	M	M
3	0	1	0,37	0	M	A	A	A	M
					M	A	M	M	M
					M	M	M	A	M
4	1	1	0,53	1	A	A	A	A	B
					A	A	A	A	M
					A	A	A	A	M
5	1	1	0,83	1	M	M	A	M	B
					A	A	M	M	B
					M	B	M	M	B
6	1	1	0,50	1	A	A	A	M	A
					M	B	A	M	A
7	1	1	0,60	1	A	A	M	M	B
					A	M	A	A	M
					M	M	M	A	B
8	1	1	0,50	1	M	M	A	M	A
					B	A	M	M	B
					M	M	A	M	A

A Turma 2 era constituída de alunos da disciplina de Análise Financeira da Faculdade de Gestão e Negócios da Universidade Federal de Uberlândia (FAGEN - UFU). Nesta turma, a atividade aplicada pelo professor era composta por 12 questões relacionadas à aplicação prática dos conhecimentos da disciplina, como cálculo de balanço patrimonial. Os grupos formados tinham de 3 a 5 alunos. A única exceção foi um aluno que fez a atividade sozinho.

Como na Turma 1, os alunos tiveram que responder às questões usando a ferramenta Documentos Google. Assim, a avaliação quanto à conclusão da tarefa (TC) e trabalho em conjunto (EC), foi feita da mesma forma que na Turma 1, ou seja, analisando o histórico de revisão dessa ferramenta. Ao contrário do que ocorreu na Turma 1, todos os grupos que entregaram a tarefa receberam nota máxima, caso contrário, tiveram nota zero.

Todos os grupos tiveram prazo de uma semana para concluir a atividade e, portanto, foi realizada em horário extra-sala, não permitindo ao professor observar as interações nos grupos. Assim, a comunicação nos grupos (I) foi observada por meio dos comentários deixados no documento. Os dados coletados na Turma 2 são apresentados na Tabela 7, onde escore com valor '-' significa a ausência de respostas no questionário *Big Five*.

Tabela 7 – Dados dos grupos da Turma 2.

Grupo	TC	I	G	EC	O	C	E	A	N
9	1	0.3	1	0	A	M	M	M	A
					-	-	-	-	-
					A	A	M	A	A
					M	M	M	A	M
10	1	0.3	1	0	M	A	M	M	A
					M	M	M	M	M
					M	A	A	A	B
					M	A	A	M	B
11	1	0.5	1	0	A	M	M	A	M
					A	B	M	A	A
					A	A	M	A	B
					A	A	M	A	A
12	1	0.5	1	1	A	M	M	A	M
					A	B	B	B	M
					A	A	M	A	M
13	0	0.3	0	0	-	-	-	-	-
					-	-	-	-	-
					A	A	A	M	M
					-	-	-	-	-
14	0	0	0	0	M	M	M	A	M

Os alunos da Turma 3 estavam matriculados na disciplina de Dados e Informações Empresariais da FAGEN - UFU. A atividade proposta pelo professor era composta de questões relacionadas ao conteúdo da disciplina e deveriam ser realizadas durante o período de aula e sem consulta. Os alunos deveriam resolver a atividade somente com a ajuda dos alunos do grupo, que era composto de no máximo 5 alunos. Nessa turma, nenhuma ferramenta de escrita compartilhada foi utilizada e todos os valores das métricas foram obtidos de relatório enviado pelo professor.

De acordo com o relatório, todos os grupos cumpriram o prazo definido (TC) e contri-

buíram para a atividade (EC), recebendo notas máximas. Por outro lado, de acordo com as observações do professor e *feedback* dos alunos, em alguns grupos a interação por meio de conversa não foi satisfatória. Alguns membros, por exemplo, não compartilharam as opiniões e se limitaram a resolver a respectiva parte da atividade. Os dados coletados nessa turma são apresentados na Tabela 8.

Tabela 8 – Dados dos grupos da Turma 3.

Grupo	TC	I	G	EC	O	C	E	A	N
15	1	0.8	1	1	A	A	M	A	M
					A	M	M	A	M
					-	-	-	-	-
					A	M	M	M	M
					M	M	A	M	A
16	1	0.8	1	1	M	B	A	A	A
					-	-	-	-	-
					M	B	A	M	A
					-	-	-	-	-
					A	M	A	A	M
17	1	0.8	1	1	M	A	M	M	A
					M	M	A	A	M
					-	-	-	-	-
18	1	0.8	1	1	M	M	A	M	A
					M	A	M	A	M
					M	M	M	M	M
					A	B	M	A	A
					M	M	M	A	B
19	1	1	1	1	A	A	A	A	B
					M	M	M	A	B
					B	A	A	A	B
					M	B	A	A	M
					M	B	A	A	M
20	1	0.6	1	1	M	M	A	A	A
					A	B	M	B	B
					-	-	-	-	-
					A	M	A	M	A
					M	B	A	A	M

Alunos da disciplina de Gestão da Qualidade de Processo do curso de pós graduação em Engenharia de Software da Unitri formaram a Turma 4. A atividade consistia em responder 4 questões teóricas relacionadas a processo de gestão e auditoria no desenvolvimento de software, que foi atribuída aos alunos através do AVA Moodle e cujo prazo de entrega era o horário de aula.

Além de entregar as respostas das atividades, também tiveram que discutir as possíveis soluções nos *chats* e fóruns do Moodle. A avaliação quanto à interação (I), entrega da atividade completa (TC), contribuição de todos os membros (EC) e a nota atribuída pelo

professor (G) foi obtida do banco de dados do Moodle. Os dados coletados na Turma 4 são exibidos na Tabela 9.

Tabela 9 – Dados dos grupos da Turma 4.

Grupo	TC	I	G	EC	O	C	E	A	N
21	1	0.6	0.59	0	A	A	A	A	B
					M	M	B	M	A
					A	A	M	M	M
22	1	0.53	0.64	0	M	A	M	A	B
					A	A	M	A	B
					A	A	A	M	B
23	1	0.67	0.69	0	M	M	M	M	B
					A	A	M	M	A
					M	A	A	A	B
24	1	0.67	0.65	1	M	A	M	A	B
					A	A	M	M	B
					M	A	B	M	A
25	1	0.4	0.73	1	-	-	-	-	-
					-	-	-	-	-

O experimento com as quatro turmas permitiu concluir que tamanho do grupo e prazo de entrega influenciam de maneira significativa a qualidade do grupo, escore M para Neuroticismo negativo para colaboração, escore A para Abertura induz ao engajamento, combinar escores M para Neuroticismo e A para Abertura contribui para finalização da atividade, escores A, B e M para Conscienciosidade combinados conduzem a ótima interação. Estas são apenas algumas conclusões baseadas nas seguintes observações:

- ❑ Em relação às características dos grupos e ao desempenho, a maioria das combinações obteve bons resultados.
- ❑ Em relação à colaboração, o resultado não foi bom quando o prazo para entrega era muito longo.
- ❑ Apenas o grupo de 3 componentes da Turma 2 interagiu e trabalhou junto, quando o prazo era de uma semana.
- ❑ Nos demais grupos, quando o prazo era menor, os grupos com mais de 3 membros apresentaram desempenho satisfatório em relação à colaboração e interação.
- ❑ Em grupos com mais de três membros, prazo longo não implicou em grupo bem sucedido, independente das características dos membros.
- ❑ Quando o prazo foi menor, as características dos indivíduos podem ter influenciado as interações. Os grupos da Turma 1 eram principalmente formados por alunos com escores A e M em todos os traços, exceto Neuroticismo.

- ❑ O único grupo (grupo 3) da Turma 1 em que os alunos não interagiram, os membros apresentaram escore M para Neuroticismo e, além disso, um dos membros não participou da atividade. Isso sugere que ter indivíduos com escore A ou B em Neuroticismo é positivo para a colaboração.
- ❑ Na Turma 2, em que praticamente todos os grupos não tiveram sucesso, o único grupo bem sucedido foi aquele em que todos tinham escore M para Neuroticismo e A para Abertura (grupo 12). Abertura alta pode aumentar o grau de engajamento dos indivíduos, especialmente se a atividade for interessante e conduzir a novos conhecimentos.
- ❑ Na Turma 3, os grupos não eram formados apenas de alunos com escore M para Neuroticismo, que parece ter influência negativa na interação.
- ❑ Os membros do grupo 19 apresentaram escores A, M e B para Conscienciosidade, tendo esse grupo apresentado o maior nível de interação.
- ❑ A falta de colaboração influenciou a nota de todos os grupos. Entretanto, na Turma 4, apesar dos grupos terem atingido nota satisfatória, nem todos os membros participaram da atividade ou interagiram. Assim, não podem ser considerados bons grupos do ponto de vista da colaboração.
- ❑ O grupo 22 foi ruim em relação à colaboração, tendo todos os membros escore B para Neuroticismo.
- ❑ O grupo 5 tem a mesma combinação em relação ao Neuroticismo, além de bom grupo em relação à interação, foi aquele que obteve maior nota na Turma 1. A principal diferença entre os grupos 5 e 22 é o traço de Conscienciosidade. Os grupo 5 e 19 eram semelhantes, com escores A, M e B nesse traço. A distribuição dos indivíduos com escores diferentes para esse traço no grupo faz com que existam alunos mais focados e com maior capacidade de improvisação, produzindo as melhores soluções após as interações.

5.2 Formação de Grupos com RBC

O experimento de formação de grupos usando o RBC foi conduzido na turma de Dados e Informações Empresariais da FAGEN - UFU. Após fazer uma atividade em grupo, onde os alunos puderam escolher os membros dos grupos, o professor propôs outras 3 atividades, mas, dessa vez, usando um critério de agrupamento. Os grupos foram formados a partir de sugestões resultantes de consultas ao RBC proposto neste trabalho. Todas as atividades foram realizadas em sala de aula e sem o uso de um AVA.

Os experimentos considerando os grupos apresentados nesta seção foram todos conduzidos na mesma turma e pelo mesmo professor. A elaboração das atividades aplicadas e a forma como os grupos foram avaliados foram deixadas a critério do professor da turma. A intenção foi interferir de forma menos invasiva possível no planejamento da disciplina e na rotina dos alunos. O professor elaborou o relatório do qual foram coletadas as informações sobre desempenho do grupo. Nenhuma entrevista estruturada ou formulário foram solicitados ao professor.

5.2.1 Sugestão de Agrupamento

O perfil de cada aluno da turma foi usado para montar uma consulta com os seguintes parâmetros: nota maior que 60% da nota total ($G > 0,6$), interação entre os alunos maior que 50% ($I > 0,5$) e todos participaram da solução ($EC = 1$). Os casos recuperados que atendiam às três restrições foram ordenados pelo grau de similaridade com as características do aluno da consulta. Os grupos foram formados inserindo no mesmo grupo aqueles alunos cujas características eram mais similares ao mesmo caso da base.

Os grupos de alunos criados a partir das consultas foram usados como sugestões de agrupamento pelo professor. Na Tabela 10, a coluna “Grupo” identifica os grupos sugeridos. Alguns alunos não constavam da lista entregue pelo professor, assim foram identificados como “Sem Grupo”. Para cada aluno, os escores dos cinco traços do *Big Five* são exibidos nas colunas O, C, E, A e N. A última coluna indica o caso mais similar a cada aluno, usado para agrupá-los no mesmo grupo. Cada caso corresponde a um grupo das Tabelas 6, 7, 8 e 9.

Em relação aos resultados apresentados na Tabela 10, as seguintes observações foram feitas:

- ☐ Nos grupos C, E e H que correspondem aos casos 7, 15 e 19, respectivamente, foi observado que os traços dos membros do grupo e do caso eram bem próximos.
- ☐ Nos demais grupos, A, B, D, F e G, não se verificou a observação do item anterior.
- ☐ Alguns grupos apresentaram os padrões dos bons grupos observados nos experimentos da Seção 5.1.
- ☐ Exceto pelo grupo F, não ocorreu grupo tendo todos os indivíduos com escore M para Neuroticismo.
- ☐ Na maioria dos grupos da Tabela 10 havia mais de um membro com escore A para Abertura, aspecto considerado bom para a colaboração.
- ☐ Grupos onde todos apresentavam escore A para Conscienciosidade e Extroversão também foram evitados, sendo o grupo H o único exemplo dessa situação. Cada um

Tabela 10 – Sugestões de agrupamentos resultantes de consultas na BC.

Grupo	Aluno	O	C	E	A	N	Caso
A	1	A	A	B	M	M	2
	2	M	A	M	M	M	
	3	A	A	M	M	A	
B	4	A	A	M	A	B	5
	5	A	A	B	M	A	
	6	M	M	B	M	B	
	7	M	M	M	M	B	
	8	A	M	M	M	A	
C	9	A	M	M	A	B	7
	10	M	A	B	M	M	
	11	M	M	M	A	B	
	12	M	M	M	A	B	
	13	A	M	M	A	B	
D	14	A	M	M	A	B	12
	15	A	M	M	M	A	
	16	A	B	M	M	M	
	17	A	M	M	A	B	
	18	M	M	M	A	M	
E	19	A	M	M	M	M	15
	20	A	M	M	M	M	
	21	M	M	A	A	A	
	22	A	A	M	A	A	
	23	A	M	M	M	M	
F	24	M	A	A	A	M	17
	25	M	M	B	A	M	
	26	M	M	A	A	M	
	27	M	M	A	A	M	
G	28	M	A	M	A	M	18
	29	M	A	M	A	M	
	30	M	M	M	M	B	
	31	M	M	M	M	M	
	32	M	M	B	M	M	
H	33	B	A	A	A	M	19
	34	A	A	A	A	B	
	35	M	A	A	A	B	
	36	A	A	A	A	B	
	37	A	A	A	A	B	
Sem Grupo	38	M	M	A	A	M	
	39	M	M	B	M	M	
	40	M	A	M	M	M	

desses traços em alto grau isolado pode conduzir a situações indesejadas. Alta Extroversão conduz à distração com conversas e alta Conscienciosidade ao isolamento. Contudo, a ocorrência de ambos simultaneamente pode reduzir os efeitos negativos das duas características, isoladamente.

5.2.2 Grupos Formados para Atividade Colaborativa

Alguns alunos faltaram ou chegaram muito atrasados para as aulas nas quais o professor aplicou as atividades colaborativas. Por essa razão, alguns grupos sofreram alteração em relação às sugestões. No dia da primeira atividade, o professor decidiu por apenas remover os faltantes dos grupos sem fazer trocas entre os grupos. Assim, os grupos D, G e H ficaram com menos membros que os grupos sugeridos na Tabela 10, mas com combinação ainda semelhante ao sugerido. A formação adotada pelo professor na primeira atividade e o desempenho observado quanto à interação ($I \in [0, 5]$) pode ser visto na Tabela 11.

Na primeira atividade, todos os grupos mostraram boa interação ($I=5$). A única exceção foi o grupo G que perdeu o aluno 32, o único que tinha escore B para Extroversão. Na segunda atividade, os grupos B, D, E e H foram mantidos os mesmos. Alunos que não participaram da primeira atividade precisaram ser inseridos em um grupo, para a realização da segunda atividade. Também ocorreu de alunos presentes na primeira atividade não compareceram na segunda, por exemplo, aluno 3 de A e três alunos do grupo F.

Como o grupo A ficaria com apenas dois membros, os alunos 25 (único membro do antigo grupo F) e 38 (aluno novo) foram inseridos em A. O grupo F recebeu outro aluno novo (39) e um aluno removido do grupo G, que teve desempenho ruim na atividade 1. Para que o grupo F tivesse pelo menos 3 integrantes, o aluno 10 do grupo C foi emprestado para o grupo F na atividade 2. A composição e a interação dos grupos são apresentadas na Tabela 12.

Para a terceira atividade os grupos da atividade 2 foram mantidos, exceto C e F. O aluno 10 retornou ao grupo C e o F ficou com os alunos que estavam presentes na aula e não pertenciam a nenhum grupo. Os alunos e a interação de cada grupo (I) estão na Tabela 13.

De acordo com o relatório sobre as atividades colaborativas enviado pelo professor:

- ❑ Todos os grupos foram capazes de terminar as atividades propostas.
- ❑ Na primeira atividade, os grupos formados conforme sugestão do RBC foram mais eficientes e melhores em relação à interação quando comparados aos grupos formados livremente.
- ❑ Apesar de questionamentos sobre as formações de grupo designadas pelo professor, a percepção dos alunos quanto ao próprio desempenho foi positiva considerando o novo grupo.

Tabela 11 – Grupos da primeira atividade.

Grupo	Aluno	O	C	E	A	N	I
A	1	A	A	B	M	M	5
	2	M	A	M	M	M	
	3	A	A	M	M	A	
B	4	A	A	M	A	B	5
	5	A	A	B	M	A	
	6	M	M	B	M	B	
	7	M	M	M	M	B	
	8	A	M	M	M	A	
C	9	A	M	M	A	B	5
	10	M	A	B	M	M	
	11	M	M	M	A	B	
	12	M	M	M	A	B	
	13	A	M	M	A	B	
D	14	A	M	M	A	B	5
	15	A	M	M	M	A	
	16	A	B	M	M	M	
E	19	A	M	M	M	M	5
	20	A	M	M	M	M	
	21	M	M	A	A	A	
	22	A	A	M	A	A	
	23	A	M	M	M	M	
F	24	M	A	A	A	M	5
	25	M	M	B	A	M	
	26	M	M	A	A	M	
	27	M	M	A	A	M	
G	28	M	A	M	A	M	1
	29	M	A	M	A	M	
	30	M	M	M	M	B	
	31	M	M	M	M	M	
H	33	B	A	A	A	M	5
	34	A	A	A	A	B	
	35	M	A	A	A	B	
	36	A	A	A	A	B	

- ❑ Grupos que trabalharam juntos em atividade anterior, em geral, continuaram a mostrar interação satisfatória nas atividades seguintes.
- ❑ Os grupos B, D, E e H permaneceram inalterados em relação às características dos membros, nas três atividades, e apresentaram combinações positivas. Por exemplo, grupos de alunos com escores extremos (A ou B) ou grupos de alunos com escores A, M e B combinados para o traço de Neuroticismo.
- ❑ O grupo G, na primeira atividade teve interação ruim, era principalmente formado por alunos com escores M na maioria dos traços. A remoção de um membro redu-

Tabela 12 – Grupos da segunda atividade.

Grupo	Aluno	O	C	E	A	N	I
A	1	A	A	B	M	M	5
	2	M	A	M	M	M	
	25	M	M	B	A	M	
	38	M	M	A	A	M	
B	4	A	A	M	A	B	5
	5	A	A	B	M	A	
	6	M	M	B	M	B	
	7	M	M	M	M	B	
	8	A	M	M	M	A	
C	9	A	M	M	A	B	5
	11	M	M	M	A	B	
	12	M	M	M	A	B	
	13	A	M	M	A	B	
D	14	A	M	M	A	B	5
	15	A	M	M	M	A	
	16	A	B	M	M	M	
E	19	A	M	M	M	M	5
	20	A	M	M	M	M	
	21	M	M	A	A	A	
	22	A	A	M	A	A	5
	23	A	M	M	M	M	
F	31	M	M	M	M	M	4
	39	M	M	B	M	M	
	10	M	A	B	M	M	
G	28	M	A	M	A	M	5
	29	M	A	M	A	M	
	30	M	M	M	M	B	
H	33	B	A	A	A	M	5
	34	A	A	A	A	B	
	35	M	A	A	A	B	
	36	A	A	A	A	B	

ziu o escore M para Conscienciosidade, Afabilidade e Neuroticismo, melhorando a interação.

Mesmo após a alteração para ajustar aos alunos presentes no dia em que as atividades 1, 2 e 3 foram aplicadas, os demais grupos também sofreram alterações que, contudo, não influenciaram de forma negativa a interação e nem resultaram em grupos ruins. As alterações feitas pelo professor mantinham alguma semelhança com grupos das atividades anteriores e com os casos da base. No entanto, apresentavam também diferenças em alguns traços e, portanto, puderam também ser usados para popular a base de casos como novos exemplos de bons grupos.

Tabela 13 – Grupos da terceira atividade.

Grupo	Aluno	O	C	E	A	N	I
A	1	A	A	B	M	M	5
	2	M	A	M	M	M	
	25	M	M	B	A	M	
	38	M	M	A	A	M	
B	4	A	A	M	A	B	5
	5	A	A	B	M	A	
	6	M	M	B	M	B	
	7	M	M	M	M	B	
	8	A	M	M	M	A	
C	9	A	M	M	A	B	4
	10	M	A	B	M	M	
	11	M	M	M	A	B	
	12	M	M	M	A	B	
	13	A	M	M	A	B	
D	14	A	M	M	A	B	5
	15	A	M	M	M	A	
	16	A	B	M	M	M	
E	19	A	M	M	M	M	5
	20	A	M	M	M	M	
	21	M	M	A	A	A	
	22	A	A	M	A	A	
F	23	A	M	M	M	M	4
	40	M	A	M	M	M	
	24	M	A	A	A	M	
	26	M	M	A	A	M	
G	28	M	A	M	A	M	4
	29	M	A	M	A	M	
	30	M	M	M	M	B	
H	33	B	A	A	A	M	5
	34	A	A	A	A	B	
	35	M	A	A	A	B	
	36	A	A	A	A	B	

5.3 Abrangência da Base de Casos

Neste trabalho, uma consulta à base de casos pode se destinar a resolver dois tipos de problemas: (a) determinar o provável desempenho de grupo conhecido ou (b) identificar grupos mais adequados para inserção de novo membro. Em (a), a consulta consiste de um grupo, se houver alta similaridade entre esse grupo e algum caso da base e, além disso, o desempenho do grupo do caso for bom, pode-se concluir um provável bom desempenho do grupo da consulta. Em (b), a consulta contém os traços de um indivíduo, mas há a necessidade de filtros como por exemplo, nota e iteração, a fim de encontrar bons grupos para inserção desse novo indivíduo.

Contudo, ressalta-se que nas duas situações acima, as soluções serão cada vez melhores, à medida que a base for abrangente e atualizada com novos casos. Assim, os experimentos descritos nesta seção têm o objetivo de verificar a capacidade da base de responder aos dois tipos de problema e a expansão dessa capacidade à medida que os resultados dos experimentos de grupo foram inseridos na base.

5.3.1 Casos x Casos

Com o objetivo de verificar se os casos da base são similares entre si e se a inserção de novos dados aumenta a diversidade na base, cada grupo dos casos da base foi convertido em consulta. As similaridades obtidas como respostas às consultas foram usadas para contabilizar a proporção de casos da base com 100%, 90%, 80% e 70% de similaridade com a própria base.

A primeira avaliação foi feita com a inserção dos dados extraídos da Turma 1 (Tabela 6). A base inicial de casos foi composta de 8 casos e, então, 8 consultas foram realizadas. Como pode ser observado na Figura 5, todos os grupos da primeira turma são muito similares. Todos têm similaridade de 80% com pelo menos um dos demais casos da base e 50% dos casos têm similaridade acima de 90% com algum caso da base.

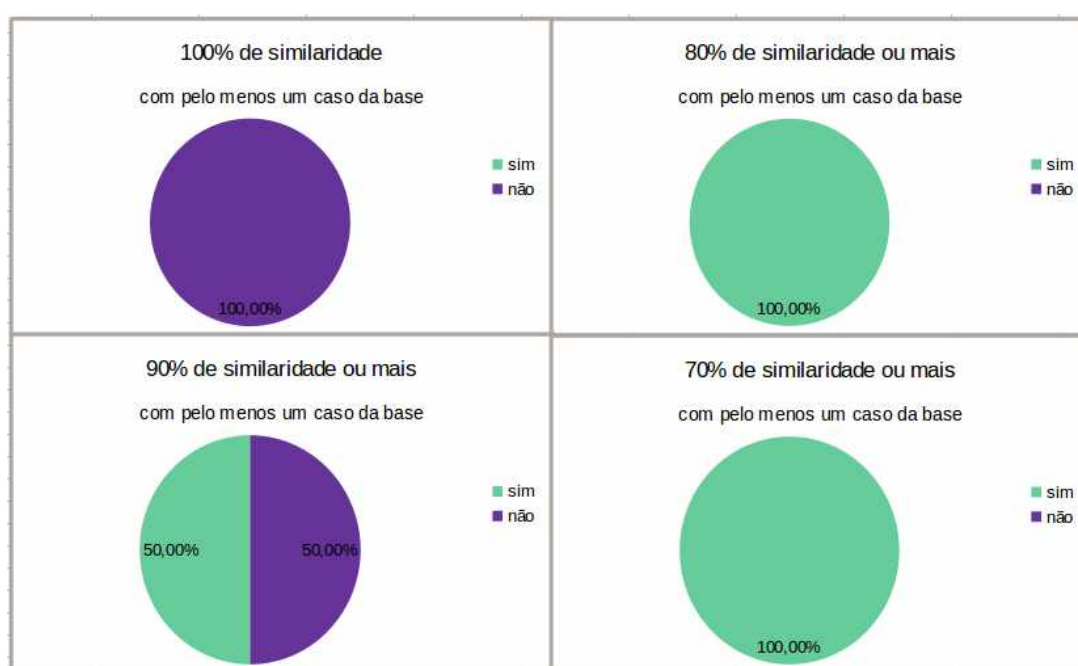


Figura 5 – Similariadade entre os casos da Turma 1.

A segunda avaliação ocorreu com a inserção dos dados da Turma 2 (vide Tabela 7) na base de casos, totalizando 14 casos e 14 consultas. Os resultados deste experimento mostraram que cerca de 7% dos casos têm similaridade menor que 80%, como pode ser visto na Figura 6. A redução no número de casos muito similares indicam que grupos com combinações diferentes dos 8 casos anteriores foram adicionados à base.

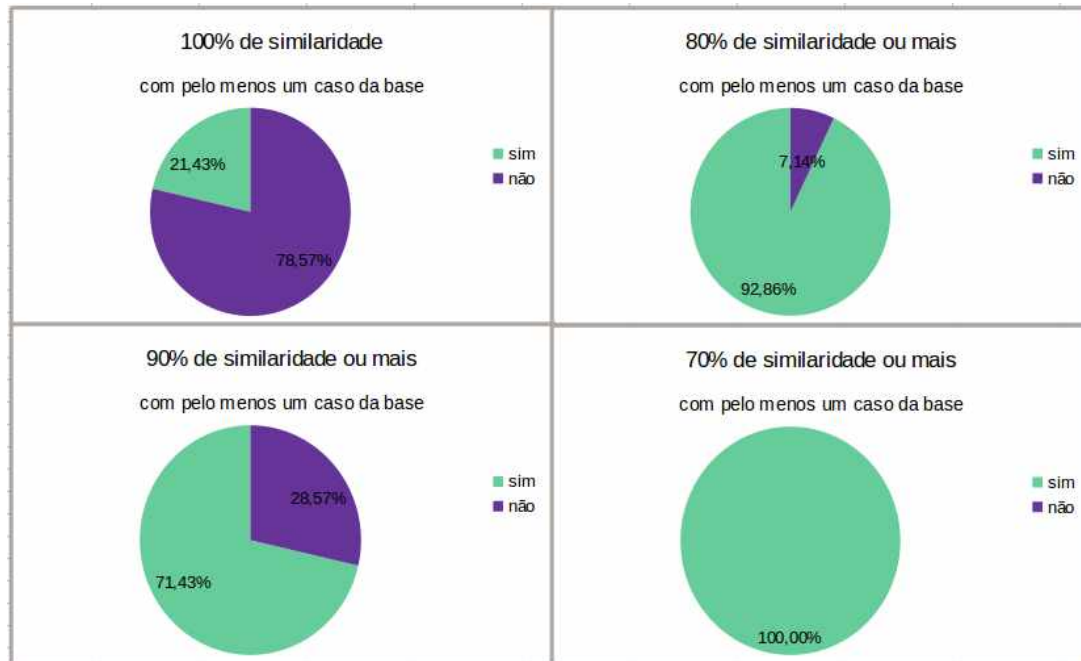


Figura 6 – Similaridade entre casos das Turmas 1 e 2.

A inserção dos dados das Turma 3 (Tabela 8) e Turma 4 (Tabela 9), aumentou a quantidade de casos com similaridade maior que 90%, conforme mostra a Figura 7. Por outro lado, também foram inseridos novos casos com menos similaridade: 4% dos casos agora apresentam menos que 70% de similaridade com os demais casos da base.

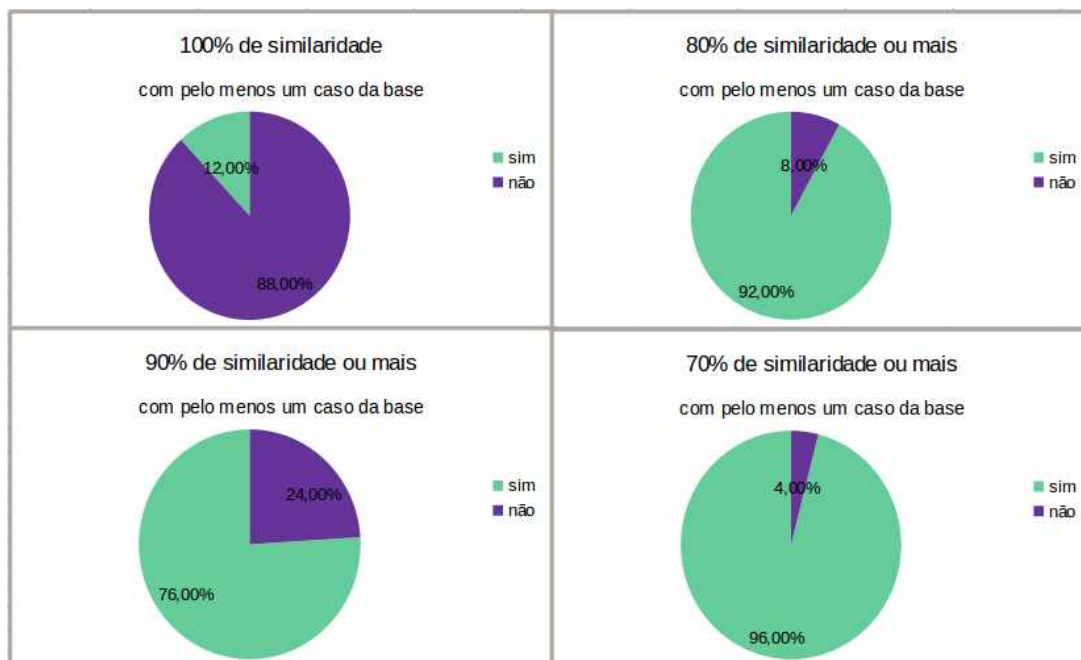


Figura 7 – Similaridade entre casos das Turmas 1, 2, 3 e 4.

A última atualização da base de casos foi com a inserção dos dados resultantes dos experimentos para formação de grupos a partir do RBC. No total foram 13 combinações

distintas inseridas na base, a partir daqueles grupos, totalizando 38 casos e 38 consultas. Esses novos casos, apesar de acrescentarem mais diferenças de características entre os indivíduos e diferentes combinações, também aumentaram a quantidade de casos muito similares, já que na Figura 8 observa-se uma redução de 24% para cerca de 21% de casos com menos que 90% de similaridade com os demais.

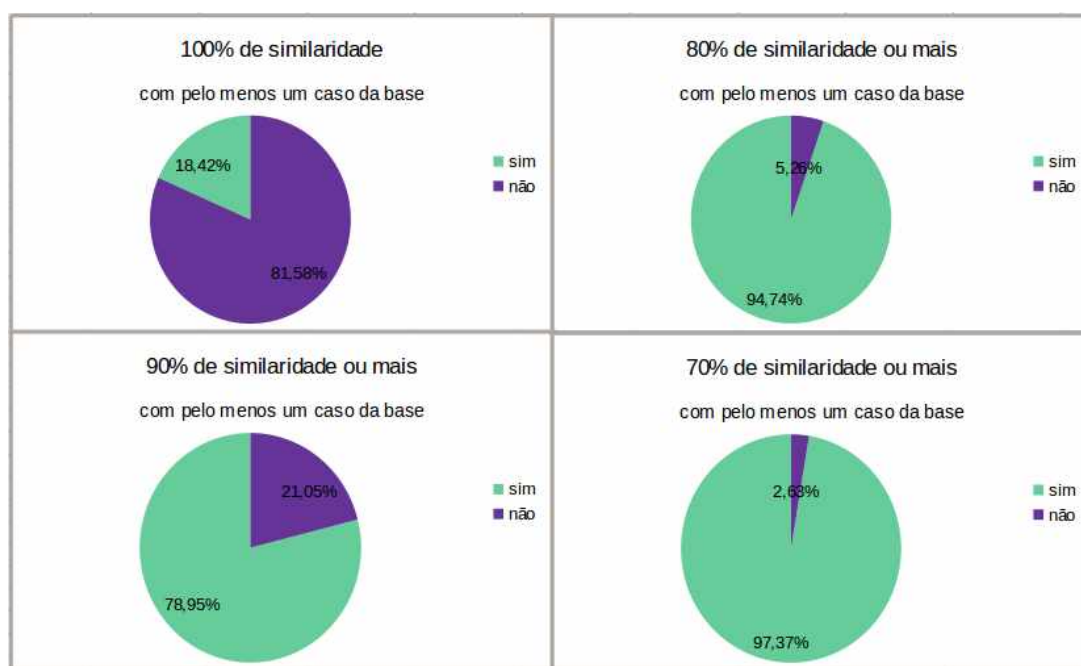


Figura 8 – Similaridade após inserção dos grupos sugeridos pelo RBC

Após a inserção de casos com os grupos da Turma 2, similaridade de 100% entre os casos da base surgiram. Por exemplo, na Figura 6, há cerca de 21% de casos com 100% de similaridade. Isso ocorre pois o tamanho do grupo não foi inserido no momento da consulta como variável para o cálculo da similaridade. Assim, se as combinações de traços de um grupo de 3 alunos, por exemplo, estiverem presentes em um grupo de 5 alunos, os grupos serão considerados 100% similares.

Os cerca de 22% de casos que apareceram na Figura 6, são os grupos 3 (Turma 1), 9 e 14 (Turma 2). O único membro de 14 tinha perfil idêntico a algum membro de 3 e 9. No entanto, os três foram inseridos na base como casos distintos por terem tamanhos distintos. O grupo 14 poderia ser considerado um candidato à remoção da base, pois tem apenas um membro, cujas características aparecem em dois outros grupos.

Por mais que haja repetição de alguns perfis em cada grupo, o tamanho do grupo pode ser significativo, conforme visto nos experimentos de formação de grupo. O grupo G (Tabela 11) teve interação ruim na Atividade 1 e apresentou boa interação com um indivíduo a menos na Atividade 2 (Tabela 12). Esses dois grupos foram inseridos nas porcentagens da Figura 8 como tendo 100% de similaridade quanto à composição dos grupos, mas têm quantidade de indivíduos e desempenhos distintos.

Considerando grupos de tamanhos iguais, similaridade de 100% não ocorre pois, se o novo grupo inserido na base tem a mesma combinação e mesmo número de membros, não há inserção de novo caso e, sim atualização das métricas. Por exemplo, o grupo C nas Atividades 1 e 3, tinha a mesma combinação de alunos, mas a interação mudou de 5 para 4 na última atividade. Na base, ambos serão o mesmo caso, porém com a métrica I atualizada para a média da interação, isto é, 4.5.

5.3.2 Consultas com Todos os Perfis de Indivíduo Possíveis

Para que fosse possível usar o RBC na formação de grupos, a base de casos deveria conter pelo menos um caso suficientemente similar para ser usado como solução. Assim, deveria ser possível encontrar uma boa combinação de indivíduos para formar grupo com cada perfil de indivíduo, em uma consulta à base de casos. Como os perfis dos alunos são determinados por cinco traços de personalidade que podem assumir os escores A, M, ou B, há 243 tipos de perfis possíveis.

Nos experimentos de construção da base, observou-se que alguns perfis são menos comuns ou podem não ser encontrados entre os alunos que participaram dos experimentos e, portanto, não foram representados na base. Os perfis ausentes podem ainda não ser similares a casos da base. Logo, é fácil pensar que para determinados perfis, pode não ser possível encontrar um caso que possa ser usado.

Para verificar se isso ocorre e o quanto ocorre na base construída com os experimentos de grupo, um experimento considerando consultas compostas por cada um dos perfis possíveis foi conduzido. Cada consulta era formada por um dos 243 perfis e as seguintes restrições ou filtros: número de membros maior ou igual a três (tamanho ≥ 3), interação maior ou igual a 60% ($I \geq 0,6$) e participação de todos os membros na atividade ($EC=1$).

Seguindo o mesmo procedimento do experimento anterior, as consultas foram feitas com a base inicial de 8 casos com os grupos da Turma 1 e, depois, inserindo os casos dos grupos da Turma 2 e assim por diante. Na Figura 9, para cada uma das bases, com 8, 14, 20, 25 e 38 casos, a porcentagem de perfis para o qual foi possível encontrar um caso contendo um indivíduo com o mesmo perfil da consulta (sim) e a porcentagem de perfis usados como consulta para os quais não há indivíduo com o mesmo perfil na base de casos (não).

À medida que os novos casos foram inseridos na base, novos grupos com novos perfis de indivíduos foram sendo inseridos na base. De 6% dos 243 perfis na base com 8 casos, passou a conter 13% dos 243 perfis na base com 38 casos. No entanto, há ainda 87% dos perfis que não estão presentes em nenhum dos casos da base que poderiam formar bons grupos. Por outro lado, mesmo sem correspondência exata, se encontrado um caso onde os perfis dos indivíduos fossem similares, seria possível usá-lo como solução.

Na Figura 10 são apresentadas as porcentagens dos 243 perfis para os quais é possível encontrar um caso com pelo menos 70% de similaridade (sim) e aqueles para os quais

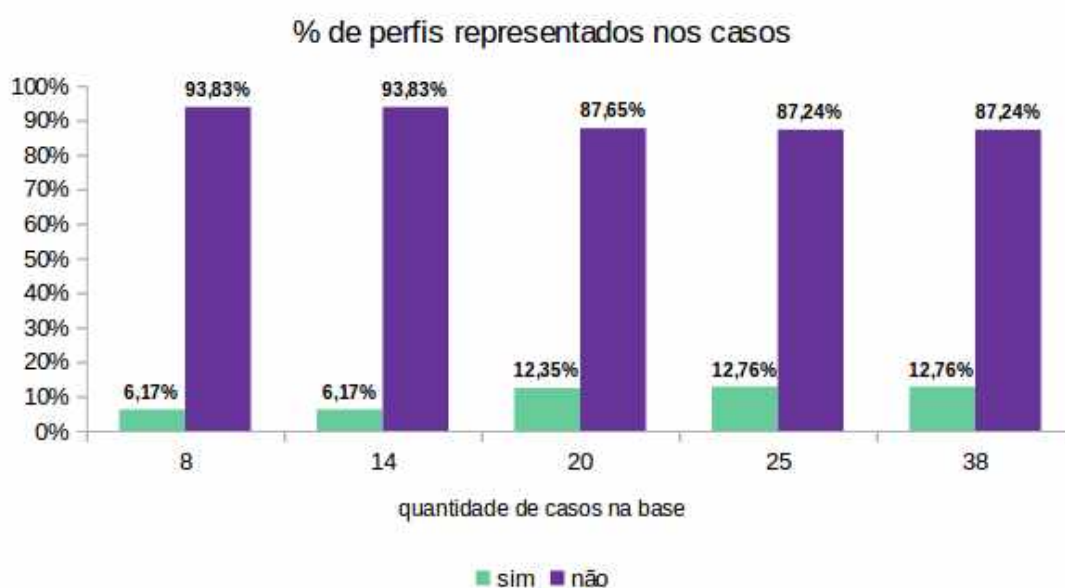


Figura 9 – Perfis possíveis presentes em algum caso da base.

não foi possível (não). Com a inserção dos novos casos, a quantidade de perfis para os quais não havia caso com mais de 70% de similaridade foi sendo reduzida. Na base com 38 casos, para todos os 243 possíveis perfis de aluno, é possível encontrar pelo menos um caso. Assim, mesmo que o perfil não exista na base, ainda é possível usar as respostas das consultas.

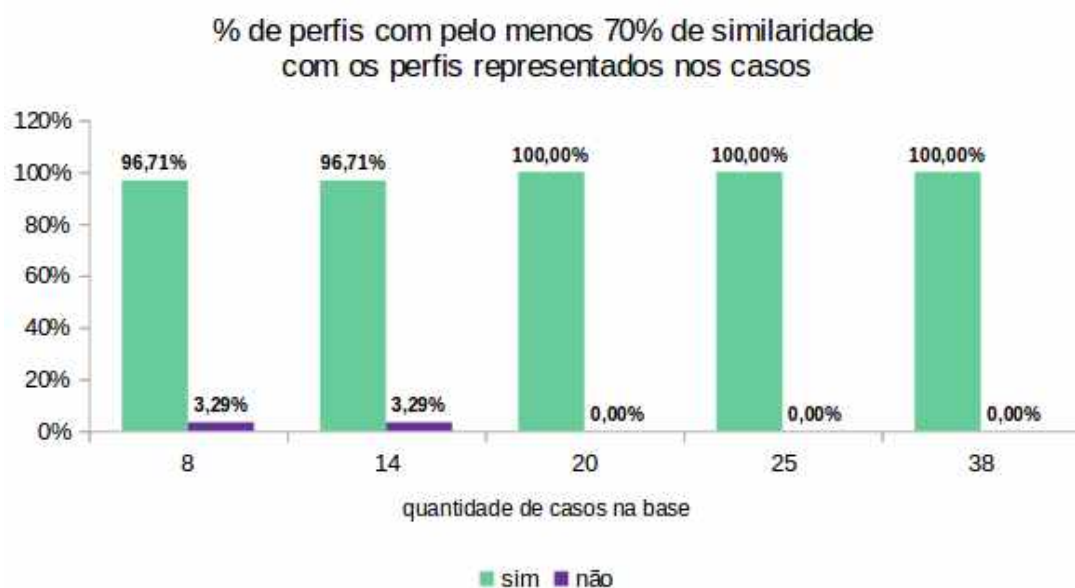


Figura 10 – Porcentagem de perfis com pelo menos 70% de similaridade com algum caso da base.

5.3.3 Efeitos da Mudança de Cálculo de Similaridade no Resultado da Consulta

Nos experimentos anteriores, todas as consultas foram feitas utilizando as mesmas funções para o cálculo da similaridade bem como a mesma estratégia para abordar valores desconhecidos. No entanto, dadas as características do RBC proposto neste trabalho, é possível trocar as funções de cálculo de similaridade e, conseqüentemente, causar um efeito no valor da similaridade.

Uma estratégia para tratar valor desconhecido é considerar que consulta e caso são 50% similares. Outra estratégia é aquela que considera 100% de similaridade, se há valores desconhecidos no caso e na consulta e nenhuma similaridade quando apenas um valor é desconhecido, seja na consulta ou no caso.

A estratégia para valor desconhecido empregada nos experimentos anteriores foi considerar 50% de similaridade para atributos sem valor conhecido. Considerar um perfil desconhecido como igual, faria com que encontrássemos vários casos com 100% de similaridade. No entanto, há uma grande chance do perfil desconhecido no caso ser diferente da consulta. Como o objetivo dos experimentos era observar a presença de todos os possíveis perfis na base, era preciso garantir que os casos trazidos com 100% de similaridade realmente eram compostos de alunos com perfil idêntico ao da consulta.

Na base, os escores A, M e B são representados, respectivamente, pelos valores inteiros 3, 2 e 1. Após normalização para o intervalo $[0, 1]$, os escores A, M e B correspondem a 1, 0.5 e 0, respectivamente. Essa representação foi adotada devido ao processo de comparação de indivíduos para calcular a similaridade de grupo. Por exemplo, considere o traço Abertura e a função linear (Eq. 4) para cálculo de similaridade. A comparação entre $O=A$ e $O=M$ resulta em 0.5, isto é, 50% similares. A distância (diferença) entre $O=A$ e $O=B$ é 1, sendo 0% similares. Por fim, quando os escores são iguais, a distância é zero e a similaridade 100%. Assim, quanto maior a distância, menor a similaridade.

Dependendo da medida de similaridade adotada, dentre aquelas disponíveis no RBC implementado, a distância entre opostos e a similaridade entre próximos são ainda mais acentuadas. A Tabela 14 apresenta os efeitos na similaridade ao mudar as funções para cálculo da similaridade entre atributos (FS Atributo) e entre objetos (FS Objeto). Em todos os testes, foram utilizados a atribuição de 50% de similaridade como estratégia de valor desconhecido, a função de diferença linear e o mesmo caso.

A similaridade utilizando a função *Simple Matching* (Eq. 8) não foi alterada ao alternar as demais funções. Independente do método de cálculo, essa função só contabiliza os atributos iguais. Para comparação dos traços não é muito indicada pois, para esta aplicação do RBC, a diferença entre extremos de cada traço é mais significativa do que a diferença entre médio e extremos.

Por outro lado, se houver a inclusão de algum atributo binário ou classe representada por texto, o *Simple Matching* pode ser uma boa opção e, portanto, foi mantido entre as

Tabela 14 – Similaridade entre caso e consulta com a troca das funções.

FS Atributo	FS Objeto	Similaridade%
Função Threshold	Minkowski grau 1	65,00%
	Minkowski grau 2	69,72%
	Minkowski grau 3	71,42%
	Simple Matching	40,00%
Função Linear	Minkowski grau 1	70,00%
	Minkowski grau 2	71,10%
	Minkowski grau 3	71,89%
	Simple Matching	40,00%
Função Exponencial	Minkowski grau 1	71,07%
	Minkowski grau 2	71,73%
	Minkowski grau 3	72,26%
	Simple Matching	40,00%
Função Sigmoid	Minkowski grau 1	70,00%
	Minkowski grau 2	71,10%
	Minkowski grau 3	71,89%
	Simple Matching	40,00%

funções de cálculo de atributo. Para o cálculo de similaridade entre objetos, a função Minkowski (Eq. 7) se adequa melhor, mantendo o peso da distância entre dois atributos.

Aumentar o grau da função de Minkowski, aumenta também a similaridade, conforme Tabela 14, onde os graus 1, 2 e 3 foram testados para cada par de FS Atributo e FS Objeto. Pode não ser muito vantajoso grau maior que 3, pois aumentará o número de casos retornados com alta similaridade que, na verdade, não são tão parecidos. O uso das funções de grau 2 e 3 para comparar membros de um grupo, faz com que a similaridade entre escore médio e escores extremos seja acentuada. O que pode ser interessante, já que indivíduos de escore médio apresentam características dos dois extremos e deveriam, até certo ponto, ser considerados similares.

5.4 Regras de Formação de Grupos

O funcionamento do RBC proposto prevê duas formas de atualização da base. Na primeira, se os membros do novo grupo e de algum caso da base forem iguais, apenas as métricas de grupo do caso são atualizadas. A segunda forma insere um novo grupo na base, se a combinação de alunos que o compõe for diferente em um único traço e tiver sido avaliado em atividade colaborativa.

Como são 243 possíveis perfis de alunos, se todos os possíveis grupos de dois alunos fossem inseridos como novos casos na base, considerando somente os grupos de dois alunos a base já estaria populada com mais de 59 mil casos. Se todos os possíveis de grupos tamanho três fossem inseridos, seriam quase 14.5 milhões de casos na base. Como nos experimentos, tiveram grupos de até cinco membros, a base poderia ser muito grande.

Considerando o possível crescimento da base e a dificuldade de encontrar um caso para responder uma consulta, a alternativa proposta é minerar regras de formação de grupo, a partir dos dados da base. Tais regras poderiam ser usadas tanto para formar novos grupos como para avaliar uma dada combinação de indivíduos. A extração das regras e formação de grupos são apresentadas nesta seção.

5.4.1 Mineração de Regras Base de Casos

A fim de minerar as regras de agrupamento, o algoritmo aplicado à base foi o *Predictive Apriori* (SCHEFFER, 2001), utilizando a classe *Predictive Apriori* do Weka (FRANK; HALL; WITTEN, 2016). Como os casos podem ter grupos com tamanhos de 1 a 5 indivíduos, para fornecer uma entrada com o mesmo número de atributos para o algoritmo, os dados foram pré-processados, gerando apenas cinco colunas que resumem os traços do grupo do caso.

Esse resumo, como o nome sugere, é uma representação condensada de cada um dos cinco traços de personalidade dos membros do grupo. O valor inserido em cada um dos traços foi atribuído seguindo as regras descritas na Tabela 15. A primeira coluna refere-se aos escores apresentados pelos membros do grupo (Escore no Grupo) e a segunda coluna o valor que é atribuído ao traço no resumo (Traço).

Tabela 15 – Valores do resumo dos traços de personalidade.

Escore no Grupo	Traço
Todos ou maioria dos membros obtiveram escore alto no traço	A
Todos ou maioria dos membros obtiveram escore baixo no traço	B
Todos ou maioria dos membros obtiveram escore médio no traço	M
Parte dos alunos têm escore médio e a outra parte, escore baixo no traço	MB
Parte dos alunos têm escore médio e a outra parte, escore alto no traço	MA
Todos os membros têm escores alto ou baixo no traço	AB
Os membros têm escores alto, médio e baixo no traço	AMB

Além do resumo dos traços de personalidade, a entrada foi composta também pelas métricas de grupo tamanho, prazo e o desempenho do grupo (GP), que pode ser Bom ou Ruim. A partir desses dados, o algoritmo foi executado sobre a base atualizada com os grupos provenientes do experimento de formação de grupos. A execução do *Predictive Apriori* resultou em 57 regras. No entanto, algumas das regras foram extraídas com baixa precisão. Todas as regras com precisão menor que 60% foram removidas da lista de regras apresentadas na Tabela 16.

Toda regra retornada possui uma condição que, quando satisfeita, indica se o grupo tem mais chance de obter desempenho Bom ou Ruim. Na regra 2, por exemplo, a condição diz que se o grupo for formado por indivíduos com escore M para Afabilidade (A=M), então o desempenho do grupo na atividade colaborativa deverá ser bom (GP=BOM).

Essa regra foi extraída da base com mais de 98% de precisão (Precisão). A regra 16 da Tabela 16 foi extraída com precisão de mais de 95% e diz se o grupo tem quatro membros (tamanho=4) e o prazo é de uma semana (prazo=W), esse grupo pode falhar (GP=RUIM). Assim, cada uma delas pode ser usada para verificar se um grupo já formado funciona bem ou se pode ser um grupo ruim.

Tabela 16 – Regras extraídas da base atualizada com os 38 casos.

Id	Regra	Precisão
1	E=MB ==> BOM	98,60%
2	A=M ==> BOM	98,60%
3	N=MB ==> BOM	97,97%
4	tamanho=3 O=MA ==> BOM	97,97%
5	E=A ==> BOM	97,39%
6	N=MA ==> BOM	97,39%
7	tamanho=3 C=A ==> BOM	97,39%
8	tamanho=3 N=AB ==> BOM	97,39%
9	O=A ==> BOM	96,44%
10	O=AMB ==> BOM	96,44%
11	C=AMB ==> BOM	96,44%
12	E=M ==> BOM	96,44%
13	A=A ==> BOM	96,44%
14	tamanho=3 ==> BOM	95,04%
15	A=AMB ==> BOM	94,75%
16	tamanho=4 prazo=W ==> RUIM	94,75%
17	tamanho=4 E=AMB ==> BOM	94,75%
18	tamanho=4 E=MA ==> RUIM	94,75%
19	tamanho=4 N=M ==> BOM	94,75%
20	tamanho=5 prazo=W ==> RUIM	94,75%
21	prazo=C ==> BOM	84,32%
22	C=A ==> BOM	79,75%
23	N=M ==> BOM	79,75%
24	C=MA ==> BOM	76,93%
25	prazo=W ==> RUIM	75,83%
26	O=M ==> BOM	75,83%
27	O=MA ==> BOM	73,39%
28	tamanho=5 E=MA ==> BOM	70,85%
29	A=MA N=AMB ==> BOM	70,85%
30	A=MA ==> BOM	68,79%
31	N=AB ==> BOM	67,23%
32	tamanho=5 ==> BOM	65,01%
33	E=MA ==> BOM	65,01%
34	tamanho=4 ==> BOM	64,43%
35	E=AMB ==> BOM	64,14%
36	tamanho=4 N=AB ==> RUIM	64,14%
37	O=MA E=MA N=AB ==> RUIM	64,14%
38	N=AMB ==> BOM	63,54%

5.4.2 Aplicação na Formação de Grupos

Além de serem empregadas para verificar se um grupo formado é bom ou ruim, as regras extraídas da base poderiam auxiliar na formação de grupos. A maioria dos casos em que os grupos tiveram prazo de uma semana ($\text{prazo} = W$) foram classificados como grupos ruins. Assim, as regras extraídas sugerem que prazo longo, aumenta a probabilidade de falha. De acordo com as regras 16 e 20, essa possibilidade é maior se o grupo tem quatro ou cinco membros.

Para o traço de Afabilidade, as regras com maior precisão e que seriam usadas preferencialmente, são aquelas onde todos os indivíduos têm somente escore M (regra 2), somente escore A (regra 13), ou uma combinação de A, M e B (regra 15). Mas, de acordo com a Tabela 10, os grupos sugeridos, em relação a esse traço, combinaram escores M e A, combinação que aparece com menos precisão no conjunto de regras (regra 30).

Em relação ao traço Extroversão, os grupos seriam preferencialmente formados por combinações de indivíduos com escores M e B, todos com escore A ou todos com escore M. Quase todos os grupos sugeridos da Tabela 10 atendiam a essas restrições. Assim, para o traço Extroversão, os grupos formados com as regras foram similares àqueles formados com consultas. As regras extraídas com maior precisão para Conscienciosidade e Neuroticismo resultariam em grupos também semelhantes aos sugeridos na Tabela 10.

Considerando as regras extraídas da base, as sugestões de grupos apresentados ao professor poderiam apresentar combinações de indivíduos diferentes das que foram obtidas com as consultas à base. No entanto, não significa que, se submetidos a atividades colaborativas, os grupos formados a partir das regras teriam desempenho ruim. Os grupos sugeridos poderiam, por exemplo, ser formados apenas de indivíduos com escores A ou M para Afabilidade. Por outro lado, considerando Extroversão, Conscienciosidade e Neuroticismo, as sugestões seriam bastante parecidas com as da Tabela 10.

5.5 Discussão dos Resultados

Os grupos formados a partir das consultas no RBC (Seção 5.2) mostraram resultados promissores no sentido de auxiliar na formação de bons grupos para colaboração. A maioria deles atingiram alto nível de interação. Do ponto de vista do professor, as sugestões de formação de grupos foram positivas e os resultados foram melhores do que aqueles apresentados em atividades anteriores. E mais ainda, os resultados positivos foram obtidos em turmas em que o professor tinha a vantagem do contato com os alunos. Acredita-se que a sugestão de formação de grupos em ambientes onde professores e alunos não se conhecem pode ser ainda mais promissora.

Usando as consultas na base de casos, alguns dos grupos formados contrariam a ideia de boa combinação de indivíduos indicada nos trabalhos correlatos apresentados no Capítulo 2. Por exemplo, os grupos formados com alta Extroversão. No entanto, tais trabalhos não

consideravam a influência que um traço de personalidade pode ter no grau de influência do outro traço no mesmo indivíduo. Essa influência é considerada no RBC pois, na recuperação de casos, os cinco traços são avaliados.

A base de casos pode crescer muito com a inserção de novos casos. A alternativa proposta e experimentada neste trabalho foi a mineração de regras, a partir da base. Como as regras apontam condições para decidir se um grupo é bom ou ruim, poderiam ser usadas para substituir as consultas na tarefa de decidir o provável desempenho de um grupo. Até certo ponto, as regras extraídas formariam grupos semelhantes aos sugeridos pelas consultas.

Como as regras são construídas a partir de um resumo dos traços, a maioria delas pode ser aplicada para grupos com diferentes tamanhos, facilitando sua inclusão em aplicações que lidam com o agrupamento de alunos. Assim, foi possível iniciar o desenvolvimento de uma aplicação *Web* que faz uso das regras extraídas da base para formar grupos e deve ser integrada ao Moodle como ferramenta de suporte à colaboração. Como uso de regras não permite formar grupos com perfis diferentes daqueles representados nelas, como ocorre no RBC, a aplicação descrita no Apêndice A considera agrupamento aleatório quando não há regra que possa ser aplicada.

Nos experimentos onde a similaridade dos casos da base com a própria base de casos foi avaliada, notou-se que há muita similaridade entre os casos da base. Muita similaridade implica num grande número de casos que podem ser usados para responder o mesmo tipo de consulta. Nesse sentido, há espaço para implementação de rotinas para melhoria da base. Por exemplo, escolha dos casos mais abrangentes e remoção de outros que possam ser muito similares.

A avaliação da qualidade da base como fonte de dados demonstrou o quanto a base atual é capaz de responder as consultas e o quanto essa capacidade melhora com a inserção de novos casos. À medida que novos casos foram sendo inseridos na base não somente o volume de casos similares ao já existentes aumentaram, como também a quantidade de casos diferentes. Em outras palavras, a inserção de novos casos na base aumenta a variabilidade da base. Nesse sentido, o resultado é bastante positivo, pois sinaliza que a base pode se adaptar e ter mais abrangência e, portanto, tornando-se mais adequada à sua função dentro da proposta apresentada neste trabalho.

Conforme observado nos experimentos apresentados no Capítulo 4, os alunos que participaram do experimento apresentavam mais escores M e A para Afabilidade, Conscienciosidade e Abertura. E para os demais traços, há maior concentração de alunos com escores M. Como os alunos que participaram dos experimentos de grupo são amostras de tais perfis, era esperado que a base de casos não tivesse representantes de todos 243 perfis possíveis.

Por outro lado, muitos dos casos inseridos na base são compostos por alunos com escore M em vários traços. Um indivíduo com escore M tende a apresentar características

dos dois extremos de cada traço sendo, portanto, em certo grau, próximo a ambos os perfis A e B. No cálculo da similaridade entre dois alunos, tal proximidade é levada em consideração e pode ter mais ou menos peso conforme a função de similaridade escolhida. De toda forma, a presença de indivíduos com escore M facilita a busca de caso similar quando a consulta contém perfil ausente da base.

Independente do perfil que um indivíduo possa ter, deve ser possível encontrar pelo menos um caso que indique um grupo que lhe seja adequado. Nos experimentos conduzidos a fim de verificar a viabilidade da base, notou-se que os membros dos grupos da base representam cerca de 13% dos 243 possíveis perfis. Por outro lado, com o uso do RBC, é possível considerar como solução os casos retornados com similaridade alta. Assim, se for possível encontrar na base algum caso que tenha alta similaridade com os perfis desconhecidos, esse caso também pode ser usado como uma possível solução.

Considerando similaridade de 70%, a base com 38 casos é capaz de responder às consultas, recuperando casos similares aos 243 perfis. Em outras palavras, o RBC proposto será capaz de retornar uma solução baseada em casos anteriores, mesmo para consultas feitas com perfis desconhecidos. Se outros algoritmos de aprendizagem fossem utilizados, como por exemplo algoritmos de clusterização, poderia ser mais difícil encontrar um grupo adequado em tal situação.

Conclusão

Este trabalho foi desenvolvido com o objetivo de explorar as características de cada um dos cinco grandes traços de personalidade do *Big Five* (Abertura, Extroversão, Neuroticismo, Afabilidade e Conscienciosidade) para apoiar sugestões de formações de grupo para colaboração. Para atingir o objetivo proposto, foram conduzidos experimentos para detecção do perfil do aluno por meio do texto e experimentos de formação de grupos, envolvendo o uso de RBC e mineração de regras da base.

Para capturar texto e conhecer o perfil dos alunos, um questionário eletrônico foi disponibilizado. O questionário era composto do inventário de 44 itens do *Big Five* e outras cinco perguntas nas quais o aluno deveria descrever suas reações e possíveis respostas, usando a mesma linguagem que usa com seus colegas. Assim, além de classificados quanto aos cinco traços, os textos coletados também eram semelhantes àqueles que os alunos normalmente escrevem usando as ferramentas de comunicação nos AVAs.

Como havia maior concentração de amostras em apenas uma das faixas de escore nos traços da base de treinamento, a taxa de acerto para as outras faixas foi baixa. Assim, os valores de precisão encontrados também foram baixos. Por outro lado, foi possível notar que alguns algoritmos, com um ajuste dos parâmetros, conseguiam se ajustar para prever todas as classes. Conseguindo uma base com maior volume de amostras em cada uma das faixas de escore e com ajuste de parâmetros adequado, o valor da precisão pode ser melhorado. Como é possível obter modelos de predição mais precisos, confirma-se a viabilidade do uso do texto para prever os traços de personalidade de forma automática.

Uma vez que os traços de personalidade podem ser capturados, podem ser disponibilizados para que o professor possa empregá-los no planejamento de ações pedagógicas. Uma das ações que podem ser tomadas para apoiar o processo de aprendizagem dos alunos é a formação de grupos para colaboração. Como os traços de personalidade agrupam um conjunto de tendências de comportamento que ajudam a entender como o aluno se comporta em grupo, é uma ferramenta útil na formação de grupos para colaboração.

A proposta apresentada neste trabalho incluiu o uso de RBC para apoiar a sugestão de grupos. O RBC deriva respostas para novos problemas a partir das respostas a problemas

anteriormente resolvidos e armazenados na sua base de casos. Essa característica ajuda a formar grupos com base nos traços de personalidade, ainda que não se conheça o efeito de um determinado traço no desempenho do grupo ou do indivíduo. Como o RBC incluiu quatro etapas cíclicas por meio das quais a base de casos é atualizada, a capacidade de formar bons grupos vai melhorando também. Assim, desenvolveu-se sistema de RBC, considerando suas quatro etapas e uma base de casos planejadas para apoiar a formação de grupos.

Para formação da base de casos, foram conduzidos experimentos com turmas de alunos nas quais os grupos foram avaliados quanto a interação, nota e capacidade dos indivíduos de trabalharem juntos para resolver e entregar a atividade colaborativa concluída dentro do prazo proposto pelo professor. A base gerada foi analisada quanto ao efeito dos traços de personalidade no desempenho dos grupos e, a partir dessa análise, foi possível identificar algumas características que têm influência na colaboração.

Quanto à qualidade e abrangência da base, os experimentos conduzidos confirmaram a hipótese de que haveria melhora com a atualização da base. A partir dos experimentos foi possível também confirmar a eficácia do uso do RBC na formação de grupos. Os grupos resultantes das consultas na base, ainda que não muito parecidos com os casos a partir dos quais foram agrupados, ainda assim foram agrupados e produziram bons resultados quanto à interação nas atividades colaborativas. Uma vez que os novos grupos foram formados com algumas características diferentes dos casos presentes na base, a base foi atualizada com mais 13 novos casos.

Confirmada a hipótese de que, com o uso do RBC, haverá melhora contínua da qualidade das sugestões de agrupamento, é possível agora empregar o agrupamento segundo a proposta apresentada neste trabalho, no desenvolvimento de aplicações que integram a formação de grupo para colaboração em AVAs, a fim de apoiar e motivar os alunos nesses ambientes. Nesse sentido, deu-se início ao desenvolvimento de uma aplicação *Web* que deverá ser integrada ao ambiente Moodle.

Com a atualização, que é parte do processo de funcionamento do RBC, a base de casos pode crescer muito, fazendo as consultas computacionalmente mais pesadas e demoradas. Por essa razão, também foram conduzidos experimentos no sentido de extrair regras de formação de grupos da base de casos. Os resultados sugerem que a extração das regras é uma alternativa viável para formação de grupos, em caso de aumento do volume de casos.

6.1 Principais Contribuições

O trabalho proposto visou validar duas hipóteses principais. A primeira hipótese era referente à relação entre os traços de personalidade e textos escritos. Se essa relação existe, então é possível encontrar um algoritmo adequado para classificar os alunos dentro dos cinco traços de personalidade do *Big Five*. A segunda hipótese sobre o uso dos traços

de personalidade e RBC na formação de bons grupos para colaboração. Ao confirmar as hipóteses propostas, as seguintes contribuições podem ser apontadas:

- ❑ Criação de um modelo de detecção automática e transparente de traços de personalidade dos alunos a partir de seus textos que pode ser integrado a um AVA, oferecendo suporte ao professor/tutor na condução de atividades colaborativas ou a sistemas de recomendação baseados no perfil do aluno.
- ❑ Definição de proposta que viabiliza o emprego dos traços de personalidade na formação de bons grupos para colaboração.
- ❑ Definição dos casos e estratégias de funcionamento do RBC capazes de sugerir bons grupos.
- ❑ Definição de abordagem que permite melhora contínua da qualidade das sugestões de grupos para colaboração, por meio do ciclo de funcionamento do RBC.
- ❑ Criação e manutenção de uma base de casos atualizadas pela inclusão de novos casos e da qual é possível minerar regras de formação de grupos mais precisas.
- ❑ Extração de conjunto de regras de formação de grupo de alunos, baseados em traços de personalidade, que recomendem grupos com interação e troca de conhecimento satisfatórias, durante a execução de atividade colaborativa.
- ❑ Elaboração de ferramenta de apoio ao processo de aprendizagem por meio da colaboração, sem que o aluno seja envolvido na avaliação dos traços de personalidade e dos grupos.

6.2 Trabalhos Futuros

O próximo passo a partir do trabalho até aqui desenvolvido é coletar mais dados para melhorar a qualidade da classificação dos modelos de previsão de traços de personalidade. A base que foi coletada para desenvolver este trabalho era formada de textos muito curtos, com média de aproximadamente 82 palavras por texto. Um volume de texto maior seria mais interessante para treinar os modelos.

Além disso, a maioria dos participantes tinham escore de médio a alto para Abertura, Conscienciosidade e Afabilidade e também escore médio para Extroversão e Neuroticismo. Assim, seria interessante refazer os testes a partir de uma base com novas amostras de maior representatividade nos cinco traços (em quantidade de exemplos). Um experimento incluindo seleção de atributos, baseado nas categorias com maior correlação com os traços, também deveria ser considerada para tentar melhorar a qualidade da predição dos traços de personalidade a partir de texto.

Quanto ao uso do RBC, ainda que a base tenha se mostrado abrangente no sentido de entregar uma solução com alta similaridade, a inserção de novos casos com aqueles perfis que ainda não estão representados na base, poderá torná-la ainda mais eficiente para responder às novas consultas. A implementação atual do RBC não prevê a seleção de casos. Assim, como trabalho futuro, propõe-se a inclusão de uma estratégia de seleção de casos mais representativos e abrangentes, bem como a exclusão de casos que parecem repetir informação já contida na base, visando tornar o processo de escolha dos melhores casos menos custoso.

A inclusão de uma estratégia de indexação dos casos ou a associação do RBC com outra técnica de IA para auxiliar na seleção de casos pode trazer benefícios à proposta e deverá ser testada em experimentos futuros. Em relação à sugestão de grupos também devem ser conduzidos novos experimentos em outras turmas de alunos, a fim de validar a efetividade das regras extraídas da base na formação de grupos para colaboração.

A definição da estrutura do caso, dentro da proposta apresentada neste trabalho, considerou como característica do aluno apenas seus traços de personalidade. Também não foram incluídas dentro da estrutura do caso as características da atividade, como seu nível de dificuldade. A estrutura definida permite que outras informações sejam incluídas, por exemplo, se o aluno já trabalhou em grupo antes, suas preferências, habilidades e, também, sobre a tarefa, como seu grau de dificuldade. Tal inclusão poderia ser considerada em trabalho futuro como forma de aprimorar a sugestão de novos grupos.

Os resultados dos experimentos conduzidos neste trabalho já permitem sua inclusão em uma aplicação voltada para a formação de grupos. Assim, devem ser incluídos em uma aplicação *Web* disponível para o professor como uma ferramenta para gerar e avaliar grupos de alunos. Essa aplicação deverá também ser acoplada a AVAs, como o Moodle. Descrição de como a aplicação deve ser implementada encontra-se no Apêndice A.

6.3 Contribuições em Produção Bibliográfica

O trabalho desenvolvido resultou na publicação de dois artigos. Um artigo resultante do trabalho desenvolvido junto com um estudante de graduação, bolsista de iniciação científica, e outro com os resultados dos experimentos iniciais que serviram de insumo para o desenvolvimento da proposta apresentada nesta dissertação. São eles:

- *Development of a Big-Five personality traits classification approach via analysis of texts in Brazilian Portuguese*, publicado nos anais do IV *Workshop de Iniciação Científica em Sistemas de Informação (WICSI)* do XIII Simpósio Brasileiro de Sistema de Informação, em 2017 (NETO; FERREIRA; FERNANDES, 2017).

- ❑ **Detecção de traços de personalidade em textos para apoiar a formação de grupos para colaboração**, publicado nos anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE), em 2017 (FERREIRA; FERNANDES, 2017).
- ❑ **Detecção automática de traços de personalidade e recomendação de agrupamento com o modelo Big Five**, aceito para publicação, em 2018 (FERREIRA et al., 2018).

Referências

- AKHTAR, R. et al. The engageable personality: Personality and trait EI as predictors of work engagement. **Personality and Individual Differences**, v. 73, p. 44–49, 2015. DOI: <<https://doi.org/10.1016/j.paid.2014.08.040>>.
- ALAVI, M. Computer-mediated collaborative learning: An empirical evaluation. **Journal MIS Quarterly**, v. 18, p. 159–174, 1994. DOI: <<https://doi.org/10.2307/249763>>.
- ALTANOPOULOU, P.; TSELIOS, N. How does personality affect wiki-mediated learning? In: **Proceedings of International Conference on Interactive Mobile and Communication Technologies and Learning**. Thessaloniki, Greece: IEEE, 2015. p. 16–18. DOI: <<https://doi.org/10.1109/IMCTL.2015.7359546>>.
- ANDRADE, J. M. **Evidências de Validade do Inventário dos Cinco Grandes Fatores de Personalidade para o Brasil**. Tese (Doutorado) — Instituto de Psicologia - Universidade de Brasília, 7 2008.
- BERGMANN, R.; STAHL, A. Similarity measures for object-oriented case representations. In: **European Workshop on Advances in Case-Based Reasoning: Advances in Case-Based Reasoning**. Berlin, Heidelberg: Springer, 1998. v. 1488, p. 25–36. DOI: <<https://doi.org/10.1007/BFb0056319>>.
- BIRD, S.; KLEIN, E.; LOPER, E. **Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit**. Sebastopol, CA: O'Reilly, 2009.
- BOZIOELOS, G. The relationship of the big-five with workplace network resources: More quadratic than linear. **Personality and Individual Differences**, v. 104, p. 374–378, 2017. DOI: <<https://doi.org/10.1016/j.paid.2016.08.036>>.
- BRINDLEY, J. E.; WALTI, C.; BLASCHKE, L. M. Creating effective collaborative learning groups in an online environment. **The International Review of Research in Open and Distributed Learning**, v. 10, 2009.
- BUIAR, J. A.; PIMENTEL, A. R.; OLIVEIRA, L. E. S. Identificação de estilo de aprendizagem: Um modelo de inferência automatizado baseado no perfil de personalidade identificado nos textos produzidos pelo aluno. In: **Proceedings of Brazilian Symposium on Computers in Education**. Recife, PE, Brasil: Sociedade Brasileira de Computação (SBC), 2017. p. 1157–1166. DOI: <<http://dx.doi.org/10.5753/cbie.sbie.2017.1157>>.

CARRO, R. M.; SANCHEZ-HORREO, V. The effect of personality and learning styles on individual and collaborative learning: Obtaining criteria for adaptation. In: **Global Engineering Education Conference (EDUCON), 2017 IEEE**. Atenas, Grécia: IEEE, 2017. p. 1585–1590. DOI: <<https://doi.org/10.1109/EDUCON.2017.7943060>>.

CHOWDHURY, G. Natural language processing. **Annual Review of Information Science and Technology**, v. 37, p. 51–89, 2003. DOI: <<https://doi.org/10.1002/ar.1440370103>>.

COCEA, M.; MAGOULAS, G. D. User behaviour-driven group formation through case-based reasoning and clustering. **Expert Systems with Applications**, v. 39, p. 8756–8768, 2012. DOI: <<https://doi.org/10.1016/j.eswa.2012.01.205>>.

DIGMAN, J. M. Personality structure: emergence of the five-factor model. **Annual Review of Psychology**, v. 41, p. 417–440, 1990. DOI: <<https://doi.org/10.1146/annurev.ps.41.020190.002221>>.

DILLENBOURG, P. What do you mean by collaborative learning. **Collaborative-learning: Cognitive and Computational Approaches**, Oxford: Elsevier, p. 1–19, 1999.

EVERSON, M. What can you gain by teaching online? **eLearn Magazine**, ACM, v. 2011, n. 1, p. 3, 2011.

FERREIRA, T.; FERNANDES, M. Detecção de traços de personalidade em textos para apoiar a formação de grupos para colaboração. In: **Proceedings of Brazilian Symposium on Computers in Education**. Recife, PE, Brasil: Sociedade Brasileira de Computação (SBC), 2017. p. 1627–1636. DOI: <<http://dx.doi.org/10.5753/cbie.sbie.2017.1627>>.

FERREIRA, T. B. et al. Detecção automática de traços de personalidade e recomendação de agrupamento com o modelo big five. Aceito para publicação. 2018.

FILHO, P. P. B.; ALUÍSIO, S. M.; PARDO, T. A. S. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In: **Proceedings of Brazilian Symposium in Information and Human Language Technology**. Fortaleza, CE, Brazil: Sociedade Brasileira de Computação (SBC), 2013. p. 215–219.

FRANK, E.; HALL, M. A.; WITTEN, I. H. **The WEKA Workbench Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques**. [S.l.]: Morgan Kaufmann, 2016.

GOLDBERG, L. R. Language and individual differences: The search for universal in personality lexicons. **Review of personality and social psychology**, v. 2, p. 141–166, 1981.

GRANT, A. M. Rethinking the extravert sales ideal: The ambivert advantage. **Psychological Science**, Sage Publications Sage CA: Los Angeles, CA, v. 24, n. 6, p. 1024–1030, 2013. DOI: <<https://doi.org/10.1177/0956797612463706>>.

HRON, A.; FRIEDRICH, H. F. A review of web-based collaborative learning: factors beyond technology. **Journal of Computer Assisted Learning**, v. 19, p. 70–79, 2003. DOI: <<https://doi.org/10.1046/j.0266-4909.2002.00007.x>>.

- JOHN, O.; SRIVASTAVA, S. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. **Handbook of personality: Theory and research**, v. 2, p. 102–138, 1999.
- KENT, C.; LASLO, E.; RAFAELI, S. Interactivity in online discussions and learning outcomes. **Computers and Education**, v. 97, p. 116–128, 2016. DOI: <<https://doi.org/10.1016/j.compedu.2016.03.002>>.
- KOLODNER, J. L. An introduction to case-based reasoning. **Artificial Intelligence Review**, v. 6, p. 3–34, 1992. DOI: <<https://doi.org/10.1007/BF00155578>>.
- _____. Educational implications of analogy a view from case-based reasoning. **American Psychologist**, v. 52, p. 57–66, 1997. DOI: <<https://doi.org/10.1037/0003-066X.52.1.57>>.
- KOP, R. The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. **The International Review of Research in Open and Distance Learning**, v. 12, n. 3, p. 19–38, 2011. DOI: <<https://doi.org/10.19173/irrodl.v12i3.882>>.
- KUMAR, R.; ROSE, C. Triggering effective social support for online groups. **ACM Transactions on Interactive Systems**, v. 3, p. artigo 24, 2014. DOI: <<https://doi.org/10.1145/2499672>>.
- KWANTES, P. J. et al. Assessing the big five personality traits with latent semantic analysis. **Personality and Individual Differences**, v. 102, p. 229–233, 2016. DOI: <<https://doi.org/10.1016/j.paid.2016.07.010>>.
- KWON, K.; LIUB, Y.; JOHNSONC, L. P. Group regulation and social-emotional interactions observed in computer supported collaborative learning: Comparison between good vs. poor collaborators. **Computers and Education**, v. 78, p. 185–200, 2014. DOI: <<https://doi.org/10.1016/j.compedu.2014.06.004>>.
- LAI, E. R. **Collaboration: A Literature Review**. [S.l.]: Pearson, 2011. 49 p.
- LAMPRECHT, J.; SIEMON, D.; ROBBA-BISSANTZ, S. Cooperation isn't just about doing the same thing – using personality for a cooperation-recommender-system in online social networks. **Collaboration and Technology**, p. 131–138, 2016. DOI: <https://doi.org/10.1007/978-3-319-44799-5_10>.
- LIU, B. Sentiment analysis and subjectivity. In: INDURKHYA, N.; DAMERAU, F. J. (Ed.). **Handbook of Natural Language Processing, Second Edition**. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2010. ISBN 978-1420085921.
- MACHADO, A. A. et al. Personalitatem lexicon: Um léxico em Português Brasileiro para mineração de traços de personalidade em textos. In: **Anais do XXVI Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação (SBC), 2015. p. 1122–1126. DOI: <<http://dx.doi.org/10.5753/cbie.sbie.2015.1122>>.
- MAGNISALIS, I.; DEMETRIADIS, S.; KARAKOSTAS, A. Adaptive and intelligent systems for collaboration learning support: A review of the field. **IEEE Transactions on Learning Technologies**, v. 4, p. 5–20, 2011. DOI: <<https://doi.org/10.1109/TLT.2011.2>>.

MANSKE, S. et al. Using differences to make a difference: a study on heterogeneity of learning groups. In: **Proceedings of International Conference on Computer Supported Collaborative Learning**. Gothenburg, Sweden: International Society of the Learning Sciences, Inc. (ISLS), 2015. p. 182–189.

MAPLES, J. L. et al. A test of the international personality item pool representation of the revised neo personality inventory and development of a 120-item ipip-based measure of the five-factor model. **Psychological Assessment**, v. 26, p. 1070–1084. DOI: <<https://doi.org/10.1037/pas0000004>>.

MCCRAE, R. R.; JOHN, O. An introduction to the five-factor model and its applications. **Journal of Personality**, v. 60, p. 175–215, 1992. DOI: <<https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>>.

NETO, A. T.; FERREIRA, T.; FERNANDES, M. Development of a big-five personality traits classification approach via analysis of texts in Brazilian Portuguese. In: **Anais do IV WICSI - XIII Simpósio Brasileiro de Sistema de Informação**. Lavras, MG, Brasil: Sociedade Brasileira de Computação (SBC), 2017. p. 57–60.

PACHECO, L.; SISTO, F. F. Aprendizagem por interação e traços de personalidade. **Psicologia Escolar e Educacional**, scielo, v. 7, p. 69 – 76, 06 2003. ISSN 1413-8557.

PAIM, A.; CAMATI, R.; ENEMBRECK, F. Inferência de personalidade a partir de textos em Português utilizando léxico linguístico e aprendizagem de máquina. In: **Anais do XIII Encontro Nacional de Inteligência Artificial e Computacional**. Recife, PE, Brasil: Sociedade Brasileira de Computação (SBC), 2016. p. 481–492.

PASQUALOTTI, P. R.; VIEIRA, R. Wordnet affectbr: uma base lexical de palavras de emoções para a Língua Portuguesa. **Revista Novas Tecnologias na Educação**, v. 6, n. 1, 2008. DOI: <<https://doi.org/10.22456/1679-1916.14693>>.

POOLE, D. L.; MACKWORTH, A. K. **Artificial Intelligence - Foundations of Computational Agents**. Vancouver: Cambridge University Press, 2010. DOI: <<https://doi.org/10.1017/CBO9780511794797>>.

REIS, R. C. D. et al. Estado da arte sobre afetividade na formação de grupos em ambientes colaborativos de aprendizagem. **Revista Brasileira de Informática na Educação**, v. 23, p. 113–130, 2015. DOI: <<https://doi.org/10.5753/rbie.2015.23.03.113>>.

_____. Relação entre os estados afetivos e as teorias de aprendizagem na formação de grupos em ambientes cscl. In: **Anais do XXVI Simpósio Brasileiro de Informática na Educação**. Maceio, AL, Brazil: Sociedade Brasileira de Computação (SBC), 2015. p. 1012–1021. DOI: <<http://dx.doi.org/10.5753/cbie.sbie.2015.1012>>.

RICCI, F.; AVESANI, P. Learning a local similarity metric for case-based reasoning. In: **International Conference on Case-Based Reasoning (ICCBR): Case-Based Reasoning Research and Development**. Sesimbra, Portugal: Springer, 1995. v. 1010, p. 301–312. DOI: <https://doi.org/10.1007/3-540-60598-3_27>.

ROBERTS, S. G. B. et al. Individual differences and personal social network size and structure. **Personality and Individual Differences**, v. 44, p. 954–964, 2008. DOI: <<https://doi.org/10.1016/j.paid.2007.10.033>>.

RUTHERFOORD, R. H. Using personality inventories to form teams for class projects – a case study. In: **Proceedings of SIGITE'06 Proceedings of the 7th conference on Information Technology Education**. Canterbury, United Kingdom: ACM, 2006. p. 73–76. DOI: <<https://doi.org/10.1145/1168812.1168817>>.

SANTOS, O. C. et al. Helping the tutor to manage a collaborative task in a web-based learning environment. In: **Supplementary Proceedings of International Conference on Artificial Intelligence in Education**. Sidney, Austrália: [s.n.], 2003. v. 4, p. 153–162.

SCHEFFER, T. Finding association rules that trade support optimally against confidence. In: **5th European Conference on Principles of Data Mining and Knowledge Discovery**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 424–435. DOI: <https://doi.org/10.1007/3-540-44794-6_35>.

SILVA, F. E. O.; MOTTA, C. L. R.; SANTORO, F. M. Team composer: Assembling groups through social matching. In: **Proceedings of the International Conference on Computer Supported Cooperative Work in Design**. Shanghai, China: IEEE, 2010. p. 128–133. DOI: <<https://doi.org/10.1109/CSCWD.2010.5471990>>.

SPOELSTRA, H. et al. Team formation instruments to enhance learner interactions in open learning environments. **Computers in Human Behavior**, v. 45, p. 11–20, 2015. DOI: <<https://doi.org/10.1016/j.chb.2014.11.038>>.

STAHL, A. **Learning of Knowledge-Intensive Similarity Measures in Case-Based Reasoning**. Tese (Doutorado) — Departamento de Ciência da Computação da Universidade de Kaiserslautern, 10 2003.

STAHL, G. **Group Cognition: Computer Support for Building Collaborative Knowledge**. London, England: MIT Press, 2006.

SURMA, J.; VANHOOF, K. Integrating rules and cases for the classification task. In: . Berlin, Heidelberg: Springer, 1995. v. 1010, p. 325–334. DOI: <https://doi.org/10.1007/3-540-60598-3_29>.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. **Journal of Language and Social Psychology**, v. 29, p. 24–54, 2010. DOI: <<https://doi.org/10.1177/0261927X09351676>>.

VYGOTSKY, L. **Mind in Society: The Developmant of Higher Psychological Processes**. Cambridge, MA: Harvard University Press, 1978.

WANG, D. Y.; LIN, S. S. J.; SUN, C. T. Diana: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. **Computers in Human Behavior**, v. 23, p. 1997–2010, 2007. DOI: <<https://doi.org/10.1016/j.chb.2006.02.008>>.

WATSON, D.; CLARK, L. A. Chapter 29 - extraversion and its positive emotional core. In: HOGAN, R.; JOHNSON, J.; BRIGGS, S. (Ed.). **Handbook of Personality Psychology**. San Diego: Academic Press, 1997. p. 767 – 793. ISBN 978-0-12-134645-4.

WEN, M.; YANG, D.; ROSE, C. P. Linguistic reflections of student engagement in massive open online courses. In: **Proceedings of the International Conference on Weblogs and Social Media**. Ann Arbor, Michigan, USA: The AAAI Press, 2014. p. 525–534.

WU, W.; CHEN, L. Implicit acquisition of user personality for augmenting movie recommendations. In: **Proceedings of International Conference on User Modeling, Adaptation and Personalization**. Dublin, Ireland: Springer, 2015. p. 302–314. DOI: <https://doi.org/10.1007/978-3-319-20267-9_25>.

XIE, K.; MILLER, N. C.; ALLISON, J. R. Toward a social conflict evolution model: Examining the adverse power of conflictual social interaction in online learning. **Computers and Education**, v. 63, p. 404–415, 2013. DOI: <<https://doi.org/10.1016/j.compedu.2013.01.003>>.

XING, W.; KIM, S. M.; GOGGIN, S. Modeling performance in asynchronous cscl: An exploration of social ability, collective efficacy and social interaction. In: **Proceedings of International Conference on Computer Supported Collaborative Learning**. Gothenburg, Sweden: International Society of the Learning Sciences, Inc. (ISLS), 2015. v. 1, p. 182–189.

Apêndices

Aplicação Web para Formação de Grupos

Os resultados dos experimentos obtidos neste trabalho foram empregados no desenvolvimento de uma aplicação *Web*. A aplicação deve ser integrada ao Moodle por meio de um *plugin*. A aplicação, ainda em desenvolvimento, foi planejada com o objetivo de permitir a formação de grupos usando as regras extraídas da base de casos do RBC. As regras são também usadas para alertar o professor sobre os grupos que possuem características que aumentam a chance de falha do grupo. O acesso do aluno também está previsto.

Uma, vez integrada ao Moodle, o banco de dados da aplicação deve ser populado por meio da transferência periódica dos dados do banco da plataforma. De acordo com a estrutura proposta, a transferência dos dados (*Data Transfer*) será executada de periodicamente e de forma independente dos demais componentes por um processo que executa em *background*. Os dados que devem ser carregados para o banco da aplicação são:

- ❑ Textos gerados pelos alunos das atividades no Moodle;
- ❑ Informação de qual professor é responsável por qual turma;
- ❑ Quais alunos pertencem a cada turma;
- ❑ Avaliação das atividades feitas pelo professor.

Conforme apresentado na Figura 11, o texto capturado durante a transferência de dados passa por tratamento de texto (*Text Processing*) e o resultado do processamento é então usado para calcular ou atualizar os traços de personalidade do *Big Five* de cada aluno (*Big Five Classifier*), no *NLP module*. O *Grouping support module* é o responsável por executar as rotinas relacionadas a manutenção da base de casos do RBC, como atualizar os casos a partir de novas informações de grupos que trabalharam em conjunto, e atualização da base de regras de agrupamento.

Quando o professor (ou o aluno) aciona o *Plugin* no moodle, há o redirecionamento para a interface *Web* da aplicação (*Web Interface*). O *plugin*, além da função de direcionar

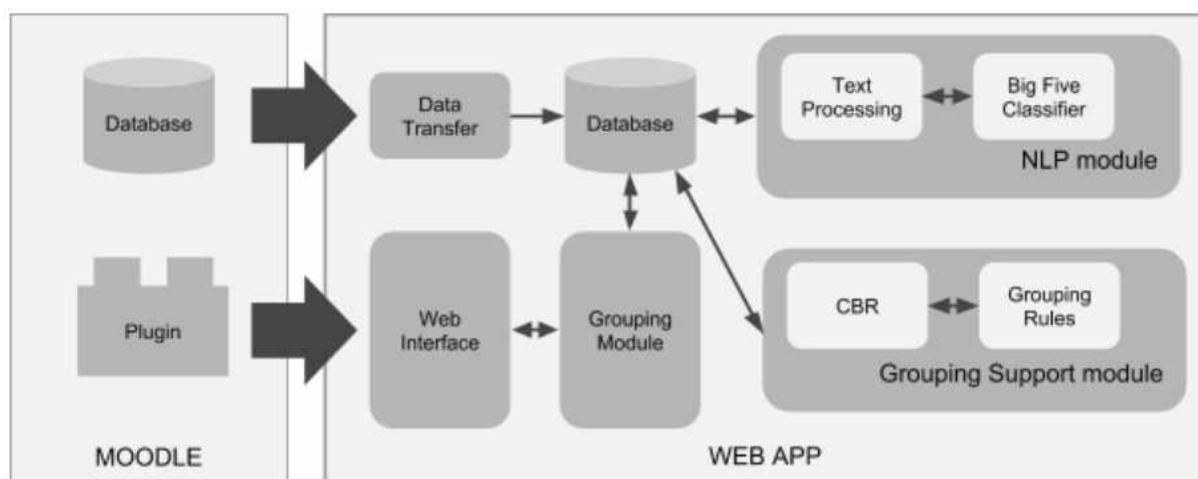


Figura 11 – Integração da aplicação com o Moodle.

o usuário para a aplicação de formação de grupos, deve também usar as informações de usuário do Moodle para fazer *login* na aplicação. Assim, dentro da aplicação, se o usuário do Moodle tinha perfil de professor, verá as opções de professor. Se o perfil era de aluno, vai ver somente as opções de aluno.

Todas as ações na interface são processadas pelo *Grouping Module* que é o módulo que executa o algoritmo de formação de grupos usando as regras de agrupamento calculadas no *grouping support module*, busca perfil dos alunos calculados no *NLP module*, calcula as estatísticas sobre os grupos que devem ser mostradas na interface.

Na interface web, tanto professor quanto aluno poderão ver estatísticas sobre os grupos formados em turmas ativas ou em turmas cujas aulas já foram encerradas. A opção de opção de filtrar as turmas por status (ativa/encerrada) e período pode ser usada no momento da seleção de uma turma. Ao selecionar uma turma o aluno, vai ter acesso somente as informações de desempenho relacionadas ao grupo do qual é membro. Já o professor o professor tem opção de acessar a mesma visão de estatísticas comparativas que o aluno tem, ou ver estatísticas da turma como um todo. O perfil de professor também tem acesso ao perfil de cada um dos alunos da turma. Além das estatísticas, ao selecionar uma turma, o professor tem acesso as ações de formar grupos e de reorganizar os grupos já formados, alterando manualmente a composição dos grupos.

A ação de formar grupos dispara as ações do *Grouping Module* conforme exibido na Figura 12. Para formar grupos, o professor seleciona uma turma (*Select a class*) e solicita o início do agrupamento (*Group students*). Ao iniciar o fluxo de formação de grupos, a aplicação procura pelos traços de personalidade dos alunos (*Retrieve students traits*) da turma selecionada pelo professor.

Se existir algum aluno cujos traços de personalidade são desconhecidos (*Someone unknown = yes*), envia um aviso para o professor sobre os alunos com perfil desconhecido

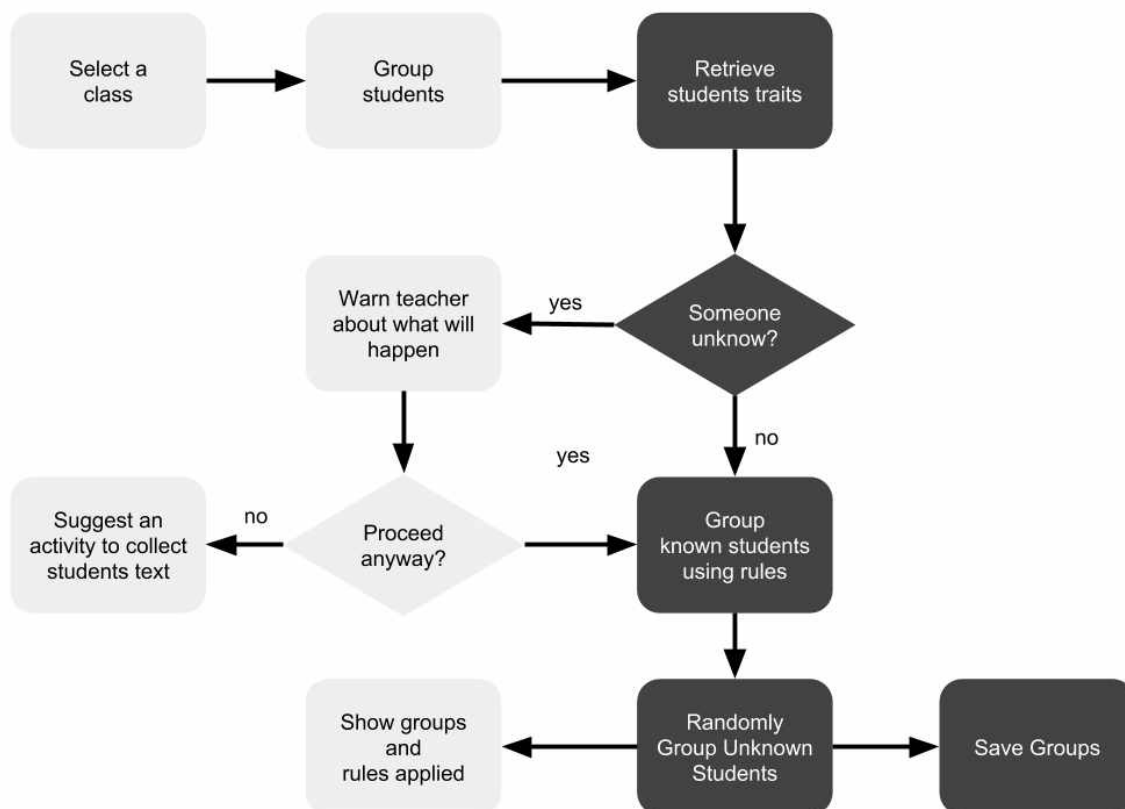


Figura 12 – Formação de grupos dentro da aplicação.

e que vai agrupar tais alunos de forma aleatória. Se o professor permitir continuar, os alunos conhecidos são agrupados com as regras (*Group known students using rules*) e o restante de forma aleatória (*Randomly group unknown*). Se o professor decidir parar o processo, avisa sobre o que deve ser feito para gerar dados para classificar os alunos desconhecidos: uma atividade que gere texto na plataforma. Se todos os alunos são conhecidos, agrupa todos com as regras (*Group known students using rules*) e continua sem executar nenhuma ação em *Randomly group unknown students*.

Uma vez que os grupos estejam formados, são salvos no banco e exibidos na interface para o professor. Todos os grupos para os quais o perfil de cada membro é conhecido são associados às regras de agrupamento. Assim, ao serem exibidos na interface, as regras que se aplicam aos grupos também são exibidas. Se casa com uma regra apontando a possibilidade de desempenho ruim, o grupo é marcado com um sinal de aviso. Assim o professor pode reagrupá-los manualmente, se desejar. O processo de reagrupar os alunos de um grupo é apresentado da Figura 13.

Caso o professor decida realocar estudantes em outros grupos, ao mover um estudante para outro grupo (*Move a student to another group*), a aplicação vai verificar quais as

regras se aplicam aos grupos alterados (*poor group rule matched*) para reexibir os grupos com as regras e avisar o professor caso a mudança tenha uma chance de gerar grupo de baixo desempenho (*Warn teacher*). Quando o professor terminar suas alterações (*Continue changing = yes*) e confirmá-las (*Confirme changes*) o sistema salva as alterações no banco da aplicação (*Save changes*).

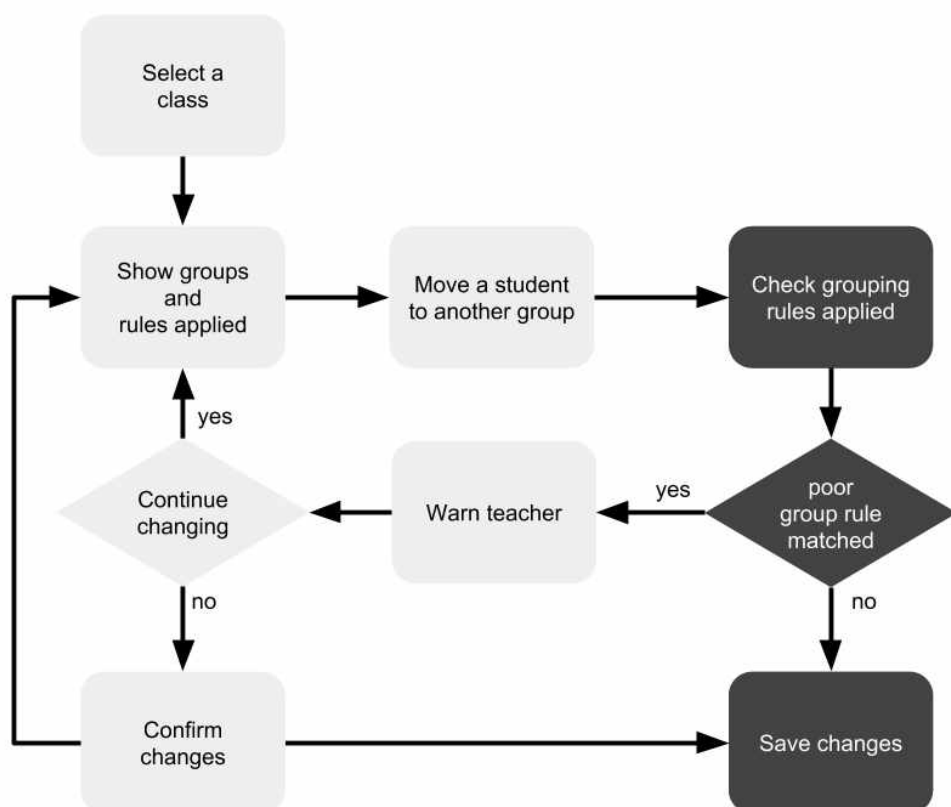


Figura 13 – Reordenação manual dos grupos formados.

No estágio atual em que se encontra o desenvolvimento da aplicação, já é possível formar e reorganizar os grupos. A formação dos grupos considerando as regras, a conexão com o moodle e a integração com os módulos de processamento de linguagem natural e de suporte à formação de grupos não está concluída. No entanto, as interfaces de acesso aos cursos disponíveis, os alunos e os grupos dentro de um curso, a criação de novos grupos, as regras que indicam a possível performance do grupo e a reordenação dos alunos após os grupos terem sido formados já foram implementadas e podem ser observadas nas Figuras 14, 15, 16, 17, 18 e 19.

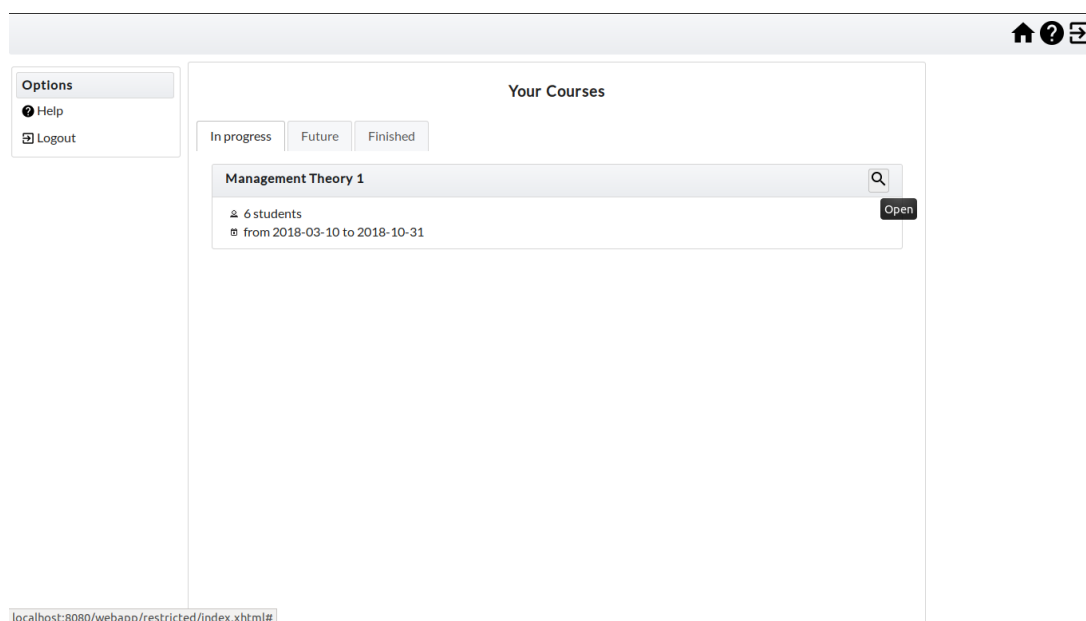


Figura 14 – Turmas as quais o usuário está relacionado e tem permissão de visualizar.

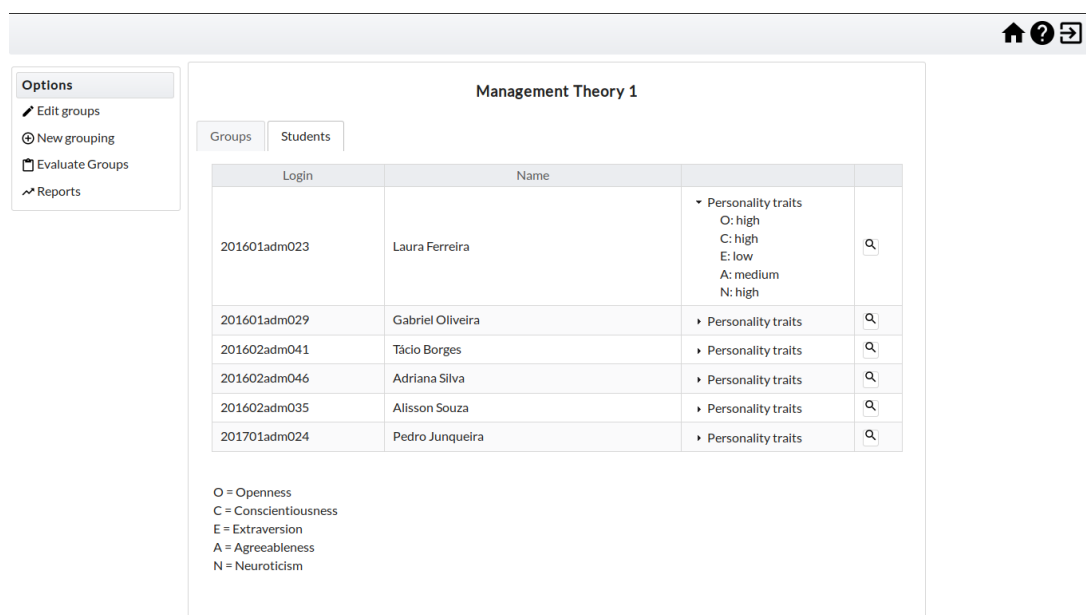


Figura 15 – Interface do curso com a lista de alunos matriculados e as opções de ação disponíveis para professor.

Options

- Edit groups
- New grouping
- Evaluate Groups
- Reports

Group 1

- 2 members
 - Students
 - 201601adm023 (Laura Ferreira)
 - 201601adm029 (Gabriel Oliveira)
 - Group personality traits summary
 - O: M or H
 - C: M or H
 - E: extremes
 - A: M or L
 - N: H
 - Average group performance (Good Group)
 - Grade: 100.0%
 - Tasks completed: 100.0%
 - Working together on tasks: 100.0%
 - Interaction (members communication): 100.0%
 - Tasks submitted late: 0.0%
- 1 group tasks evaluated

Group 2

- 2 members
 - Students
 - Group personality traits summary
- 1 group tasks evaluated
 - Average group performance (Good Group)

Group 3

- 1 members
 - Students

Figura 16 – Informações dos grupos formados no curso (visão do professor).

Options

- New grouping
- Discart groups
- Apply to course
- Course page

Groups

Groups	O	C	E	A	N	Actions
Group 1	M or H	M or H	extremes	M or L	H	X
Group 2	M or L	M or H	H	L	M or L	X
Group 3	M	M	M	M	M	X

Ungrouped students

Name	E	A	N	Actions
Alisson	medium	high	medium	

New grouping

Group Size: 2 members

Create groups Cancel

Figura 17 – Seleção do tamanho dos grupos para gerar novo agrupamento na turma.

The screenshot shows a web application interface for managing student groups. On the left, there is a sidebar with 'Options' including 'New grouping', 'Discart groups', 'Apply to course', and 'Course page'. The main area displays a table titled 'Groups' with columns: Groups, O, C, E, A, N, and Actions. The table lists three groups: Group 50, Group 51, and Group 52. Group 52 is highlighted with a yellow warning triangle. A tooltip is visible over Group 52, listing five rules with their accuracy scores.

Groups	O	C	E	A	N	Actions
Group 50	M or H	M or H	M or H	extremes	M or H	X
Gabriel	medium	medium	high	low	high	↺
Alisson	high	high	medium	high	medium	↺
Group 51	M or L	M or H	H	L	M or L	X
Tácio	medium	medium	high	low	low	↺
Adriana	low	high	high	low	medium	↺
Group 52	M or H	M or H	M or L	M	M or H	X
	medium	medium	medium	medium	medium	↺
	high	high	low	medium	high	↺

Regra 4: A = M => Good group (acc = 0.98252)
 Regra 5: N = M or H & task deadline = class time => Good group (acc = 0.98252)
 Regra 10: E = M or L => Good group (acc = 0.96124)
 Regra 18: C = M or H & task deadline = one week => Poor group (acc = 0.96124)
 Regra 23: N = M or H => Good group (acc = 0.80521)

localhost:8080/webapp/restricted/course/groups.xhtml#

Figura 18 – Regras associadas ao grupo 52 do novo agrupamento. O grupo está marcado com um sinal de alerta indicando a associação das características do grupo com uma regra que indica possibilidade de desempenho ruim.

The screenshot shows the same 'Groups' management interface. A 'Move student' dialog box is open in the foreground. The dialog asks 'Where do you want to move the student?' and provides a dropdown menu with 'Group 54' selected. There are 'Move' and 'Cancel' buttons. The background table shows groups 54, 55, and 56. Group 55 has a yellow warning triangle, and Group 56 has a red error icon.

Move student

Where do you want to move the student?
 The option "ungroup" will exclude the student from his/her group.

Group 54 ▼ Move Cancel

Groups	O	C	E	A	N	Actions
Group 54	M	M	H	L	H	X
Gabriel	medium	medium	high	low	high	↺
Group 55	M or L	M or H	H	L	M or L	X
Group 56	M	M	M	M	M	X

Figura 19 – Reordenação manual dos grupos movendo um aluno.