

# O uso seletivo de classificadores binários na solução de problemas multirrótulos

Adriano Rivolli\*<sup>†</sup> e André C. P. L. F. de Carvalho\*

\*Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (ICMC-USP)

Av. Trabalhador São-carlense, 400, São Carlos, São Paulo, Brazil

<sup>†</sup>Universidade Tecnológica Federal do Paraná (UTFPR)

Av. Alberto Carazzai, 1640, Cornélio Procópio, Paraná, Brazil

**Resumo**—Algumas tarefas de classificação permitem que exemplos pertençam a mais de uma classe simultaneamente, uma delas é chamada de classificação multirrótulo. Uma forma simples e eficiente de solucionar problemas desta natureza consiste em transformá-los em vários problemas binários e tratá-los independentemente. Em geral, o mesmo classificador base é usado para induzir os diversos modelos, sem considerar seu viés e as particularidades de cada conjunto binário. Todavia, nesse estudo, investigamos a hipótese de que utilizar o classificador adequado para cada conjunto binário melhora a classificação multirrótulo. Utilizando o método de transformação *Binary Relevance*, foi adotada uma estratégia de meta-aprendizado para recomendar o classificador adequado para cada subproblema. Os resultados experimentais validam a hipótese investigada e mostram o potencial da abordagem utilizada. Além disso, a estratégia proposta é genérica, de modo que, pode ser aplicada em outros problemas de transformação multirrótulo.

**Keywords**—Classificação Multirrótulo, Meta-aprendizado, Métodos de transformação multirrótulo, *Binary Relevance*

## I. INTRODUÇÃO

Classificação multirrótulo é uma linha de pesquisa em expansão na área de aprendizado de máquina [1]. Neste tipo de tarefa, um classificador prediz um ou mais rótulos para cada instância classificada, o que torna a tarefa, de um modo geral, mais complexa do que a classificação de um único rótulo. Os principais desafios da classificação multirrótulo podem ser resumidos em questões como suportar adequadamente o número exponencial das combinações dos rótulos e a dependência entre eles, e também manipular dados desbalanceados e com alta dimensionalidade [2].

Uma forma simples e de baixo custo computacional, de solucionar um problema multirrótulo é transformá-lo em vários subproblemas de classificação binário, resolver separadamente cada um destes e posteriormente unificar os resultados obtidos [3]. É comum utilizar um único algoritmo de classificação na indução dos modelos de cada subproblema, sem no entanto, considerar as suas particularidades. Por exemplo, em [4] foram utilizados árvores de decisão e em [5] foram utilizados *Support Vector Machines* (SVM), como classificadores base.

Ao relevar que o viés de cada algoritmo o torna mais ou menos apropriado para um problema específico, um especialista tende a escolher pelo classificador que lhe parece ser a melhor opção. Todavia, no contexto multirrótulo, no qual o problema original é decomposto em diversos novos problemas mais simples, as escolhas são feitas de modo generalizado, com base no todo. A partir desta perspectiva, a hipótese

investigada neste estudo consiste de que a escolha seletiva dos classificadores binários, em detrimento do uso de um único classificador, leva a uma melhora nos resultados do aprendizado multirrótulo.

Recomendar de antemão o algoritmo de aprendizado mais adequado para um problema é uma tarefa não trivial, que exige entre outras coisas, conhecimento do domínio e dos algoritmos em questão. Considerando ainda, que um problema multirrótulo pode ser convertido em centenas de subproblemas binários, a recomendação dos classificadores é também um dos desafios investigados nesta pesquisa. Neste aspecto, a solução investigada é uma abordagem baseada em meta-aprendizado, que por meio da tarefa de classificação, torna possível a recomendação de algoritmos, a partir de experiências prévias [6].

As contribuições deste estudo consistem principalmente na validação da hipótese inicial, de que a escolha seletiva dos classificadores binários melhora o resultado da predição multirrótulo, e na proposição de estratégias de seleção e recomendação de algoritmos neste contexto. Além disso, os resultados experimentais mostram o potencial da proposta investigada, uma vez que novas possibilidades de extensão e melhorias foram identificadas.

O restante do artigo está organizado da seguinte maneira: na Sessão II os principais conceitos multirrótulos são apresentados; na Sessão III são abordados assuntos referentes ao meta-aprendizado, juntamente com a proposta do estudo e os trabalhos relacionados; na Sessão IV são descritos os procedimentos experimentais adotados nesta pesquisa; por fim, na Sessão V os resultados obtidos são apresentados e discutidos e na Sessão VI as considerações finais e trabalhos futuros são destacados.

## II. CLASSIFICAÇÃO MULTIRRÓTULO

Classificação multirrótulo é uma tarefa de aprendizado supervisionado que busca encontrar uma função objetivo  $h(x) \rightarrow Y$ , a partir de um conjunto de dados de treinamento, com o propósito de predizer um ou mais rótulos para novos exemplos. Assim, seja  $D$  um conjunto de instâncias rotuladas  $E$ , tal que  $D = \{E_1, E_2, \dots\}$ . Todas as instâncias rotuladas  $E_i = (x_i, Y_i)$  são compostas por um conjunto de atributos  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ , que descreve a posição deste exemplo no espaço  $\mathbb{R}^m$ , e  $Y_i \subseteq \{\lambda_1, \lambda_2, \dots, \lambda_q\}$ , o qual é um subconjunto de todos os  $q$  distintos rótulos  $\lambda$ , aqui chamado de classe.

A função  $h$  é a generalização de um modelo que pode ser obtido de três formas distintas [5]:

- *Transformação de problemas*: consiste na conversão de um problema multirrótulo para um conjunto de subproblemas binários ou multi-classe. Esta solução também é chamada de *algoritmo independente* [7] porque permite que algoritmos tradicionais de classificação possam ser reutilizados na solução de problemas multirrótulos;
- *Adaptação de algoritmos*: oferece suporte intrínseco à predição de múltiplos rótulos, por meio da modificação dos algoritmos tradicionais ou proposição de novos, por isto é também chamado de *dependente de algoritmo*, pois, utilizam algoritmos específicos para classificação multirrótulo;
- *Comitês*: combinam um ou mais algoritmos e/ou modelos para realizar predições multirrótulos, podendo também combinar ambas as abordagens anteriores.

Estas estratégias possuem diversos métodos/algoritmos que permitem predizer a classe de um exemplo com múltiplos rótulos. *Binary Relevance* [8], embora simples, é um método de transformação que tem sido utilizado como referência para comparação e extensão de novos métodos [9], [10], [11]. Seu uso tem levado à obtenção de resultados competitivos, em relação aos demais métodos [5], por isso, a sua escolha como método base neste trabalho.

Utilizando a estratégia um-contra-todos, cada rótulo é mapeado para um conjunto de dados que compartilha os mesmos atributos preditivos do conjunto original e possui um atributo alvo binário, indicando se os exemplos possuem ou não o rótulo em questão. Com isso, um modelo de classificação binária é induzido para predizer cada rótulo do problema original e, no contexto multirrótulo, um exemplo é classificado com todos os rótulos preditos como positivos. É comum, nesse tipo de solução, utilizar o mesmo classificador base para a indução de todos os modelos.

A avaliação de predições multirrótulos é realizada por meio de medidas específicas, como apresentadas em [7]. As métricas apropriadas para avaliar a qualidade das classes preditas são organizadas em dois grupos:

- *baseada nos exemplos*: as medidas originais de acurácia, precisão e revocação foram adaptadas para avaliação multirrótulo e novas medidas como acurácia exata e *Hamming Loss* são calculadas para cada exemplo separadamente e depois obtida a média de todo o conjunto;
- *baseadas nos rótulos*: as medidas de precisão, revocação, medida F1 e *Area Under Curve* (AUC) são empregadas na forma original, sem adaptações em suas fórmulas, apenas na composição dos valores da matriz de confusão (MC). Nos casos em que é utilizada uma MC para cada rótulo, a medida é denominada *macro* e é obtida pela média dos valores de cada rótulo. Por outro lado, quando a medida é calculada utilizando uma única MC, composta da soma dos valores de todos os rótulos, a medida é chamada *micro* e é obtida pela aplicação da fórmula.

Esta seção não tem por objetivo cobrir toda a literatura referente ao aprendizado multirrótulo, apenas contextualizar os assuntos abordados nesta pesquisa. Uma revisão mais detalhada sobre o tema pode ser encontrada em [2], [3], [7], [12].

### III. META-APRENDIZADO

O uso de meta-aprendizado é utilizado nesta pesquisa como abordagem para, em um nível meta, predizer quais classificadores binários utilizar na solução de um problema multirrótulo. Devido ao viés dos algoritmos de aprendizado de máquina, algumas escolhas são feitas durante a indução do modelo de classificação que restringem o espaço de busca do algoritmo e privilegia algumas situações em detrimento de outras. Com isso, não existe uma solução única que resolva todos os problemas de classificação da melhor forma possível [13].

Embora um especialista em aprendizado de máquina possa intuitivamente ou empiricamente fazer boas escolhas para um problema específico, as vezes isso se torna impraticável ou exige um alto custo computacional. Os métodos de transformação de problemas correspondem a um destes cenários, pois um único conjunto de dados multirrótulo pode ser decomposto em centenas de problemas binários. Para estes casos, acreditamos que o uso de meta-aprendizado pode ser mais adequado.

O meta-aprendizado destaca-se das estratégias convencionais de aprendizado de máquina ao permitir que possa ocorrer adaptações no processo [14]. Isso ocorre, devido a experiências previamente obtidas por meio da aplicação de algoritmos distintos em vários conjuntos de dados. Em um nível base são extraídas características dos conjuntos de dados, chamadas de meta-características [15], e estas são mapeadas para as técnicas investigadas que obtiveram o melhor desempenho, para algum critério de qualidade. O resultado obtido é chamado de *meta-base*.

No nível meta, são realizadas predições referentes a quais algoritmos utilizar para induzir modelos para novos conjuntos de dados, utilizando para isso o conhecimento armazenado na meta-base. Em geral, a adoção de meta-aprendizado acontece como parte de uma solução mais específica, como ocorre neste estudo.

No restante da seção, a estratégia desenvolvida é apresentada, seguida de uma breve análise do uso de meta-aprendizado no contexto multirrótulo, ressaltando as principais diferenças com a abordagem proposta.

#### A. Estratégia proposta

Ajustando a abordagem utilizada em [4] para o nosso contexto, o uso de meta-aprendizado acontece em seis etapas:

- 1) Decompor o problema multirrótulo original em problemas binários;
- 2) Induzir e avaliar um modelo para cada tarefa de classificação binária e conjunto de dados investigados;
- 3) Extrair as meta-características de cada um dos conjuntos de dados binários;
- 4) Criar uma meta-base usando as meta-características e o método mais apropriado para cada exemplo;
- 5) Induzir um meta-classificador utilizando a meta-base;

- 6) Usar o meta-classificador, obtido na etapa anterior, para recomendar o algoritmo adequado para um novo caso.

Inicialmente, foram considerados o uso de quatro classificadores distintos: *Support Vector Machine* (SVM); *Random Forest* (RF), *Naive Bayes* (NB); e, *k-Nearest-Neighbors* (kNN), no entanto, devido ao baixo número de vezes que NB obteve o melhor desempenho, para os critérios de avaliação utilizados, optamos por descartá-lo dos experimentos, tornando o sistema de recomendação mais simples.

A partir dos conjuntos multirrótulos, as bases binárias são geradas utilizando o método *Binary Relevance*, e são caracterizadas pelos atributos apresentados na Tabela I. Estas meta-características consistem em informações representativas dos conjuntos binários e compreendem: *medidas simples*, como número de atributos e a dimensionalidade da matriz de dados; *medidas estatísticas* do conjunto, como obliquidade e correlação dos atributos; e, *medidas de recuperação de informação*, como ruído e entropia dos dados. O uso destas medidas, possibilita a representação única de cada conjunto de dados sem exigir, em contrapartida, elevados custos computacionais. Os detalhes sobre estas características estão disponíveis em [15].

A meta-base gerada possui como exemplos (linhas) cada uma das bases binárias e como atributos preditivos (colunas) as meta-características extraídas. O atributo alvo é o nome do classificador que obteve o melhor desempenho preditivo, dada uma medida de avaliação. Foram utilizadas três métricas de avaliação, resultando conseqüentemente em três meta-bases, distintas entre si apenas pelo valor do atributo alvo. As medidas utilizadas foram acurácia, acurácia balanceada e medida F1, dando origem as bases denominadas *mtb-Acc*, *mtb-BalAcc* e *mtb-F1* respectivamente. Dada a semelhança dos resultados obtidos entre as duas últimas, alguns testes foram realizados com apenas uma delas, a *mtb-F1*, que foi melhor para um número maior de casos.

A partir da meta-base, um modelo de classificação é induzido para prever o algoritmo mais adequado para novos problemas binários. Assim, diante de um problema multirrótulo, aplica-se o método de transformação gerando os conjuntos binários; se extrai as características destes conjuntos; classifica cada um deles utilizando um dos meta-classificadores; utilizando o resultado da recomendação, dá-se continuidade ao processo de solução da forma tradicional.

### B. Trabalhos relacionados

O uso de meta-aprendizado na predição de algoritmos é uma abordagem frequentemente encontrada na literatura [4], [6], [16]. Entretanto, há poucos trabalhos que relacionam o uso de meta-aprendizado no cenário multirrótulo.

Em Chekina et al. [4], meta-aprendizado foi empregado para a recomendação de um método particular de classificação multirrótulo. Quando o método recomendado consiste em uma solução de transformação como o *Binary Relevance*, um único classificador base é utilizado para solucionar todos os subproblemas gerados. Visando otimizar as diversas métricas de avaliação, foram construídos meta-bases para cada medida multirrótulo e com isso, meta-modelos foram induzidos para minimizar/maximizar cada uma delas.

Tabela I. META-CARACTERÍSTICAS UTILIZADAS NA GERAÇÃO DA META-BASE

Atributo	Descrição
Atr	Número de classes
Spl	Número de exemplos de treinamento
Dim	Número de exemplos dividido pelo número de atributos
NumRate	Proporção de atributos numéricos
NomRate	Proporção de atributos nominais do conjunto de dados
Lfq	Frequência relativa dos valores da classe positiva
Sks, SksP	Obliquidade do conjunto de dados (original e normalizado)
Kts, KtsP	<i>Kurtose</i> do conjunto de dados (original e normalizado)
SymMin, SymMax	Número mínimo e máximo de valores simbólicos do conjunto
SymMean, SymSd	Média e desvio padrão de valores simbólicos do conjunto
SymSum	Total de valores simbólicos do conjunto
AbsC	Correlação entre os atributos
CanC	Correlação canônica entre atributos e classes
ClEnt, NClEnt	Entropia do conjunto de dados (original e normalizada)
AtrEnt	Média da entropia dos atributos do conjunto de dados
JEnt	Entropia dos atributos combinadas par a par com a classe
MutInf	Medida de informação mútua entre atributos e classes
EAttr	Relação entre as medidas de entropia e informação mútua
NoiSig	Nível de ruído do conjunto de dados

Um modo distinto do uso de meta-aprendizado na solução de problemas multirrótulo é encontrado no método 2BR, também chamado de MetaBR (MBR) [10]. Esta abordagem consiste na execução do método *Binary Relevance* duas vezes, utilizando o resultado da primeira predição de cada rótulo como atributo da meta-base, que é utilizada na segunda predição. Esta abordagem explora o relacionamento entre os rótulos. porém neste caso também são utilizados um único classificador base para todos os subproblemas. Outras extensões e variações do método 2BR são citados em [2]. Esta estratégia difere completamente da proposta deste trabalho, no entanto, sua citação ilustra outras formas de combinar meta-aprendizado e multirrótulo.

Quanto a escolha do classificador base em problemas de transformação multirrótulo, não há um consenso geral, porém SVM [5], [9] e Árvores de Decisão [4], [17] têm sido frequentemente adotados. Experimentos comparativos com classificadores distintos foram realizados em [3], [11], no primeiro caso a comparação é referente ao desempenho dos classificadores base, enquanto que no segundo, o objetivo é comparar métodos de transformação.

## IV. MATERIAIS E MÉTODOS

Nesta seção é descrita a metodologia seguida neste estudo. Inicialmente, são apresentados os conjuntos multirrótulos utilizados, seguidos pela descrição das ferramentas e parâmetros dos algoritmos. Por fim, a metodologia de condução dos experimentos e avaliação dos resultados são descritas.

### A. Conjunto de dados

Foram utilizados treze conjuntos de dados multirrótulos, disponíveis no *web site* do projeto Mulan<sup>1</sup> [18]. Estes são provenientes de diferentes domínios de aplicação e possuem características diversificadas como pode ser observado na Tabela II. As colunas *Treino* e *Teste* contém o número de instâncias de treinamento e de testes respectivamente, as colunas *Atr* e *Rótulos* descrevem o número de atributos e rótulos distintos de cada conjunto. Por fim, as colunas *LCard* e *labelsets* são duas medidas multirrótulos que descrevem respectivamente a média

<sup>1</sup><http://mulan.sourceforge.net>

do número de rótulos por instância, chamada de cardinalidade, e o número de combinações distintas entre os rótulos.

Com exceção do conjunto *CAL500*, todos os demais possuem as instâncias originalmente separadas em treinamento e teste. Para dividi-lo, foi adotada a solução utilizada em [5], destinando de forma aleatória 2/3 dos exemplos para treinamento e 1/3 para teste. Para os demais conjuntos, foram mantidas as divisões originais.

### B. Configuração dos experimentos

Os experimentos foram divididos em dois níveis: *base* e *meta*. Em ambos, foram utilizadas a linguagem R [19], juntamente com a biblioteca *MLDR* [20] que oferece o suporte básico para manipulação e avaliação de dados multirrótulo.

No nível base ocorre a decomposição do conjunto multirrótulo original em conjuntos binários, utilizando o método *Binary Relevance*; a extração das meta-características destes conjuntos; a execução dos classificadores SVM, RF e kNN para cada um deles; e por fim, a indução dos meta-classificadores, no qual optou-se pelo uso do SVM.

Foram utilizadas as implementações disponíveis nas bibliotecas *e1071*, *randomForest* e *FNN* para os algoritmos SVM, RF e kNN respectivamente. Com relação aos parâmetros, para o SVM<sup>2</sup> e RF foram utilizados o padrão definido pelas bibliotecas, cujos valores são  $kernel = radial$ ,  $cost = 1$  e  $gamma = 1/numero\_de\_atributos$  para o SVM, e  $ntree = 500$  para o RF. Para o kNN, foram realizados alguns testes preliminares para a escolha do valor de  $k$ , que foi definido em  $k = 3$ .

Por sua vez, o nível meta consiste na recomendação dos classificadores para os conjuntos binários; a utilização destas recomendações para geração do modelo de classificação de cada rótulo e as previsões dos dados de teste; e finalmente, a junção das previsões para compor o resultado final multirrótulo.

As avaliações de desempenho dos classificadores binários e do meta-aprendiz utilizaram a biblioteca *caret*. A avaliação da classificação multirrótulo utilizou a biblioteca *MLDR*.

### C. Metodologia de avaliação

Embora o objetivo principal da solução proposta seja obter melhores previsões multirrótulos, avaliar o meta-classificador é necessário para verificar a viabilidade da estratégia. Deste modo, as avaliações foram realizadas sob duas perspectivas: a qualidade do meta-classificador e a qualidade do resultado multirrótulo obtido.

Para avaliação do meta-classificador, os dados foram particionados segundo a estratégia *leave-one-out* por conjunto multirrótulo, no qual em cada iteração, exemplos de onze conjuntos são utilizados para treinamento, enquanto que, os exemplos dos outros dois são utilizados para validação e teste, separadamente. Para ilustrar este cenário, assumamos que  $\{A, B, C, D\}$  são conjuntos multirrótulos distintos e que possuem  $\{5, 15, 20, 10\}$  rótulos respectivamente. Neste caso, a meta-base gerada possui ao todo 50 exemplos. Para avaliar as

Tabela II. CONJUNTOS DE DADOS MULTIRRÓTULOS UTILIZADOS

Nome	Treino	Teste	Atr	Rótulos	LCard	Labelsets
bibtex	4880	2515	1995	159	2.38	2058
birds	322	323	279	19	1.06	89
CAL500	335	167	242	174	26.05	335
corel5k	4500	500	873	374	3.52	2925
emotions	391	202	78	6	1.81	26
enron	1123	579	1054	53	3.38	545
flags	129	65	26	7	3.42	45
genbase	463	199	1213	27	1.26	29
mediamill	30993	12914	221	101	4.36	4913
medical	333	645	1494	45	1.25	61
scene	1211	1196	300	6	1.06	14
tmc2007-500	21519	7077	522	22	2.22	1172
yeast	1500	917	117	14	4.22	164

previsões referentes aos exemplos do conjunto  $A$ , são utilizados os 35 exemplos de treinamento provenientes dos conjuntos  $\{B, C\}$  e os 10 exemplos do conjunto  $D$  são utilizados como validação. Em seguida, os exemplos de teste, validação e treinamento são trocados pelos próximos conjuntos, repetindo o processo até completar o ciclo.

Com isso, é possível prever o comportamento do meta-classificador diante de novos conjuntos multirrótulos, uma vez que, as recomendações são realizadas sem a interferência de exemplos do mesmo conjunto. Os resultados de validação foram utilizados para a escolha e refinamento dos meta-classificadores. As previsões do teste foram utilizadas como recomendações dos classificadores empregados na previsão e avaliação multirrótulo.

## V. RESULTADOS

De forma a validar a hipótese inicial, na qual a escolha seletiva de classificadores binários melhora os resultados da avaliação multirrótulo, são realizadas comparações dos resultados obtidos de diferentes formas de seleção dos classificadores. Em seguida, são analisados os resultados da recomendação de classificadores utilizando meta-aprendizado. Por fim, são apresentados os resultados multirrótulos obtidos com a utilização dos algoritmos recomendados pelo meta-classificador.

### A. Validação da hipótese

A partir de algumas restrições, para solucionar um problema multirrótulo, como utilizar o método de transformação *Binary Relevance* e os classificadores binários SVM, RF e kNN<sup>3</sup>:

*É possível combinar o uso dos classificadores, na solução dos subproblemas de classificação binário, e melhorar significativamente o desempenho das previsões multirrótulos em relação ao uso de um único classificador?*

Para responder a esta questão, alguns arranjos de classificadores, convencionalmente chamados de estratégia de combinação, foram utilizados para a previsão dos 13 conjuntos multirrótulos. Para cada cada métrica de avaliação, a melhor estratégia tradicional (aquela que gerou o melhor resultado utilizando um único classificador) foi comparada individualmente com cada estratégia proposta, utilizando o teste de Wilcoxon definido com um intervalo de confiança de 95%.

<sup>2</sup>em ambos os momentos, classificação binária e meta-classificador

<sup>3</sup>Utilizando os parâmetros definidos na Seção IV

Tabela III. MÉDIA E DESVIO PADRÃO DOS RESULTADOS MULTIRRÓTULO OBTIDOS COM DIFERENTES COMBINAÇÕES DE CLASSIFICADORES BINÁRIOS

Medida	SVM	RF	KNN	Rand	Acc	BalAcc	F1	Acc <sub>t1</sub>	BalAcc <sub>t1</sub>	F1 <sub>t1</sub>
Acurácia	0.910 ± 0.10	0.917 ± 0.09	0.904 ± 0.10	0.910 ± 0.10	<b>0.921 ± 0.09</b>	0.915 ± 0.09	0.916 ± 0.09	10	0	3
Acurácia exata	0.249 ± 0.28	0.322 ± 0.33	0.294 ± 0.28	0.280 ± 0.28	0.337 ± 0.33	<b>0.341 ± 0.33</b>	0.343 ± 0.33	2	2	5
Hamming Loss	0.090 ± 0.10	0.083 ± 0.09	0.096 ± 0.10	0.09 ± 0.10	<b>0.079 ± 0.09</b>	0.085 ± 0.09	0.084 ± 0.09	12	0	0
Medida F1	0.704 ± 0.19	0.726 ± 0.21	0.706 ± 0.18	0.708 ± 0.19	<b>0.739 ± 0.20</b>	0.733 ± 0.19	<b>0.740 ± 0.19</b>	3	0	5
Precisão	0.806 ± 0.11	0.783 ± 0.18	0.639 ± 0.19	0.719 ± 0.17	<b>0.812 ± 0.12</b>	0.700 ± 0.19	0.705 ± 0.20	5	1	2
Revocação	0.409 ± 0.30	0.493 ± 0.28	0.501 ± 0.25	0.470 ± 0.27	0.515 ± 0.29	<b>0.547 ± 0.26</b>	<b>0.559 ± 0.26</b>	0	4	8
Macro-AUC	0.598 ± 0.12	0.629 ± 0.14	0.628 ± 0.11	0.619 ± 0.12	0.644 ± 0.15	<b>0.658 ± 0.15</b>	0.655 ± 0.15	0	9	4
Macro-F1	0.466 ± 0.23	0.502 ± 0.27	0.483 ± 0.23	0.486 ± 0.23	0.544 ± 0.25	0.532 ± 0.26	<b>0.536 ± 0.26</b>	3	0	5
Macro-Precisão	0.794 ± 0.16	0.743 ± 0.21	0.572 ± 0.24	0.660 ± 0.21	<b>0.819 ± 0.13</b>	0.678 ± 0.20	0.679 ± 0.21	7	0	2
Macro-Revocação	0.256 ± 0.28	0.322 ± 0.30	0.340 ± 0.26	0.309 ± 0.27	0.357 ± 0.32	0.389 ± 0.30	<b>0.394 ± 0.30</b>	0	5	8
Micro-AUC	0.680 ± 0.14	0.723 ± 0.14	0.717 ± 0.12	0.706 ± 0.13	0.733 ± 0.14	<b>0.747 ± 0.13</b>	<b>0.749 ± 0.13</b>	0	3	10
Micro-F1	0.471 ± 0.28	0.565 ± 0.26	0.543 ± 0.23	0.532 ± 0.25	0.586 ± 0.26	<b>0.596 ± 0.24</b>	<b>0.603 ± 0.24</b>	1	0	12
Micro-Precisão	0.806 ± 0.11	0.782 ± 0.18	0.636 ± 0.18	0.716 ± 0.17	<b>0.818 ± 0.11</b>	0.695 ± 0.18	0.699 ± 0.19	5	1	2
Micro-Revocação	0.393 ± 0.29	0.483 ± 0.28	0.489 ± 0.24	0.455 ± 0.26	0.504 ± 0.29	<b>0.540 ± 0.26</b>	<b>0.551 ± 0.26</b>	0	3	10

Os resultados são apresentados na Tabela III, na qual as linhas compreendem as medidas multirrótulo e as colunas compreendem as diferentes estratégias analisadas. Na parte esquerda, são apresentados a média e o desvio padrão de cada métrica. As colunas *SVM*, *RF* e *kNN*, correspondem às abordagens convencionais, na qual um único classificador é utilizado para todos os subproblemas. Na coluna *Rand*, a escolha do classificador foi realizada de forma aleatória, apenas para efeito de comparação. As colunas *Acc*, *BalAcc* e *F1* compreendem às estratégias propostas, cujas escolhas foram realizadas com base no classificador binário que obteve a melhor acurácia, acurácia balanceada e medida F1 respectivamente. Os valores em negrito indicam a estratégia que obteve o melhor resultado para cada medida. Os valores sublinhados indicam os casos definidos como significativos pelo teste de hipótese. No lado direito da tabela, nas colunas *Acc<sub>t1</sub>*, *BalAcc<sub>t1</sub>* e *F1<sub>t1</sub>* são informados o número de conjuntos multirrótulos que as respectivas estratégias propostas obtiveram o melhor resultado.

Ao analisar a tabela é possível observar que, embora nenhuma das estratégias propostas obteve o melhor desempenho para todas as medidas multirrótulos, todas apresentaram um número satisfatório de vitórias quando comparadas com as estratégias usando um único algoritmo de classificação. Além disso, todas as estratégias foram significativamente melhores, em relação ao uso de um único classificador, para pelo menos 3 medidas. Com relação a quantidade de vezes que cada estratégia foi a melhor opção entre todas, é possível constatar que a seleção dos classificadores com base na acurácia contribui para a melhoria das medidas baseadas em exemplo, enquanto que, a escolha dos classificadores com base na medida F1 privilegia as medidas baseadas nos rótulos, especialmente as medidas micro. O uso da acurácia balanceada na escolha dos classificadores se mostrou inferior as outras duas abordagens, contudo mais próxima à última.

Uma observação considerada relevante e não óbvia, diz respeito a relação entre as medidas de precisão e revocação e as estratégias *Acc* e *F1*. Ao utilizar a segunda estratégia, esperávamos obter um equilíbrio em ambas as medidas, uma vez que a mesma indica uma boa relação entre precisão e revocação, contudo isso não foi observado. Enquanto a medida de revocação foi melhorada a precisão foi até mesmo pior que as estratégias tradicionais. Além disso, usando a primeira estratégia, as medidas de precisão foram maximizadas sem entretanto penalizar a revocação.

Ao investigar em detalhe este comportamento, a razão

pode estar no grande número de problemas desbalanceados, que são gerados no processo de transformação multirrótulo. Em geral, o desbalanceamento ocorre porque a classe positiva possui poucos exemplos, com isso, o uso da acurácia como estratégia de seleção, tende a priorizar a classe negativa que é a majoritária, predizendo poucos exemplos como positivos. Por outro lado, ao usar a acurácia balanceada ou a medida F1, busca-se um melhor compromisso entre as predições de ambas as classes, o que leva o classificador a obter um número maior de predições falso-positivo. Com a união dos rótulos este erro se propaga, e o somatório dos mesmos impacta diretamente na qualidade das métricas relacionadas com a precisão. A revocação, ao contrário, é melhorada pois passa-se a acertar mais valores da classe positiva.

Retornando à questão inicial, a qual nos propomos a responder nesta seção, os resultados obtidos com esta simulação indicam que, é possível selecionar os classificadores a partir de critérios de qualidade dos seus resultados binários, obtendo resultados multirrótulos com significância estatística para algumas medidas. O aumento do número de classificadores disponíveis para escolha, tende a melhorar estes resultados e piorar a seleção randômica. Portanto, considerando que uma estratégia de tentativa e erro ou força bruta, para os  $n$  classificadores e  $q$  rótulos distintos, possui  $n^q$  combinações distintas, as estratégias de seleção usando as medidas de classificação binária se mostram efetivamente viáveis.

## B. Meta-classificador

Estabelecida a viabilidade do uso seletivo de classificadores em problemas de transformação multirrótulo, o próximo desafio consiste em como realizar uma escolha adequada sem a necessidade de verificar todas as possibilidades de combinação e principalmente, sem ter que treinar todos os classificadores para todos os subproblemas. Para isso, a estratégia de meta-aprendizado foi utilizada para recomendar os classificadores adequados para indução dos modelos.

Seguindo a estratégia de validação por conjunto multirrótulo (ver Seção IV-C), o meta-classificador foi treinado sem usar nenhum exemplo de treinamento referente ao conjunto a ser predito, simulando o comportamento do meta-aprendizado diante de novos conjuntos.

Nas Tabelas IV e V são apresentados os resultados obtidos na recomendação dos classificadores utilizando as meta-bases *mtb-Acc* e *mtb-F1* respectivamente. As linhas correspondem

Tabela IV. RESULTADOS DA PREDIÇÃO DO META-APRENDIZ AO RECOMENDAR OS CLASSIFICADORES COM MELHOR ACURÁCIA

Base	Ex.	Acurácia			F1		
		Maj.	Val.	Tes.	Maj.	Val.	Tes.
bibtex	159	0.692	<b>0.704</b>	<b>0.704</b>	0.273	<b>0.341</b>	<b>0.341</b>
birds	19	0.526	<b>0.526</b>	<b>0.526</b>	0.230	0.230	<b>0.238</b>
CAL500	174	<b>0.908</b>	0.345	0.351	<b>0.317</b>	0.205	0.207
corel5k	371	<b>0.951</b>	0.922	<b>0.951</b>	0.325	0.365	<b>0.521</b>
emotions	6	0	0	0	0	0	0
enron	52	<b>0.673</b>	0.654	<b>0.673</b>	0.268	0.304	<b>0.343</b>
flags	7	<b>0.714</b>	0.286	<b>0.714</b>	<b>0.278</b>	0.206	<b>0.278</b>
genbase	25	<b>0.520</b>	0.320	0.200	<b>0.228</b>	0.242	0.175
mediamill	101	0.545	0.624	<b>0.634</b>	0.235	0.394	<b>0.407</b>
medical	38	<b>0.632</b>	0.579	0.579	0.258	<b>0.411</b>	<b>0.411</b>
scene	6	0.500	0.500	0.500	0.222	0.222	0.222
tmc2007-500	22	0	0.545	<b>0.591</b>	0	0.235	<b>0.248</b>
yeast	14	0.429	<b>0.571</b>	<b>0.571</b>	0.200	<b>0.356</b>	<b>0.356</b>
Média Ponderada	-	<b>0.777</b>	0.68	0.695	0.284	0.322	<b>0.383</b>

aos conjuntos multirrótulos e as colunas às medidas de avaliação do meta-classificador. Para efeitos de comparação, o desempenho obtido por um classificador que prediz apenas a classe majoritária foi adicionado à tabela, também foram diferenciados os resultados de validação e teste. A última linha contém a média ponderada das medidas, uma vez que o número de exemplos de cada conjunto varia. Os valores em negrito correspondem ao valor do melhor resultado para cada medida.

Em ambos os cenários, a predição da classe majoritária obteve uma melhor acurácia e os resultados do teste tiveram um melhor desempenho para a medida F1. Todavia, ao comparar individualmente os conjuntos multirrótulos, o uso do meta-aprendiz obteve um melhor desempenho preditivo para um número maior de bases. A variação entre os resultados de validação e teste, são em partes decorrentes da metodologia adotada, pois os resultados de validação e teste foram obtidos de modelos de classificação diferentes e em partes do uso de poucos conjuntos multirrótulos. Uma diversidade maior de conjuntos, levaria a uma estabilidade maior dos meta-modelos, fazendo com que ao retirar exemplos de dois conjuntos simultaneamente, os dados de treinamento não percam representatividade.

Com relação aos resultados obtidos utilizando a meta-base *mtb-Acc*, três conjuntos (*CAL500*, *emotions* e *genbase*) tiveram uma acurácia abaixo de 50% no resultado de teste. O fato de *CAL500* possuir um grande número de rótulos, impactou negativamente na média ponderada. Na predição utilizando a meta-base *mtb-F1*, os resultados de validação foram sutilmente melhores do que os do teste e os valores abaixo de 50% aconteceram para um número maior de conjuntos multirrótulos.

Observando apenas os resultados de validação, foram realizadas algumas simulações visando a melhoria do desempenho preditivo. Foram utilizadas técnicas de pré-processamento, balanceamento artificial de dados e o uso de outros classificadores no meta-aprendiz, todavia os resultados obtidos foram similares ou piores. Neste sentido, acreditamos que duas medidas sejam necessárias para a melhoria do meta-aprendiz, apontadas como trabalhos futuros. As propostas consistem em explorar o uso de novas meta-características como, por exemplo, medidas de *landmarking* [21] e de complexidades [22]; e, diversificar os meta-exemplos incluindo novos conjuntos multirrótulos.

Tabela V. RESULTADOS DA PREDIÇÃO DO META-APRENDIZ AO RECOMENDAR OS CLASSIFICADORES COM MELHOR MEDIDA F1

Base	Ex.	Acurácia			F1		
		Maj.	Val.	Tes.	Maj.	Val.	Tes.
bibtex	159	0.465	<b>0.472</b>	0.459	0.212	<b>0.292</b>	0.264
birds	19	0.211	<b>0.526</b>	<b>0.526</b>	0.116	<b>0.485</b>	0.412
CAL500	174	<b>0.609</b>	0.316	0.316	<b>0.252</b>	0.165	0.173
corel5k	371	0.919	0.884	<b>0.903</b>	0.319	0.467	0.476
emotions	6	0	<b>0.667</b>	0.500	0	<b>0.472</b>	0.286
enron	52	0.558	0.596	<b>0.615</b>	0.239	<b>0.456</b>	0.429
flags	7	<b>0.714</b>	0.429	0.571	0.278	0.300	<b>0.579</b>
genbase	25	<b>0.520</b>	0.200	0.400	0.228	0.175	<b>0.326</b>
mediamill	101	0.267	0.307	<b>0.317</b>	0.141	0.187	<b>0.322</b>
medical	38	0.632	<b>0.763</b>	0.711	0.258	0.530	<b>0.583</b>
scene	6	<b>0.667</b>	0.167	0.167	<b>0.267</b>	0.095	0.095
tmc2007-500	22	0	<b>0.227</b>	0.182	0	<b>0.123</b>	0.103
yeast	14	0.214	<b>0.714</b>	0.643	0.118	<b>0.439</b>	0.421
Média Ponderada	-	<b>0.634</b>	0.591	0.599	0.247	0.341	<b>0.358</b>

### C. Multirrótulo

A partir das recomendações obtidas na meta-classificação, analisamos os resultados no contexto multirrótulo. Nesta análise, comparamos o uso dos classificadores indicados pelo meta-aprendiz em relação ao uso de um único classificador para todos os subproblemas, o SVM. Sua escolha, foi motivada por ser o classificador que obteve o melhor desempenho para um número maior de subproblemas.

Nas Tabelas VI e VII são apresentadas as diferenças entre as medidas multirrótulo obtidas em cada conjunto de dados, utilizando as meta-bases *mtb-Acc* e *mtb-F1*, respectivamente. Com exceção da medida *Hamming Loss* (HL), que é uma medida de erro, os valores positivos indicam o quanto a recomendação do meta-aprendiz foi melhor em comparação ao SVM, enquanto que para HL os valores negativos representam os melhores resultados. A última coluna indica a quantidade de métricas que a estratégia proposta obteve o melhor resultado e as últimas linhas indicam a média das diferenças e o número de bases que os resultados obtidos foram superiores. Os valores em negrito indicam que o resultado positivo foi obtido em um número maior de vezes pela nossa abordagem e os valores sublinhados indicam que houve relevância estatística no resultado obtido.

A análise dos resultados será realizada sob duas perspectivas distintas. Na perspectiva das métricas, a análise foi iniciada na Seção V-A, onde foi considerado o melhor cenário possível de cada métrica/estratégia, assim é relevante compará-las com o cenário obtido. Na perspectiva dos conjuntos multirrótulos, a análise deve ser realizada observando as distintas medidas de avaliação obtidas para cada conjunto.

### Perspectiva das métricas

Utilizando a predição dos classificadores recomendados pela acurácia, apenas as medidas relacionadas com precisão tiveram um desempenho médio inferior em relação ao uso do SVM. Para as demais medidas, foram obtidos resultados em média superiores, além de ocorrerem um maior número de vezes. As diferenças médias foram pequenas, variando entre 0.1% a 5%, contudo, para quase todas as métricas, a estratégia foi superior para um número maior de bases. As medidas mais beneficiadas nessa abordagem foram a *acurácia exata*, as

Tabela VI. DIFERENÇA DO RESULTADO ENTRE AS RECOMENDAÇÕES DO META-APRENDIZ UTILIZANDO A META-BASE *mtb-Acc* E O SVM

Base	Acur.	Acur. Ex.	HL	F1	Prec.	Rev.	Macro				Micro				Melhor
							AUC	F1	Prec.	Rec.	AUC	F1	Prec.	Rec.	
bibtex	0	0.001	0	-0.001	-0.003	0.004	0	0.003	0	0.001	0.002	0.005	-0.003	0.003	7
birds	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CAL500	-0.009	0	0.009	-0.022	-0.071	-0.008	-0.001	-0.180	-0.241	0.007	-0.007	-0.018	-0.094	-0.005	1
corel5k	0	0	0	-0.013	-0.244	0.022	0	-0.022	0.002	0.002	0.011	0.037	-0.243	0.023	6
emotions	0.020	0.030	-0.020	0.047	0.018	0.056	0.029	0.085	-0.016	0.062	0.032	0.094	0.027	0.068	13
enron	0	0.005	0	0.003	-0.017	0.033	0.002	0.019	0	0.006	0.012	0.017	-0.015	0.025	9
flags	-0.002	-0.031	0.002	0.001	0.001	-0.026	0	-0.010	-0.023	-0.016	-0.003	-0.007	0.006	-0.018	3
genbase	0.003	0.040	-0.003	-0.01	0	0.060	0.002	0.028	0	0.005	0.035	0.037	0	0.069	10
mediamill	0.001	0.008	-0.001	0.009	0.003	0.017	0.003	-0.008	-0.026	0.005	0.009	0.016	0.004	0.017	12
medical	0.008	0.304	-0.008	0.022	0.021	0.352	0.044	0.294	0.111	0.103	0.179	0.398	0.011	0.359	14
scene	-0.005	-0.053	0.005	0.003	0.033	-0.063	-0.028	-0.043	0.026	-0.063	-0.028	-0.035	0.035	-0.063	4
tmc2007-500	0.033	0.430	-0.033	0.149	0.150	0.166	0.110	0.164	0.064	0.194	0.105	0.177	0.155	0.196	14
yeast	0.002	0.011	-0.002	0.008	-0.007	0.028	0.010	-0.053	-0.064	0.030	0.009	0.013	-0.009	0.027	10
Média	<b>0.004</b>	<b>0.057</b>	<b>-0.004</b>	<b>0.015</b>	-0.009	<b>0.049</b>	<b>0.013</b>	<b>0.021</b>	-0.013	<b>0.026</b>	<b>0.027</b>	<b>0.056</b>	-0.010	<b>0.054</b>	<b>11</b>
Melhor	<b>6</b>	<b>8</b>	<b>6</b>	<b>8</b>	<b>6</b>	<b>9</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>10</b>	<b>9</b>	<b>9</b>	<b>6</b>	<b>9</b>	<b>12</b>

Tabela VII. DIFERENÇA DO RESULTADO ENTRE AS RECOMENDAÇÕES DO META-APRENDIZ UTILIZANDO A META-BASE *mtb-F1* E O SVM

Base	Acur.	Acur. Ex.	HL	F1	Prec.	Rev.	Macro				Micro				Melhor
							AUC	F1	Prec.	Rec.	AUC	F1	Prec.	Rec.	
bibtex	-0.001	-0.058	0.001	-0.082	-0.091	-0.063	-0.004	-0.026	-0.049	-0.008	-0.013	-0.043	-0.089	-0.026	0
birds	0	0.009	0	-0.002	-0.350	0.250	0.075	0.124	-0.391	0.165	0.131	0.286	-0.347	0.278	8
CAL500	-0.027	0	0.027	0.044	-0.187	0.101	0.005	-0.158	-0.283	0.067	0.028	0.051	-0.192	0.106	7
corel5k	-0.001	0.006	0.001	0.049	-0.341	0.059	0.004	0.010	-0.478	0.013	0.028	0.081	-0.348	0.058	9
emotions	0.035	0.099	-0.035	0.074	-0.162	0.201	0.089	0.234	-0.155	0.238	0.084	0.244	-0.108	0.228	11
enron	-0.006	0.019	0.006	-0.028	-0.110	0.044	0.011	-0.009	-0.049	0.031	0.010	-0.013	-0.124	0.029	6
flags	-0.013	-0.062	0.013	-0.023	-0.004	-0.065	-0.004	-0.019	-0.024	-0.033	-0.015	-0.023	0.002	-0.046	4
genbase	0.004	0.050	-0.004	-0.006	0	0.065	0.021	0.029	0	0.047	0.039	0.042	0	0.078	10
mediamill	-0.003	-0.010	0.003	-0.009	-0.089	0.032	0.003	-0.001	-0.113	0.011	0.014	-0.001	-0.101	0.032	5
medical	0.008	0.304	-0.008	0.022	0.021	0.352	0.044	0.294	0.111	0.103	0.179	0.398	0.011	0.359	14
scene	-0.024	0.015	0.024	-0.001	-0.152	0.013	-0.011	-0.048	-0.107	0.009	-0.009	-0.047	-0.146	0.014	4
tmc2007-500	0.009	0.053	-0.009	0.036	0.063	0.029	-0.028	-0.045	-0.028	-0.067	0.022	0.047	0.057	0.038	10
yeast	-0.017	0.044	0.017	0.004	-0.051	0.038	0.040	-0.077	-0.285	0.099	0	-0.002	-0.063	0.042	6
Média	-0.003	<b>0.036</b>	0.003	<b>0.006</b>	-0.112	<b>0.081</b>	<b>0.019</b>	<b>0.024</b>	-0.142	<b>0.052</b>	<b>0.038</b>	<b>0.078</b>	-0.111	<b>0.091</b>	<b>9</b>
Melhor	4	<b>9</b>	4	6	2	<b>11</b>	<b>10</b>	6	1	<b>11</b>	<b>9</b>	7	3	<b>11</b>	7

medidas de revocação e a medida *micro-F1*. Houve relevância estatística para as medidas *micro-AUC* e *micro-revocação*.

Utilizando a predição da recomendação pela medida F1, além das medidas de precisão, outras (*acurácia* e *Hamming Loss*) obtiveram resultados inferiores ao uso do SVM. As diferenças obtidas neste cenário foram maiores que o primeiro, chegando a aproximadamente 10% de melhoria (*micro-revocação*) e 15% de piora (*macro-precisão*). O número de vezes em que os resultados obtidos foram melhores, foi inferior a primeira estratégia, contudo, para as medidas de revocação foram obtidos os maiores valores, além de um número maior de medidas consideradas estatisticamente relevantes.

### Perspectiva dos conjuntos multirrótulos

Embora as diferenças médias tenham sido pequenas, para alguns conjuntos as diferenças foram maiores do que 10%, chegando em alguns casos a quase 50%.

No primeiro cenário, os conjuntos *medical* e *tmc2007-500* obtiveram em distintas métricas grandes diferenças positivas, chegando a aproximadamente 40% (*micro-F1*) no primeiro caso e 43% (*Acurácia exata*) no segundo. Houve também diferenças negativas maiores, para os conjuntos *CAL500* e *corel5k* associadas a medidas de precisão.

No segundo cenário, as diferenças individuais ocorreram

em maior número tanto positivamente quanto negativamente. Os conjuntos *emotions* e *medical* são positivamente destacados, entretanto para a maioria dos conjuntos resultados positivos e negativos foram obtidos. Isto é decorrente do bom desempenho desta abordagem para as medidas de revocação em comprometimento das medidas de precisão.

Considerando que a estratégia proposta foi melhor para um conjunto, se foi obtido um número maior de medidas positivas do que o SVM, no primeiro cenário, os resultados foram melhores para 8 dos 13 conjuntos, enquanto que no segundo foram para 7 deles. Considerando que o desempenho do meta-classificador foi de respectivamente 70% e 60%, os resultados obtidos foram considerados satisfatórios.

Portanto, a escolha entre as abordagens investigadas deve ser determinada pelos requisitos da solução. Enquanto o uso da acurácia melhora um número maior de medidas o uso da medida F1 obtém resultados mais significantes. Todavia, ambas as estratégias se mostraram não adequadas para melhoria das medidas de precisão. Ressalta-se que a melhoria do meta-classificador e a recomendação de um número maior de classificadores pode levar a obtenção de resultados ainda mais competitivos.

## VI. CONCLUSÃO

Investigamos neste estudo o uso seletivo de classificadores binários na solução de problemas multirrótulos, por meio do

uso de meta-aprendizado para recomendação dos classificadores. Inicialmente, foram consideradas três estratégias distintas de seleção, no entanto, apenas as duas mais promissoras foram investigadas em detalhes. Tais abordagens consistem na recomendação dos classificadores que obtiveram a melhor acurácia e medida F1 durante a classificação binária.

A análise realizada a partir da comparação empírica destas estratégias, em relação ao uso de um único classificador na predição multirrótulo, sugere que utilizando as recomendações do meta-classificador, em média, são obtidos resultados consistentes, ao considerar uma solução que envolve múltiplos domínios ou diferentes cenários. Os resultados obtidos no contexto multirrótulo, se mostraram competitivos, quando comparados com o método base.

Partindo da premissa de que é possível ter um sistema de recomendação ideal, comparamos os limites das estratégias propostas, determinando seu potencial em relação ao uso de três classificadores distintos (SVM, RF e kNN). Nesta análise, foi possível observar relações entre problemas binários desbalanceados e sua influência nas medidas de avaliação multirrótulo relacionadas com precisão e revocação. Todavia, como é descartada a viabilidade da premissa em questão, os resultados obtidos na avaliação do meta-classificador, como sistema de recomendação de classificadores, podem ser utilizados como referência para balizar novos trabalhos nesta direção.

A solução apresentada neste estudo não se estabelece como um novo método de classificação multirrótulo, mas como uma estratégia de melhoria dos métodos multirrótulos baseados em transformação de problemas. A escolha do método de transformação *Binary Relevance*, na parte experimental do estudo, foi motivada pela sua simplicidade e, principalmente, por este ser a base de outros métodos estado da arte [23], [11]. Nos próximos trabalhos, pretendemos explorar outros métodos de transformação binária, principalmente aqueles que utilizam outras abordagens de decomposição, como por exemplo todos-contra-todos, no qual o número de subproblemas gerado é maior e mais diversificado.

Como trabalhos futuros, além dos já citados no texto, pretendemos explorar a recomendação de um número maior de classificadores; em um nível meta, encontrar formas de identificar soluções para os problemas de dados desbalanceados, obtidos pela aplicação dos métodos de transformação; e também, investigar as relações entre as métricas de avaliação dos subproblemas binários em relação as medidas resultantes na avaliação multirrótulo.

## VII. AGRADECIMENTOS

Pesquisa desenvolvida com utilização dos recursos computacionais do Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI) financiados pela FAPESP.

## REFERÊNCIAS

- [1] E. Gibaja and S. Ventura, "Multi-label learning: a review of the state of the art and ongoing research," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 6, pp. 411–444, 2014.
- [2] —, "A Tutorial on Multilabel Learning," *ACM Computing Surveys*, vol. 47, no. 3, pp. 1–38, 2015.
- [3] G. Tsoumakas and I. Katakis, "Multi-Label Classification: An Overview," *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1–13, Jan. 2007.
- [4] L. Chekina, L. Rokach, and B. Shapira, "Meta-learning for selecting a multi-label classification algorithm," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 220–227, 2011.
- [5] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Džeroski, "An extensive experimental comparison of methods for multi-label learning," *Pattern Recognition*, vol. 45, no. 9, pp. 3084–3104, 2012.
- [6] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.
- [7] A. C. P. L. F. de Carvalho and A. A. Freitas, "A Tutorial on Multi-label Classification Techniques," in *Foundations of Computational Intelligence Volume 5*, 2009, pp. 177–195.
- [8] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.
- [9] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier Chains for Multi-label Classification," *Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science*, vol. 5782, pp. 254–269, 2009.
- [10] G. Tsoumakas, A. Dimou, E. Spyromitros, V. Mezaris, I. Kompatsiaris, and I. Vlahavas, "Correlation-based pruning of stacked binary relevance models for multi-label learning," *Proceedings of the Workshop on Learning from Multi-Label Data (MLD'09)*, pp. 22–30, 2009.
- [11] E. Alvares-Cherman, J. Metz, and M. C. Monard, "Incorporating label dependency into the binary relevance framework for multi-label classification," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1647–1655, 2012.
- [12] M.-L. Zhang and Z.-H. Zhou, "A Review on Multi-Label Learning Algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.
- [13] S. Ali and K. a. Smith, "On learning algorithm selection for classification," *Applied Soft Computing*, vol. 6, no. 2, pp. 119–138, 2006.
- [14] R. Vilalta, C. Giraud-Carrier, P. Brazdil, and C. Soares, "Using Meta-Learning to Support Data Mining," *International Journal of Computer Science Applications*, vol. 1, no. 1, pp. 31–45, 2004.
- [15] C. Castiello, G. Castellano, and A. M. Fanelli, "Meta-data: Characterization of Input Features for Meta-learning," *Modeling Decisions for Artificial Intelligence*, vol. 3558, pp. 457–468, 2005.
- [16] K. a. Smith-Miles, "Cross-disciplinary perspectives on meta-learning for algorithm selection," *ACM Computing Surveys*, vol. 41, no. 1, pp. 1–25, 2008.
- [17] E. A. Tanaka, S. R. Nozawa, A. A. Macedo, and J. A. Baranauskas, "A multi-label approach using binary relevance and decision trees applied to functional genomics," *Journal of Biomedical Informatics*, vol. 54, pp. 85–95, 2015.
- [18] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label Data," in *Data Mining and Knowledge Discovery Handbook*, 2nd ed., O. Maimon and L. Rokach, Eds. Springer US, 2010, ch. 34, pp. 667–685.
- [19] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: <http://www.R-project.org/>
- [20] F. Charte, *mldr: Exploratory Data Analysis and Manipulation of Multi-Label Data Sets*, 2015, r package version 0.1.70. [Online]. Available: <http://CRAN.R-project.org/package=mldr>
- [21] B. Pfahringer, H. Bensusan, and C. Giraud-Carrier, "Meta-Learning by Landmarking Various Learning Algorithms," *Proceedings of the Seventeenth International Conference on Machine Learning ICML2000*, vol. 951, no. 2000, pp. 743–750, 2000.
- [22] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE Transactions on Pattern Analysis and Machine Learning*, vol. 24, no. 3, pp. 289–300, 2002.
- [23] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine Learning*, vol. 85, no. 3, pp. 333–359, 2011.