

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vinícius Henrique Resende

**Combinação de Técnicas Convencionais e
Baseadas em Redes Complexas para
Classificação Multirrótulo**

Monte Carmelo, Brasil

2018

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Vinícius Henrique Resende

**Combinação de Técnicas Convencionais e Baseadas em
Redes Complexas para Classificação Multirrótulo**

Trabalho de conclusão de curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, Minas Gerais, como
requisito exigido parcial à obtenção do grau
de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Murillo Guimarães Carneiro

Universidade Federal de Uberlândia – UFU

Faculdade de Ciência da Computação

Bacharelado em Sistemas de Informação

Monte Carmelo, Brasil

2018

Vinícius Henrique Resende

Combinação de Técnicas Convencionais e Baseadas em Redes Complexas para Classificação Multirrótulo

Trabalho de conclusão de curso apresentado
à Faculdade de Computação da Universidade
Federal de Uberlândia, Minas Gerais, como
requisito exigido parcial à obtenção do grau
de Bacharel em Sistemas de Informação.

Monte Carmelo, Brasil, 19 de Dezembro de 2018

Prof. Dr. Murillo Guimarães Carneiro
Orientador

Leandro Nogueira Couto
Membro da banca

Marcos Luiz de Paula Bueno
Membro da banca

Monte Carmelo, Brasil
2018

Dedico este trabalho a minha mãe Silvânia, meu pai Fábio, minha irmã Francielle e também á todos meus amigos e leitores.

Agradecimentos

Agradeço primeiramente a Deus por guiar meus passos e iluminar minha vida trazendo diversas conquistas e pessoas incríveis que me incentivaram muito durante todo esse tempo. Aos meus pais por me dar suporte durante toda minha vida, sempre dispostos a fazer o possível e impossível para me ajudar.

Agradeço o meu orientador Prof. Dr. Murillo Carneiro, por toda atenção e comprometimento em sempre ajudar, incentivar e me inspirar a ser um bom pesquisador.

Agradeço também a Universidade Federal de Uberlândia pela oportunidade de fazer um curso gratuito e de excelente qualidade.

Resumo

A classificação se trata de um tópico importante do aprendizado de máquina e mineração de dados que têm por objetivo resolver e automatizar problemas de categorização ou predição, que, basicamente, consiste em associar a um objeto uma categoria que o melhor caracterize ou represente. Normalmente, tal categoria é denominada como *classe* ou *rótulo*. Em classificação, os algoritmos e técnicas tradicionais associam um objeto a uma única classe, impossibilitando a resolução de problemas em que os objetos possam ter múltiplos significados semânticos simultaneamente, como por exemplo, sugestão automática de tags, diagnóstico de doenças, classificação de músicas em gêneros, entre outros. A ideia de classificação multirrótulo é justamente lidar com esses problemas onde suas instâncias podem pertencer a mais de uma classe ao mesmo tempo. Em comum, a maioria dos algoritmos de classificação mono ou multirrótulo existentes na literatura analisam apenas os atributos físicos dos dados, utilizando por exemplo, medidas de similaridade ou distância, ignorando assim, características e relações importantes dos dados, como a formação de padrão. Inspirado em recentes avanços no uso de redes complexas para classificação de dados, este trabalho visa investigar uma solução que combina classificadores multirrótulo já existentes com técnicas baseadas em redes complexas, apresentando uma nova ideia de classificação multirrótulo que, além dos atributos físicos, analisa também a estrutura topológica dos dados. Resultados de experimentos realizados em quatro bases de dados reais mostram que tal abordagem é promissora.

Palavras-chave: Classificação Multirrótulo, Redes Complexas, Aprendizado de Máquina, Aprendizado Supervisionado.

Lista de ilustrações

Figura 1	– (a) Classificação Tradicional. (b) Classificação Multirrótulo.	14
Figura 2	– Grafo.	17
Figura 3	– Coeficiente de agrupamento dos vértices de um grafo.	18
Figura 4	– Exemplos de imagens da base Scene.	21
Figura 5	– Grafo gerado utilizando o algoritmo Gerar Rede para o rótulo “montanha” da base Scene com $k = 10$	23
Figura 6	– Acurácia e acurácia de subconjunto da base Birds de acordo com variação de λ	28
Figura 7	– Acurácia e acurácia de subconjunto da base Emotions de acordo com variação de λ	28
Figura 8	– Acurácia da base Emotions de acordo com variação de δ	29
Figura 9	– Acurácia da base Emotions de acordo com variação de τ	29

Lista de tabelas

Tabela 1 – Propriedades das bases de dados.	25
Tabela 2 – Acurácia e Acurácia de Subconjunto de cada algoritmo em cada base de dados.	26
Tabela 3 – Valores de acurácia para C_y e M_y para cada base de dados.	27

Lista de algoritmos

1	GERAR REDE	22
2	CONECTARMAISPROXIMOS	22
3	TÉCNICA PROPOSTA.	24

Lista de abreviaturas e siglas

BR	Binary Relevance
CC	Classifier Chain
DT	Decision Tree
MLkNN	MultiLabel k-Nearest Neighbors
NB	Naive Bayes
SVM	Support Vector Machine
WWW	World Wide Web

Sumário

1	INTRODUÇÃO	11
2	REVISÃO BIBLIOGRÁFICA	13
2.1	Classificação Multirrótulo	13
2.1.1	Aprendizado Supervisionado Monorrótulo	13
2.1.2	Aprendizado Supervisionado Multirrótulo	13
2.1.3	Algoritmos de Classificação Multirrótulo	14
2.1.4	Medidas de avaliação para classificação multirrótulo	15
2.1.4.1	Perda de Hamming	15
2.1.4.2	Acurácia do subconjunto	15
2.1.4.3	Acurácia	15
2.1.5	Medidas de avaliação para bases de dados multirrótulo	16
2.1.5.1	Cardinalidade do rótulo	16
2.1.5.2	Densidade do rótulo	16
2.2	Redes Complexas	16
2.2.1	Medidas para Redes Complexas	17
2.2.1.1	Assortatividade	17
2.2.1.2	Coeficiente de Agrupamento	17
2.2.2	Trabalhos relacionados	18
3	DESENVOLVIMENTO	20
3.1	Bases de Dados para Classificação Multirrótulo	20
3.2	Construção da Rede	21
3.3	Classificação Multirrótulo usando Redes Complexas	23
4	RESULTADOS	25
4.1	Ambiente Experimental	25
4.2	Comparação entre técnicas convencionais	26
4.3	Combinação de técnicas convencionais e baseadas em redes complexas	26
4.4	Análise dos parâmetros de combinação dos classificadores	27
5	CONCLUSÃO	30
	REFERÊNCIAS	31

1 Introdução

Com o grande volume de dados disponível em formato digital, a extração de informação relevante desses dados se torna um problema. O *Aprendizado de Máquina* é um dos campos de estudo que busca resolver esse problema por meio de algoritmos capazes de “analisar” os dados e “aprender” com os mesmos, gerando modelos automáticos que auxiliam na detecção de padrões e consequentemente na descoberta de conhecimento em grandes bases de dados (CARNEIRO, 2016). Uma das principais categorias do aprendizado de máquina é o *Aprendizado Supervisionado*. O aprendizado supervisionado tem como objetivo induzir um modelo preditivo usando dados com informações previamente conhecidas, também denominadas como *rótulo*, *classe* ou *saída* (CARNEIRO, 2016). As tarefas de aprendizado supervisionado podem ser aplicadas em diferentes problemas, como, filtragem de spam, análise de crédito, diagnóstico médico, reconhecimento facial entre outros.

O aprendizado supervisionado se caracteriza por um processo de categorização conduzido por um especialista no domínio, o qual é capaz de mapear um conjunto de amostras iniciais para seu respectivo rótulo. Em tal cenário podemos treinar um modelo capaz de prever de forma automática o rótulo de novos objetos ainda não mapeados.

O tipo de especialização necessária para rotular os exemplos depende de cada domínio. O diagnóstico de um paciente, por exemplo, é atribuído por médicos, a categorização de emails ou imagens pessoais pode ser realizada pelo próprio usuário (CHERMAN, 2013). As tarefas de predição possuem diferentes nomes, os quais chamamos de *classificação* se os valores assumidos pelos rótulos são discretos e *regressão* quando assumem valores contínuos.

Na classificação tradicional, um objeto pode ser associado a uma única classe, o que pode ser incoerente em certos domínios, pois, em vários problemas reais o comportamento dos objetos costuma ser complexo e ter múltiplos significados semânticos simultaneamente (ZHANG; ZHOU, 2014). Por exemplo, pacientes podem ter múltiplas doenças, livros podem ser associados a múltiplos gêneros (STREICH; BUHMANN, 2009). No contexto musical, por exemplo, uma música que tem características do rock e do pop, não deve ser rotulada como pertencendo a um gênero ou outro, mas sim, a ambos.

Para lidar com esse problema, utilizamos algoritmos de *classificação multirrótulo*, os quais podem associar um objeto a um *conjunto* de classes. Consequentemente, a classificação multirrótulo torna o processo de classificação mais complexo não só computacionalmente, mas como um todo. Apesar de ser um tópico desafiador e de grande relevância, é ainda pouco explorado na literatura onde grande parte dos algoritmos são adaptações da classificação tradicional.

Nos últimos anos, o estudo de redes complexas chamou bastante atenção em diversas áreas do conhecimento por resolver uma grande variedade de problemas. Estas redes fornecem um conjunto de ferramentas para a representação e análise de dados por ser capaz de descrever a interação entre propriedades topológicas, funcionais e dinâmicas dos dados (CARNEIRO et al., 2018).

O objetivo deste trabalho é combinar os algoritmos já existentes de classificação multirrótulo com conceitos de redes complexas para que, além de atributos físicos, seja também considerada a estrutura topológica dos dados. Os objetivos específicos deste trabalho são:

- Investigar os principais algoritmos e medidas de classificação multirrótulo presentes na literatura.
- Aplicar os conceitos da teoria das redes complexas.
- Desenvolver uma técnica que combine os algoritmos de classificação multirrótulo tradicionais da literatura com medidas de redes complexas.
- Demonstrar que a teoria das redes complexas pode ser usada para melhorar os algoritmos convencionais de classificação multirrótulo.

O restante deste trabalho é dividido nos seguintes capítulos:

- No capítulo 2, são apresentados conceitos fundamentais para o entendimento e desenvolvimento do trabalho. Especificamente, são revisados conceitos básicos de classificação multirrótulo e redes complexas, além de apresentar medidas de desempenho para avaliar os algoritmos apresentados.
- No capítulo 3 é mostrado tudo que foi feito para o desenvolvimento do trabalho. Especificamente são apresentados as bases de dados utilizadas, quais os parâmetros dos algoritmos e modelagem do problema.
- No capítulo 4 são exibidos os resultados do trabalho. Especificamente é mostrada uma comparação entre as técnicas convencionais e posteriormente as mesmas com a combinação do classificador via medidas de rede. São descritas também características das bases de dados e quais valores dos parâmetros foram utilizados.
- No 5º e último capítulo, são feitas as considerações finais do trabalho.

2 Revisão Bibliográfica

2.1 Classificação Multirrótulo

2.1.1 Aprendizado Supervisionado Monorrótulo

No aprendizado supervisionado tradicional ou *monorrótulo*, utilizamos duas bases de dados, uma denominada *base de treino*, onde cada elemento já está associado a sua respectiva classe, e a outra denominada *base de teste*, onde os elementos ainda não foram associados a nenhuma classe.

A base de treino pode ser formalmente descrita da seguinte maneira. Considere uma base de treino com n elementos, $B_{treino} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \mid 1 \leq y_i \leq k; 1 \leq i \leq n\}$. O par (x_i, y_i) representa o i -ésimo objeto do problema, sendo $x_i = \{a_1, a_2, \dots, a_m\}$ um vetor que descreve os m atributos do objeto, e y_i sendo sua respectiva classe dentre as k possíveis do problema. A base de teste, por sua vez, pode ser definida da seguinte maneira. Considere uma base de teste com j elementos, $B_{teste} = \{(x_{n+1}, y?), (x_{n+2}, y?), \dots, (x_{n+j}, y?) \mid 1 \leq y_i \leq k; n+1 \leq i \leq n+j\}$. As bases possuem diferentes propósitos. O objetivo da base de treino é a criação de uma função $f: x \rightarrow y$, que chamamos também de classificador, que possa determinar os rótulos dos objetos da base de teste, que, por sua vez, tem como objetivo avaliar a habilidade preditiva do classificador gerado.

2.1.2 Aprendizado Supervisionado Multirrótulo

A diferença entre o aprendizado supervisionado tradicional e o multirrótulo, está na quantidade de rótulos que podem ser atribuídos a um objeto. Enquanto no problema tradicional o objeto é associado a um único rótulo, no contexto multirrótulo um objeto pode ser associado a um *conjunto* de rótulos. Formalmente, seja $Y = \{y_1, y_2, \dots, y_k\}$ um conjunto com todas as k possíveis classes e $\alpha = 2^Y$ representando todas possíveis combinações de rótulos do problema, o objetivo de um classificador multirrótulo é gerar uma função $h: x \rightarrow \alpha$.

A Fig. 1 representa a diferença entre classificação tradicional e classificação multirrótulo. Fig. 1a ilustra a classificação tradicional, onde uma música só pode ser do gênero pop ou rock, mas nunca os dois. Fig. 1b, ilustra a classificação multirrótulo, onde os elementos que ficaram divididos entre pop e rock, são então classificados em ambas as classes.

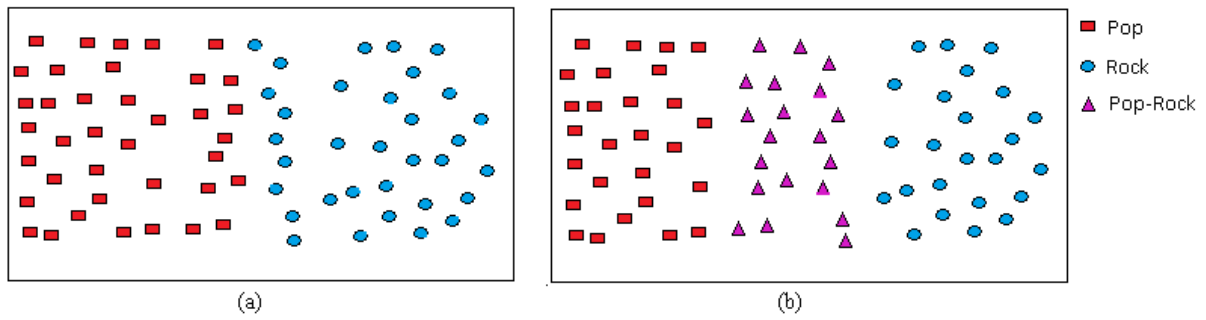


Figura 1 – (a) Classificação Tradicional. (b) Classificação Multirrótulo.

2.1.3 Algoritmos de Classificação Multirrótulo

Na literatura, métodos de classificação multirrótulo são normalmente divididos em dois grupos. *Métodos de transformação do problema* e *métodos de adaptação de algoritmos* (SPYROMITROS; TSOUMAKAS; VLAHAVAS, 2008).

No primeiro grupo, uma opção é quebrar o problema multirrótulo em um ou mais problemas monorrótulo, e então resolver individualmente cada instância de problema utilizando algoritmos tradicionais de classificação. Outra opção para o primeiro grupo é transformar o problema em uma tarefa de ranqueamento dos rótulos, onde o objetivo é computar o quão relevante é uma classe para um objeto em particular e então selecionar as classes que satisfazem um grau pretendido de relevância (MENCIA; FURNKRANZ, 2008). No segundo grupo, são abordados algoritmos específicos de classificação multirrótulo que tratam o problema diretamente e pertencem a classe de adaptação de algoritmos.

Para a classe de transformação de problema, um dos principais algoritmos é o *Binary Relevance* (BR). O algoritmo consiste em aprender um modelo binário monorrótulo para cada classe distinta da base de dados, gerando $|L|$ bases de dados da forma $\{y_i, \neg y_i\}$ onde o classificador terá que decidir se cada objeto pertence ou não a i -ésima classe do problema (TSOUMAKAS et al., 2009). Existe outro algoritmo desta classe que, diferente do BR, considera a dependência dos rótulos, considerando se um conjunto de rótulos nunca vai acontecer ou sempre vai acontecer, por exemplo. O *Classifier Chain* (CC) cria $|L|$ bases de dados, da mesma forma que o BR, entretando, a medida que um objeto é classificado em uma determinada classe, essa informação é repassada para as próximas instâncias a serem preditas expandido seu espaço de atributos, dessa forma, levando em consideração quais rótulos foram atribuídos nas instâncias anteriores.

Para a classe de adaptação de algoritmo, um exemplo é o *Multi-label k -Nearest Neighbor*, que é um dos algoritmos considerados estado da arte em classificação multirrótulo (ZHANG; ZHOU, 2014). Ele recebe esse nome pois é baseado no tradicional kNN. A idéia base é adaptar o kNN para lidar com o problema multirrótulo (ZHANG; ZHOU,

2007). Além de utilizar como parte do algoritmo o tradicional kNN, é utilizada *inferência bayesiana* para classificar os rótulos.

2.1.4 Medidas de avaliação para classificação multirrótulo

Nesta seção, serão abordadas as principais medidas de avaliação para algoritmos de classificação multirrótulo encontradas na literatura.

2.1.4.1 Perda de Hamming

Perda de Hamming ou *Hamming Loss* é a média do total de classes que foram incorretamente preditas para cada exemplo (SOKOLOVA; LAPALME, 2009). Seja D uma base de dados, Y_i o subconjunto correto de rótulos para o i -ésimo objeto e $Z_i = h(x_i)$,

$$HammingLoss(h, D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \oplus Z_i|}{k} \quad (2.1)$$

Para esta medida, os valores devem ser minimizados.

2.1.4.2 Acurácia do subconjunto

Acurácia do subconjunto ou *subset accuracy* é uma medida que considera a quantidade de exemplos que foram perfeitamente classificados, ou seja, só será levado em consideração quando o classificador acertar totalmente todos os rótulos previstos para um objeto. (TSOUMAKAS; VLAHAVAS, 2007).

$$SubsetAccuracy(h, D) = \frac{1}{|D|} \sum_{i=1}^{|D|} (Y_i = Z_i) \quad (2.2)$$

Para esta medida, os valores devem ser maximizados.

2.1.4.3 Acurácia

Acurácia ou *Accuracy* é uma medida que quantifica quantos rótulos foram previstos corretamente. Nesta medida, são considerados os acertos individuais em relação aos rótulos, ou seja, todos os rótulos que forem previstos corretamente serão considerados, independente de acertar todo o conjunto de rótulos do objeto ou não.

$$Accuracy(h) = \frac{1}{|D|} \frac{\sum_{i=1}^{|D|} |Y_i \cap h(x_i)|}{k} \quad (2.3)$$

Para esta medida, os valores devem ser maximizados.

2.1.5 Medidas de avaliação para bases de dados multirrótulo

As bases de dados multirrótulo podem variar muito de uma para outra. Em algumas bases a quantidade de rótulos de cada exemplo pode ser muito diferente em relação a outra base, e isso pode implicar na avaliação do algoritmo. Algumas medidas como *cardinalidade do rótulo* e *densidade do rótulo* podem dizer o “quão multirrótulo” é uma base (TSOUMAKAS; KATAKIS, 2007).

2.1.5.1 Cardinalidade do rótulo

Cardinalidade do rótulo ou *label cardinality* é uma medida que calcula a média da quantidade de rótulos dos exemplos de uma base. Sendo D o conjunto de dados e Y_i o conjunto de rótulos preditos para o i -ésimo objeto,

$$LabelCardinality(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} |Y_i| \quad (2.4)$$

.

2.1.5.2 Densidade do rótulo

Densidade de rótulo ou *label density* é uma medida que calcula a média da quantidade de rótulos dos exemplos de uma base, dividido pela quantidade de possíveis rótulos da base. Sendo k a quantidade de rótulos possíveis do conjunto de dados,

$$LabelDensity(D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i|}{k} \quad (2.5)$$

.

2.2 Redes Complexas

O estudo de redes complexas chamou bastante atenção nos últimos anos ao resolver problemas de sistemas sociais, biológicos e de comunicação (JEONG; NÉDA; BARABÁSI, 2003). As redes complexas podem ser entendidas como redes que possuem estrutura topológica não trivial, o que significa que seu viés de formação não segue um critério nem específico e nem totalmente aleatório (CARNEIRO, 2016).

Uma das maneiras mais comuns de representar uma rede é por meio de estruturas chamadas grafos. Um grafo pode ser representado como $G(V, E)$ sendo V um conjunto de objetos chamados de vértices e E sendo um par que representa uma relação entre dois objetos, denominado aresta. Com esta estrutura, podemos representar uma grande variedade de problemas do mundo real, como redes sociais, Internet e *World Wide Web* (WWW). Um exemplo do uso de grafos ou redes complexas é na representação de relações de amizade em redes sociais, onde cada pessoa é um vértice e as arestas representam uma

relação de amizade. Um outro exemplo é a representação da WWW, onde os vértices são documentos (páginas web) e as arestas são hiperlinks (URLs) (ALBERT; BARABÁSI, 2002). A Fig. 2 mostra um exemplo de grafo.

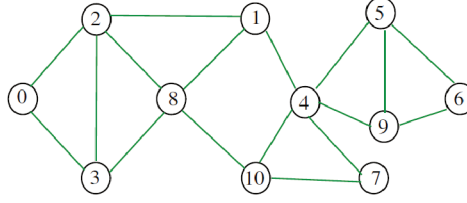


Figura 2 – Grafo.

2.2.1 Medidas para Redes Complexas

2.2.1.1 Assortatividade

Esta medida determina o quanto os vértices tendem a se conectar de maneira assortativa. A medida pode assumir valores entre $[-1, 1]$, de modo que valores positivos indicam que pares de vértices diretamente conectados estão mais suscetíveis a se comportar da mesma forma, enquanto valores negativos indicam maior probabilidade dos vértices conectados terem comportamentos distintos. (CARNEIRO, 2016). Seja L o número de arestas na rede e i_u, k_u o grau dos vértices i e k que compõem uma aresta u , a assortatividade da rede pode ser calculada por:

$$r = \frac{L^{-1} \sum_u i_u k_u - [L^{-1} \sum_u \frac{1}{2}(i_u + k_u)]^2}{L^{-1} \sum_u \frac{1}{2}(i_u^2 + k_u^2) - [L^{-1} \sum_u \frac{1}{2}(i_u + k_u)]^2} \quad (2.6)$$

2.2.1.2 Coeficiente de Agrupamento

De forma geral, esta medida quantifica o quanto os vértices tendem a agrupar-se. O coeficiente de agrupamento de um vértice mede o quão próximo ele está de formar um clique. Esta medida pode ser obtida por,

$$CC_i = \frac{|e_{us}|}{k_i(k_i - 1)} \quad (2.7)$$

onde $|e_{us}|$ representa quantos cliques de três vértices são formados a partir do vértice i , o que significa o número de conexões compartilhadas por vizinhos diretos de i . Para formar um clique, i precisa ser conectado a u , s e u também precisa estar conectado a s . k_i é o grau do vértice i .

O coeficiente de agrupamento da rede pode ser obtida por,

$$CC = \frac{1}{N} \sum_{i=1}^N CC_i \quad (2.8)$$

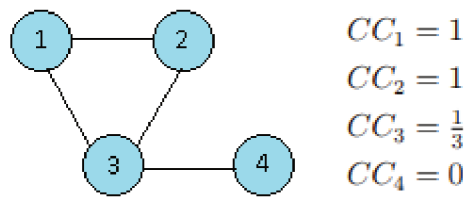


Figura 3 – Coeficiente de agrupamento dos vértices de um grafo.

A Fig. 3 apresenta o coeficiente de agrupamento para cada vértice de um grafo com coeficiente de agrupamento médio de 0.77. Os vértices 1 e 2 possuem ligações diretas entre todos os seus vizinhos enquanto o vértice 3 nem todos seus vizinhos possuem ligações diretas. O vértice 4 nem sequer possui mais de um vizinho. Vértices com grau ≤ 1 como por exemplo o vértice 4, não são considerados no cálculo para que seja evitada a divisão por zero.

2.2.2 Trabalhos relacionados

Em Zhang e Zhou (2014) é apresentada uma visão geral sobre classificação multirrótulo, trazendo conceitos, algoritmos e medidas de avaliação importantes para tal. O trabalho tem como objetivo trazer uma revisão sobre classificação multirrótulo abordando os principais algoritmos encontrados na literatura.

Apesar do trabalho proposto aqui ser um dos primeiros relacionados à classificação multirrótulo usando redes complexas, outros trabalhos para classificação *monorrótulo* contribuíram de forma significativa para o seu desenvolvimento, os quais são apresentados a seguir.

Em Carneiro (2016) são apresentados conceitos e medidas importantes sobre redes complexas e classificação de dados além de exibir uma compilação sobre técnicas de classificação envolvendo a teoria das redes complexas. O objetivo principal do trabalho é o desenvolvimento de métodos e heurísticas baseadas em teoria de redes complexas para a classificação de dados.

Carneiro et al. (2018) propõe uma estrutura de otimização de enxame de partículas que é responsável por construir uma rede a partir de um conjunto de dados baseados em vetores, otimizando uma função de qualidade guiada pela acurácia da classificação. O trabalho mostra com resultados em problemas artificiais e reais que tal abordagem apresenta melhores resultados em geral do que aqueles gerados pelos métodos clássicos de formação de redes.

Silva e Zhao (2012) apresenta uma ideia muito similar a dessa pesquisa. Neste artigo, é proposto um modelo de classificação híbrido que combina medidas de redes com-

plexas com algoritmos de classificação monorrótulo tradicionais. O objetivo do trabalho foi aprimorar a performance do classificador utilizando uma abordagem que além de analisar os atributos físicos dos dados também considera sua estrutura topológica. O trabalho exibiu melhora no desempenho da classificação monorrótulo trazendo uma nova ideia para classificação de dados.

O modelo de classificação usado nesta pesquisa teve como base o trabalho de Silva e Zhao (2012), com modificações relacionadas à normalização, atribuição de pesos, construção da rede, entre outras.

3 Desenvolvimento

A primeira etapa do trabalho consistiu basicamente na leitura de diversos artigos relacionados a teoria das redes complexas e classificação multirrótulo com o objetivo de adquirir base teórica suficiente para prosseguir com o trabalho. Com o conhecimento conquistado nesta etapa foi possível desenvolver uma habilidade crítica acerca de ambas áreas de estudo, conhecendo os principais conceitos, medidas de avaliação e algoritmos para classificação multirrótulo além de diversas aplicações e medidas de redes complexas, estas, que analisam importantes características da rede.

Em seguida foram implementadas e testadas medidas de redes complexas, sendo elas as medidas de assortatividade e coeficiente de agrupamento da rede, descritas na seção 2.2.1. Também foram implementadas e testadas medidas de avaliação de algoritmos de classificação multirrótulo, sendo elas as medidas de Hamming Loss, Subset Accuracy e Accuracy. Todas essas medidas de classificação multirrótulo foram listadas e descritas na seção 2.1.4.

A seguir são apresentadas outras etapas importantes para o desenvolvimento deste trabalho, desde a seleção e coleta das bases de dados utilizadas, passando pelo processo de construção da rede até a formulação do algoritmo de classificação multirrótulo usando redes complexas.

3.1 Bases de Dados para Classificação Multirrótulo

Para este trabalho foram selecionadas bases de dados de vários problemas de domínios distintos cujos dados estão representados na forma de vetores de atributos numéricos. As bases coletadas foram:

- Birds, é uma base de dados contendo a frequência gerada através de FFT (Fast Fourier Transform) de uma gravação de áudio do som de pássaros. O problema que essa base traz é o seguinte: dada a gravação de um áudio, dizer o conjunto de todas as espécies de pássaros presentes na gravação (BRIGGS et al., 2012).
- Emotions, esta base contém informação da frequência de músicas. A frequência extraída são do timbre e o ritmo das músicas (TROHIDIS et al., 2008), que também foram capturadas usando FFT. O problema dado pela base é o seguinte: dado o timbre e o ritmo da música, o que irá sentir quem ouvir a música? Os possíveis rótulos (sentimentos) disponíveis na base são: surpresa, felicidade, calma, continuar quieto, tristeza, raiva.

- Scene, esta é uma base de dados que possui 294 atributos descrevendo imagens. O problema dado pela base é: Dada uma imagem, o que é mostrado na imagem ? Exemplos de rótulos são praia, montanha, campo e pôr do sol (BOUTELL et al., 2004). Na Fig. 4 podemos ver exemplos de imagens da base, onde em Fig. 4a mostra uma imagem com os rótulos praia e urbano e a Fig. 4b uma imagem com os rótulos praia e montanha.



Figura 4 – Exemplos de imagens da base Scene.

- Yeast, esta é uma base de dados de levedura (fungos) com dados de expressão genética e perfis filogenéticos produzidos via *microarray de genes* (ELISSEEFF; WESTON, 2002). O problema dado pela base é: Dado a expressão genética e informações filogenéticas de uma levedura, quais são suas classes funcionais genéticas ? A base possui 14 rótulos distintos para classificação.

3.2 Construção da Rede

Para tornar possível o uso de redes complexas na solução, primeiro é preciso modelar o problema como uma rede. Para tal, é necessário gerar uma rede dos objetos de treinamento da base em que o classificador será treinado usando o Alg. 1, o qual apresenta a geração de uma rede k vizinhos mais próximos para cada classe da base de dados

correspondente.

Algoritmo 1: GERAR REDE

Entrada: B_{treino}, C, k

Saída: Grafo representando a classe C

```

1 início
2   Lista : Grafo[|Btreino|]
3   para cada  $u \in B_{treino}$  faça
4     se  $u \in C$  então
5       conectarMaisProximos(Grafo,  $u, k, C, B_{treino}$ )
6     fim
7   fim
8   Transformar o grafo em um novo não direcionado
9 fim
10 retorna Grafo

```

A função conectarMaisProximos analisa os k exemplos mais próximos de u utilizando uma medida de similaridade e o conecta apenas a outros exemplos que também estão associados a C . Neste trabalho a medida utilizada foi a distância Euclidiana. O Alg. 2 descreve a função.

Algoritmo 2: CONECTARMAISPROXIMOS

Entrada: *Grafo*, u, k, C, B_{treino}

```

1 início
2   Lista < par < real, inteiro > > : distancia
3   para cada  $v \in B_{treino}$  faça
4     se  $v \in C \wedge v \neq u$  então
5       distancia.inserir( $\{distanciaEuclidiana(B_{treino}[u], B_{treino}[v]), v\}$ )
6     fim
7   fim
8   ordenar(distancia)
9   para  $i \leftarrow 0$  até  $\min(k, |distancia[i]|) - 1$  faça
10    Grafo[ $u$ ].inserir(distancia[ $i$ ].second)
11  fim
12 fim

```

Assim que a rede é construída, é então aplicada as medidas de redes complexas no grafo. Antes que as medidas sejam aplicadas, o grafo direcionado é transformado em um novo grafo não direcionado. A Fig. 5 mostra um exemplo de grafo que foi gerado utilizando o Alg. 1. Para este trabalho as medidas aplicadas foram assortatividade e coeficiente de agrupamento, ambas descritas em 2.2.1.

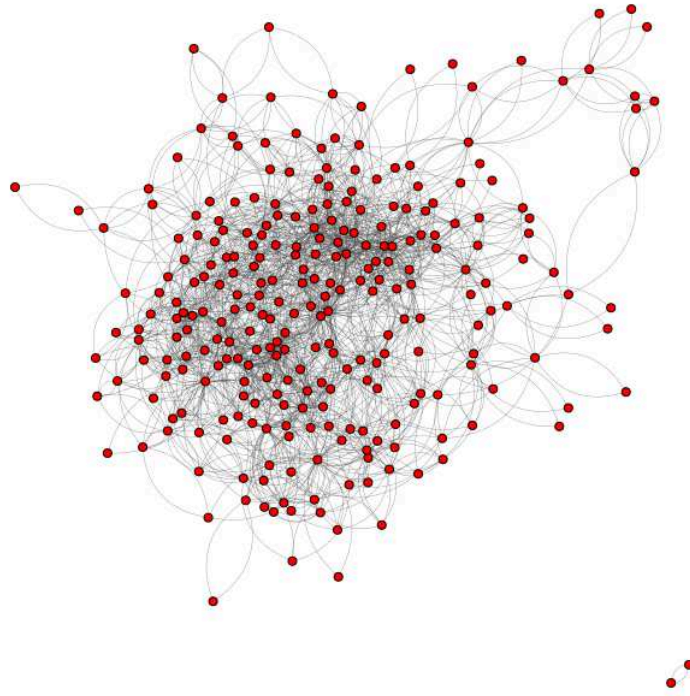


Figura 5 – Grafo gerado utilizando o algoritmo **Gerar Rede** para o rótulo “montanha” da base Scene com $k = 10$.

3.3 Classificação Multirrótulo usando Redes Complexas

Em classificação, a maioria dos algoritmos mono ou multirrótulo da literatura analisam apenas os atributos físicos dos dados ignorando características e relações importantes dos dados, como por exemplo, a formação de padrão. Uma das maneiras de capturar essas características é utilizando a teoria das redes complexas para modelar o problema. A ideia base do trabalho é combinar a classificação via medidas de rede com os algoritmos de classificação multirrótulo já existentes para que seja considerada a estrutura topológica dos dados. Uma proposta similar chamada de *framework híbrido* foi proposta por (SILVA; ZHAO, 2012) para classificação tradicional.

A partir dos cálculos das medidas de rede sobre o grafo gerado a partir da base de treino, a próxima etapa é inserir no grafo cada objeto de teste usando o Alg. 1. Entretanto, deve ser desconsiderada a verificação de classe do objeto de teste, visto que em teoria não a conhecemos. Observe que um objeto de teste pode ser inserido em grafos distintos, assim como os objetos de treino, visto que cada grafo representa uma classe e um mesmo objeto pode ter vizinhos de similaridade associados a diferentes classes.

Após a inserção do objeto de teste, é calculada a variação das medidas do grafo original com o novo grafo considerando o objeto de teste. No caso de muita variação a tendência é que o objeto não pertença à classe representada pelo grafo, do contrário existe uma chance maior do objeto ser associado àquela classe do grafo.

Após o cálculo da variação, o objeto teste é removido para que outro possa ser testado e o algoritmo se repete para todos os outros objetos de teste. A combinação das medidas de rede pode ser descrita da seguinte forma:

$$H_y^{(c)} = \sum_{u=1}^Z \delta(u) [1 - f_y^{(c)}(u)], \quad (3.1)$$

onde $\delta \in [0, 1]$, $\sum_u \delta_u = 1$ e representa um peso para cada medida de rede, e (u) representa as medidas de rede escolhidas para análise de conformidade de padrão. A função $f_y^{(c)}$ é definida como:

$$f_y^{(c)} = G_y^{(c)}(u) p^{(c)}, \quad (3.2)$$

no qual $G_y^{(c)}$ representa a variação que o objeto de teste y gerou na medida u no grafo e $p^{(c)}$ representa a proporção de objetos de treino que pertencem a C , uma estratégia para lidar com bases de dados desbalanceadas.

Combinadas as medidas de redes complexas é preciso finalizar combinando os resultados com os algoritmos próprios de classificação multirrótulo. Esta etapa pode ser formalmente descrita da seguinte maneira:

$$M_y^{(c)} = \lambda C_y^{(c)} + (1 - \lambda) H_y^{(c)} \quad (3.3)$$

onde $M_y^{(c)}$ é o valor gerado pela junção das probabilidades de um objeto y pertencer a classe c dadas por um classificador multirrótulo tradicional, denotada por $C_y^{(c)}$ e pela combinação multirrótulo das variações das medidas de rede, denotada por $H_y^{(c)}$.

Assim que é obtido M , para fazer a classificação é preciso definir um *threshold* que chamaremos de τ , onde, um objeto y será associado a uma classe c , se $M_y^c \geq \tau$, do contrário não. O Alg. 3 descreve um passo a passo da técnica proposta.

Algoritmo 3: TÉCNICA PROPOSTA.

Entrada: $k, B_{treino}, B_{teste}, \lambda, \delta, \tau$

- 1 Construir a rede da B_{treino} , como descrito no Alg. 1
 - 2 Calcular Eq. 2.6, 2.7 para os grafos gerados
 - 3 Inserir um objeto de teste y nos grafos de seus k vizinhos de treino mais próximos
 - 4 Calcular Eq. 3.3
 - 5 Classificar y em todas classes onde $M_y^c \geq \tau$
 - 6 Remover y dos grafos e repetir as etapas 3, 4, 5, 6 com o próximo objeto de teste
-

4 Resultados

Neste capítulo será feita uma análise de todos os resultados experimentais obtidos, fazendo uma comparação primeiramente entre as técnicas convencionais e depois, as mesmas técnicas com a combinação via medidas de rede. Será discutido também, os parâmetros, bases de dados e algoritmos utilizados.

4.1 Ambiente Experimental

O modelo de classificação desenvolvido nesta pesquisa foi aplicado sobre as quatro bases de dados descritas no capítulo anterior, estas, que foram divididas em treino e teste. A Tab. 1 exibe as características de cada base de dados utilizada no trabalho.

Tabela 1 – Propriedades das bases de dados.

Base	Domínio	Instâncias	Rótulos	Atributos	Cardinalidade	Densidade	Treino	Teste
Birds	Áudio	645	19	258	1.014	0.053	322	323
Emotions	Música	593	6	72	1.869	0.311	391	202
Scene	Imagem	2407	6	294	1.074	0.179	1211	1196
Yeast	Biologia	2417	14	103	4.237	0.303	1500	917

Para avaliar os resultados obtidos pela técnica proposta, serão analisadas as medidas de acurácia e acurácia de subconjunto apresentadas na seção 2.1.4. Os classificadores multirrótulo convencionais escolhidos para uso nas análises foram:

- MLkNN por ser uma das principais técnicas da literatura.
- Binary Relevance e Classifier Chain juntos ao SVM por ser tratar de combinações muito populares em classificação multirrótulo.
- Binary Relevance e Classifier Chain juntos ao Naive Bayes, por se tratar de um algoritmo simples e muito utilizado na literatura. O modelo escolhido do classificador ingênuo foi o Gaussiano por lidar melhor com objetos com atributos contínuos, como os das bases utilizadas neste trabalho.

Os parâmetros dos algoritmos foram mantidos o padrão. Sendo para o MLkNN o valor de $k = 10$, para o SVM o kernel = rbf e nenhum parâmetro para o Naive Bayes. Para a combinação das medidas de redes complexas com os algoritmos foram testadas todas combinações de λ , δ e τ considerando os valores $[0.0, 0.1, 0.2, \dots, 0.8, 0.9, 1.0]$.

A implementação dos classificadores multirrótulo convencionais usados neste trabalho pode ser encontrada em (Szymański; Kajdanowicz, 2017). O código da combinação

do classificador via medidas de rede com os algoritmos tradicionais foram implementados e testados nas linguagens C++ e Python. Os resultados obtidos sobre cada base será discutido nas próximas seções.

4.2 Comparação entre técnicas convencionais

Na Tab. 2 podemos observar que o algoritmo MLkNN obtêm um bom resultado de modo geral. O algoritmo se destaca principalmente nas últimas duas bases de dados observadas, onde, na base Scene, o valor da acurácia de subconjunto foi muito superior aos demais algoritmos analisados.

A estratégia do Binary Relevance e Classifier Chain utilizando como modelo binário o SVM obtiveram resultados muito próximos e conseguiram o melhor resultado na base Birds por uma pequena diferença em relação ao MLkNN. Mesmo obtendo um resultado satisfatório de acurácia na base Yeast, essas estratégias tiveram um péssimo desempenho na acurácia de subconjunto dessa base, acertando perfeitamente os rótulos de menos de 2% da base de teste. A técnica Classifier Chain junto ao Gaussian Naive Bayes obteve um resultado insatisfatório de acurácia de subconjunto na base Birds, um bom resultado de acurácia de subconjunto na base Scene e a melhor acurácia na base Emotions.

Tabela 2 – Acurácia e Acurácia de Subconjunto de cada algoritmo em cada base de dados.

Algoritmo	Emotions		Birds		Scene		Yeast	
	Acc.	S.Acc.	Acc.	S.Acc.	Acc.	S.Acc.	Acc.	S.Acc.
MLkNN	69.6	13.4	94.6	46.7	91.0	60.5	79.1	16.7
BR (SVM)	73.5	14.3	94.8	46.7	85.9	21.7	76.7	1.7
CC (SVM)	73.5	14.3	94.8	46.7	86.3	24.1	76.7	1.7
BR (GNB)	72.7	11.8	74.1	7.4	75.6	17.8	69.9	10.3
CC (GNB)	73.9	14.8	75.3	8.3	79.3	29.0	68.6	9.2

4.3 Combinação de técnicas convencionais e baseadas em redes complexas

A Tab. 3 mostra uma comparação da acurácia entre classificadores multirrótulos convencionais (C_y) e suas respectivas combinações com o classificador via medidas de redes complexas (M_y). Analisando a tabela, podemos observar que o classificador via medidas de rede contribuiu com a melhora dos algoritmos convencionais em pelo menos uma base de dados. Na base Birds, a técnica via medidas de rede apresentou um resultado melhor de maneira individual do que todos os algoritmos analisados, e, por esta razão, a acurácia se repetiu para todas as combinações de algoritmos. É importante observar, que, devido a

otimização dos parâmetros, caso o classificador via medidas de rede não consiga melhorar o desempenho de C_y o resultado retornado será exatamente igual a C_y .

Tabela 3 – Valores de acurácia para C_y e M_y para cada base de dados.

Algoritmo/Base	Emotions		Birds		Scene		Yeast	
	C_y	M_y	C_y	M_y	C_y	M_y	C_y	M_y
MLkNN	69.6	71.9	94.6	94.9	90.9	90.9	79.0	79.0
BR (SVM)	73.5	74.2	94.8	94.9	85.9	89.8	76.7	79.6
CC (SVM)	73.5	74.4	94.8	94.9	86.3	88.5	76.7	77.9
BR (GNB)	72.7	75.5	74.1	94.9	75.6	81.8	69.9	72.5
CC (GNB)	73.9	75.7	75.3	94.9	79.3	81.8	68.6	70.7

Existe um detalhe importante sobre os resultados das bases que precisa ser observado. Os algoritmos obtiveram melhores resultados em algumas bases mesmo estas sendo maiores e apresentando uma maior possibilidade de rótulos que outras. Como vimos na seção 2.1.5, as bases possuem duas medidas de avaliação muito importantes, a cardinalidade e densidade de rótulo. Os valores de tais medidas influenciam muito no resultado dos algoritmos, pois, elas podem dizer o “quão multirrótulo” ou “quão complicada” é uma base de dados. Observe que todos algoritmos obtiveram melhores resultados na base Scene em relação a Emotions, mesmo uma base sendo muito maior que a outra. Isto acontece pois estamos tratando de uma base “menos multirrótulo” e portanto mais fácil de realizar a classificação. A Tab. 1 mostra propriedades importantes das bases de dados, como a densidade e cardinalidade, fatores estes que influenciam diretamente na dificuldade da classificação.

4.4 Análise dos parâmetros de combinação dos classificadores

Nesta seção será discutida a contribuição de cada parâmetro do classificador via medidas de rede.

Na Fig. 6, foi definido um valor fixo para $\tau = 0.8$ e $\delta = 0.5$ e a contribuição (λ) do classificador BR (GNB) variou de $[0.0, 0.1, 0.2, \dots, 0.9, 1.0]$. Lembre-se que sempre que o λ do classificador tradicional aumenta o percentual de contribuição da classificação via medidas de rede diminui.

É possível observar que usando um λ elevado para o classificador multirrótulo tradicional os resultados ficaram piores, mostrando que, nesta base, o classificador multirrótulo via medidas de redes complexas de forma individual tem uma melhor performance em relação ao classificador multirrótulo tradicional usado. Outra observação é que as duas medidas de desempenho usadas se comportam de maneira similar em escalas diferentes. Nos resultados da Fig. 6, o grafo foi gerado utilizando $k = 13$.

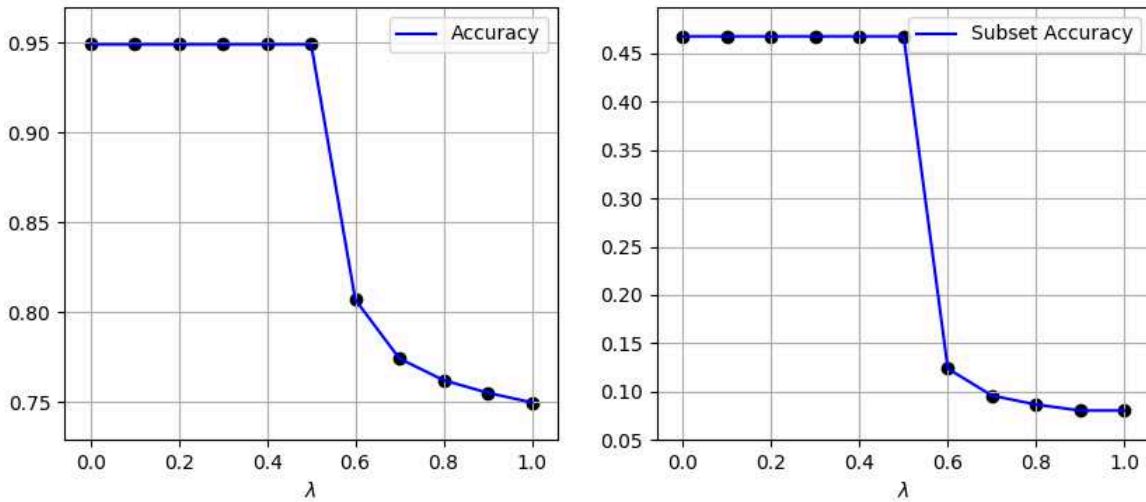


Figura 6 – Acurácia e acurácia de subconjunto da base Birds de acordo com variação de λ .

Na Fig. 7, foi definido um valor fixo para $\tau = 0.6$ e $\delta = 0.5$ e a contribuição (λ) do classificador BR (GNB) variou de $[0.0, 0.1, 0.2, \dots, 0.9, 1.0]$. É possível observar que usando apenas o classificador multirrótulo tradicional obtivemos ≈ 0.725 de acurácia e ≈ 0.16 de acurácia de subconjunto. Utilizando 0.6 do classificador multirrótulo tradicional e 0.4 da classificação via medidas de rede chegamos a ≈ 0.75 na acurácia e ≈ 0.21 de acurácia do subconjunto. Diferente do resultado mostrado na Fig. 6, o classificador multirrótulo tradicional contribuiu de maneira mais efetiva para obter o melhor resultado.

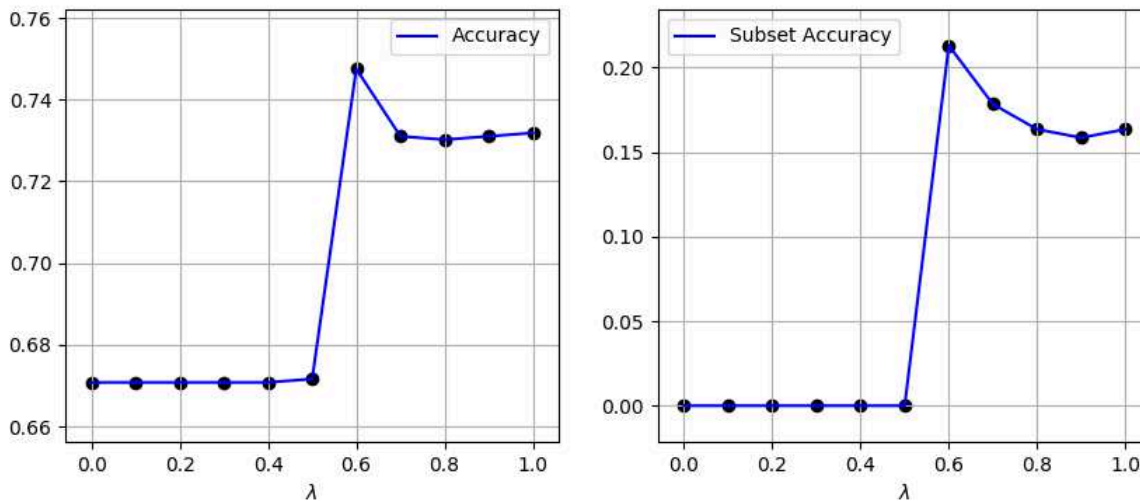


Figura 7 – Acurácia e acurácia de subconjunto da base Emotions de acordo com variação de λ .

Na Fig. 8, foi selecionado a melhor contribuição λ observando os valores de acurácia da Fig. 7 e fixado o mesmo τ . Podemos ver que o valor de δ para as medidas não foi muito

significativo. Isso aconteceu pois ambas as medidas utilizadas (assortatividade e coeficiente de agrupamento) apresentaram variações muito similares.

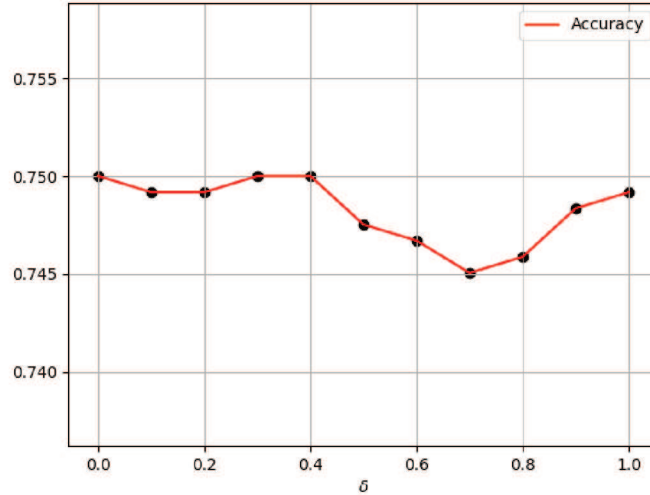


Figura 8 – Acurácia da base Emotions de acordo com variação de δ .

Na Fig. 9, foi selecionado o melhor valor λ obtido inicialmente na Fig. 7, o melhor δ obtido na Fig. 8 e o τ foi variado. É possível notar que o valor de τ para decisão de escolha é significativo nos resultados, como mostra a variação da acurácia. Para os resultados dos gráficos da base Emotions, o grafo foi gerado utilizando $k = 7$.

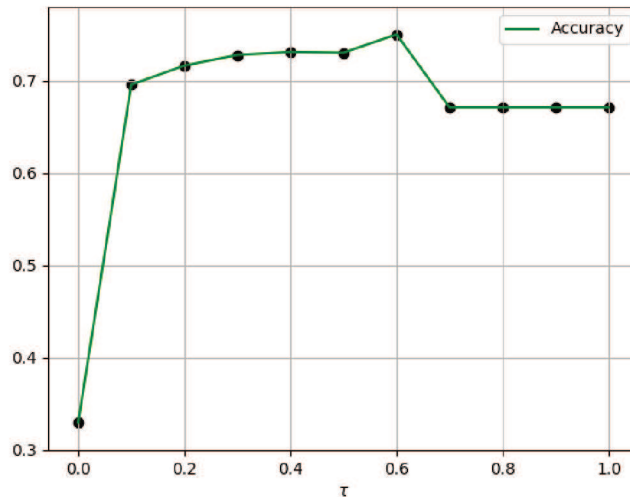


Figura 9 – Acurácia da base Emotions de acordo com variação de τ .

5 Conclusão

Como foi mostrado nos resultados, a utilização de redes complexas combinada a classificadores multirrótulo possibilita a melhora da taxa de acertos dos classificadores convencionais de maneira individual. Em nossos experimentos, obtivemos resultados consideráveis como vimos na seção de resultados onde, em alguns algoritmos, foi possível na base Birds melhorar a acurácia de 74.1% para 94.9% e na base Scene de 75.6% para 81.8%. Este é um resultado promissor que nos permite considerar mais investigações sobre o uso de redes complexas para a classificação multirrótulo. Apesar do modelo proposto aqui poder ser muito melhorado e explorado, o objetivo base de apresentar uma nova abordagem para classificação multirrótulo que considera não só os atributos físicos, mas também a estrutura topológica dos dados foi alcançado.

Para trabalhos futuros, serão consideradas novas medidas de avaliação, novas bases de dados, novas técnicas para otimização de parâmetros, como por exemplo, o *GridSearch*, afim de refinar a comparação entre os algoritmos, já que foram usados os valores padrão para cada parâmetro. Além do mais, serão investigadas novas abordagens para tratar o problema. Uma das abordagens que será discutida, se dá pelo uso de detecção de comunidades, onde será possível combinar e usar diferentes medidas em relação as que foram utilizadas neste trabalho.

Referências

- ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics*, APS, v. 74, n. 1, p. 47, 2002. Citado na página 17.
- BOUTELL, M. R. et al. Learning multi-label scene classification. *Pattern recognition*, Elsevier, v. 37, n. 9, p. 1757–1771, 2004. Citado na página 21.
- BRIGGS, F. et al. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, ASA, v. 131, n. 6, p. 4640–4650, 2012. Citado na página 20.
- CARNEIRO, M. G. *Redes complexas para classificação de dados via conformidade de padrão, caracterização de importância e otimização estrutural*. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado 4 vezes nas páginas 11, 16, 17 e 18.
- CARNEIRO, M. G. et al. Particle swarm optimization for network-based data classification. *Neural Networks*, Elsevier, 2018. Citado 2 vezes nas páginas 12 e 18.
- CHERMAN, E. A. *Aprendizado de máquina multirrótulo: explorando a dependência de rótulos e o aprendizado ativo*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado na página 11.
- ELISSEEFF, A.; WESTON, J. A kernel method for multi-labelled classification. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2002. p. 681–687. Citado na página 21.
- JEONG, H.; NÉDA, Z.; BARABÁSI, A.-L. Measuring preferential attachment in evolving networks. *EPL (Europhysics Letters)*, IOP Publishing, v. 61, n. 4, p. 567, 2003. Citado na página 16.
- MENCIA, E. L.; FURNKRANZ, J. Pairwise learning of multilabel classifications with perceptrons. In: IEEE. *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*. [S.l.], 2008. p. 2899–2906. Citado na página 14.
- SILVA, T. C.; ZHAO, L. Network-based high level data classification. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, v. 23, n. 6, p. 954–970, 2012. Citado 3 vezes nas páginas 18, 19 e 23.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, Elsevier, v. 45, n. 4, p. 427–437, 2009. Citado na página 15.
- SPYROMITROS, E.; TSOU MAKAS, G.; VLAHAVAS, I. An empirical study of lazy multilabel classification algorithms. In: SPRINGER. *Hellenic conference on artificial intelligence*. [S.l.], 2008. p. 401–406. Citado na página 14.
- STREICH, A. P.; BUHMANN, J. M. Ignoring co-occurring sources in learning from multi-labeled data leads to model mismatch. *Learning from Multi-label Data at ECML PKDD 2009*, p. 85, 2009. Citado na página 11.

Szymański, P.; Kajdanowicz, T. A scikit-based Python environment for performing multi-label classification. *ArXiv e-prints*, fev. 2017. Citado na página 25.

TROHIDIS, K. et al. Multi-label classification of music into emotions. In: *ISMIR*. [S.l.: s.n.], 2008. v. 8, p. 325–330. Citado na página 20.

TSOUMAKAS, G. et al. Correlation-based pruning of stacked binary relevance models for multi-label learning. In: *Proceedings of the 1st international workshop on learning from multi-label data*. [S.l.: s.n.], 2009. p. 101–116. Citado na página 14.

TSOUMAKAS, G.; KATAKIS, I. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining (IJDWM)*, IGI Global, v. 3, n. 3, p. 1–13, 2007. Citado na página 16.

TSOUMAKAS, G.; VLAHAVAS, I. Random k-labelsets: An ensemble method for multilabel classification. In: SPRINGER. *European conference on machine learning*. [S.l.], 2007. p. 406–417. Citado na página 15.

ZHANG, M.-L.; ZHOU, Z.-H. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, Elsevier, v. 40, n. 7, p. 2038–2048, 2007. Citado na página 15.

ZHANG, M.-L.; ZHOU, Z.-H. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, IEEE, v. 26, n. 8, p. 1819–1837, 2014. Citado 3 vezes nas páginas 11, 14 e 18.