

Capítulo

3

<https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/7/11/36-1?inline=1>

Introdução à Classificação Multirrótulo

Eduardo Corrêa Gonçalves



Classificação Multirrótulo



Introdução

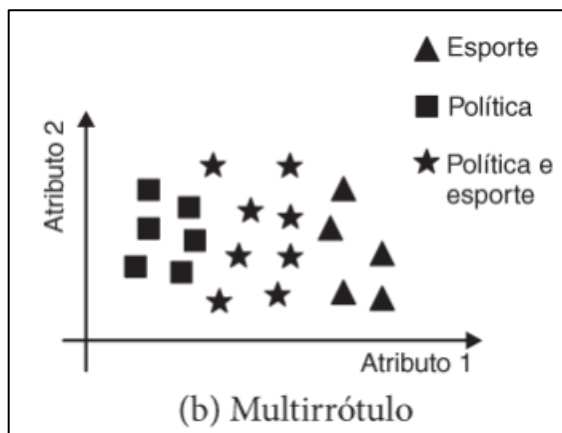
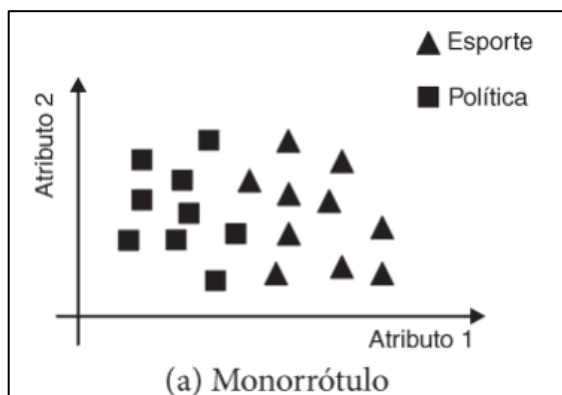
- Existem problemas de classificação, conhecidos como problemas de classificação multirrótulo, em que cada exemplo pode pertencer simultaneamente a mais de uma classe



Introdução

- Problemas de classificação multirrótulo ocorrem com frequência na área de processamento de textos, entre outras
 - Um documento pode ser classificado ao mesmo tempo nas categorias Política e Economia, por exemplo
 - Um paciente pode ser diagnosticado com duas doenças
 - etc.

Introdução



X	Metal	Jazz	Bossa	Pop
X_1	●			●
X_2		●	●	
X_3		●		
X_4	●			
X_5		●	●	●



Introdução

- O objetivo é realizar o aprendizado de uma função h (classificador) que dado um objeto não rotulado seja capaz de prever de forma efetiva o seu labelset (conjunto de rótulos de classe)



Introdução

- Neste contexto é muito comum a existência de correlação entre rótulos de classe
 - Exemplo: a probabilidade de uma canção ser rotulada como "Pop" torna-se mais forte caso ela tenha sido rotulada como "Dance" e "R&B" (Rhythm & Blues)
- Assim, é intuitivo esperar que algoritmos capazes de capturar e modelar as correlações entre os rótulos sejam mais acurados



Principais Abordagens

- Os métodos existentes podem ser divididos em duas grandes abordagens:
 - **Abordagem independente de algoritmo (+utilizados)**
 - Utiliza algoritmos tradicionais de classificação para tratar problemas multirrótulo, transformando o problema multirrótulo original em um conjunto de problemas monorrótulo
 - São flexíveis, uma vez que permitem a abstração do algoritmo de classificação base (monorrótulo) – vantagem, pois diferentes algoritmos monorrótulo são mais ou menos efetivos de acordo com os diferentes domínios de aplicação



Principais Abordagens

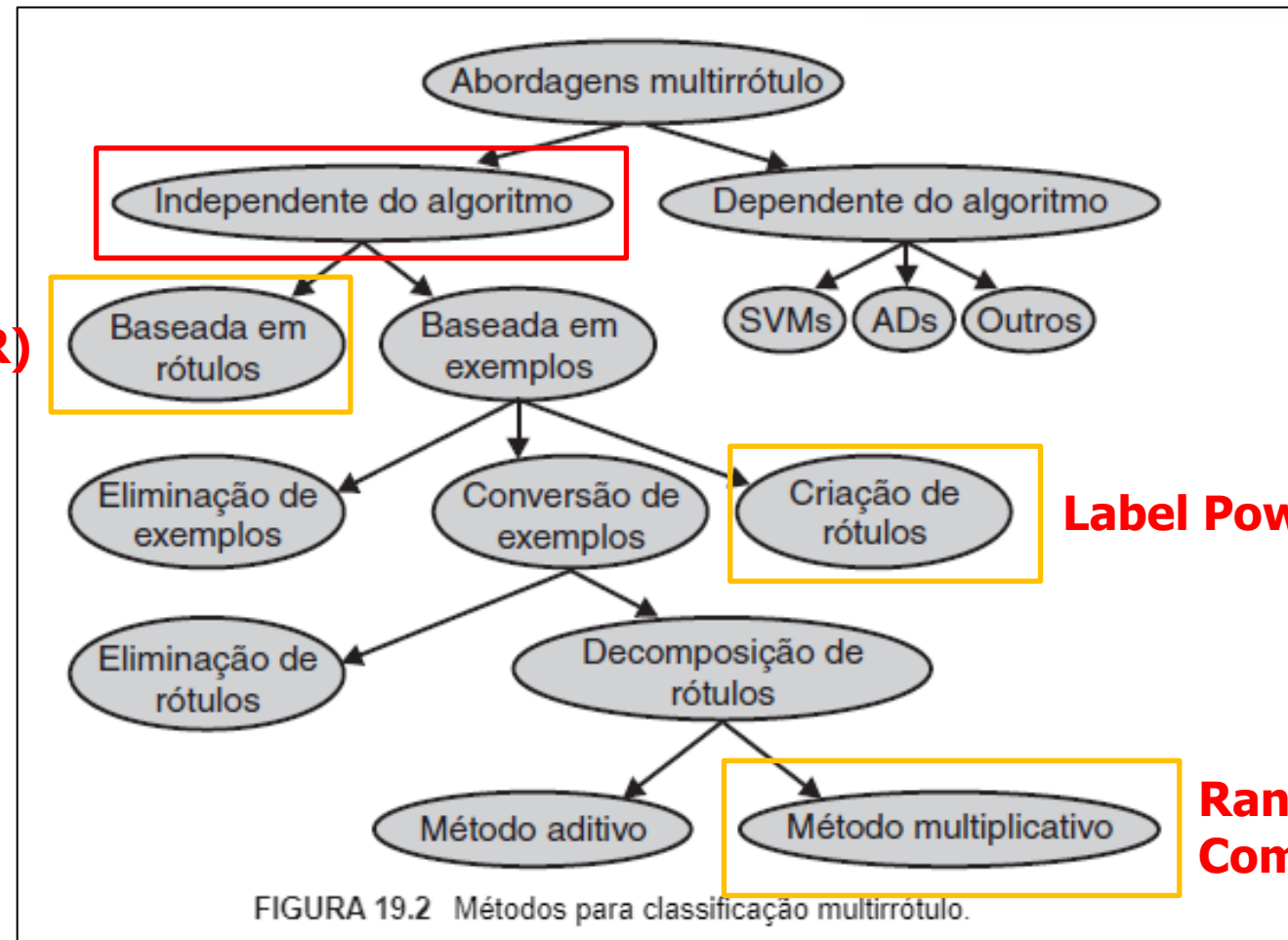
- Os métodos existentes podem ser divididos em duas grandes abordagens:
 - **Abordagem dependente de algoritmo**
 - Cria algoritmos específicos para tratar o problema multirrótulo, os quais podem ser baseados em algoritmos convencionais, como SVM e árvores de decisão, ou podem ser especificamente desenvolvidos para classificação multirrótulo

Principais Abordagens

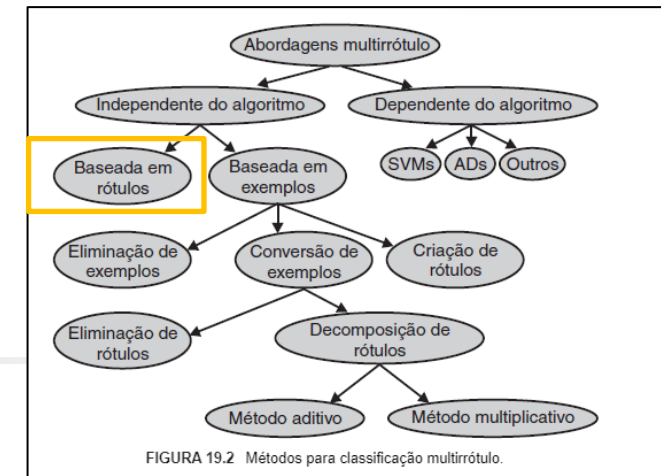
Binary Relevance (BR)

Label Powerset (LP)

Ranking by Pairwise Comparison (RPC)

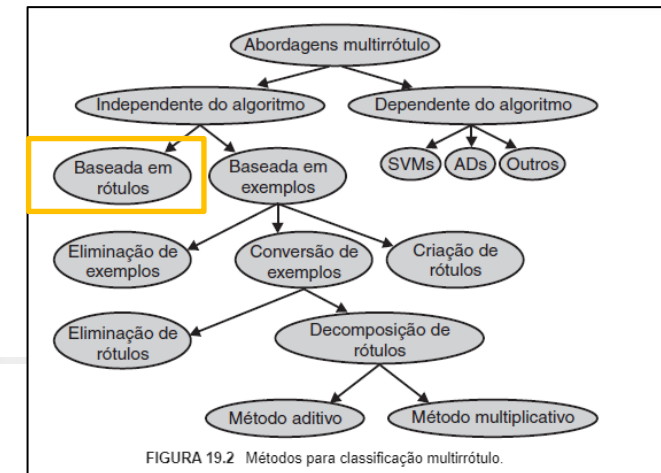


Transformação Baseada nos Rótulos das Classes (**Binary Relevance (BR)**)



- **Método mais conhecido e adotado**
- Utiliza-se k classificadores, sendo k o número de classes do problema
- Cada classificador é então associado a uma classe e treinado para resolver um problema de classificação binária, na qual é considerada a classe que ele está associado contra todas as outras classes envolvidas
- Também chamado de método binário ou um-contra-todos

Transformação Baseada nos Rótulos das Classes (**Binary Relevance (BR)**)



X	Metal	Jazz	Bossa	Pop
X ₁	•			•
X ₂		•	•	
X ₃		•		
X ₄	•			
X ₅		•	•	•

X	Y:Metal
X ₁	•
X ₂	
X ₃	
X ₄	•
X ₅	

X	Y:Bossa
X ₁	
X ₂	•
X ₃	
X ₄	
X ₅	•

X	Y:Jazz
X ₁	
X ₂	•
X ₃	•
X ₄	
X ₅	•

X	Y:Pop
X ₁	•
X ₂	
X ₃	
X ₄	
X ₅	•

Transformação Baseada nos Rótulos das Classes (**Binary Relevance (BR)**)

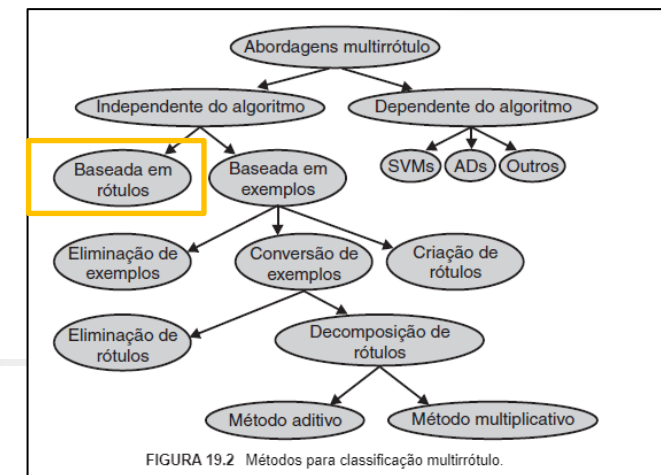


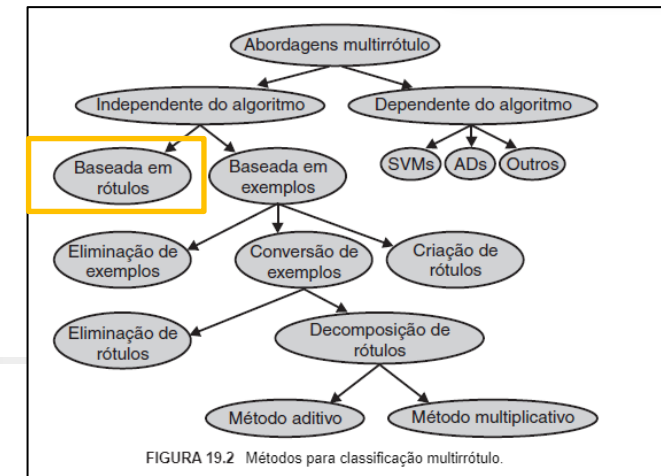
FIGURA 19.2 Métodos para classificação multirrotulo.

$t = \text{"Garota de Ipanema"} - \text{Tom Jobim \& Frank Sinatra}$

Classificadores	Classificações
$f_1: X \rightarrow \{\text{Metal}, \sim\text{Metal}\}$	$\sim\text{Metal}$
$f_2: X \rightarrow \{\text{Jazz}, \sim\text{Jazz}\}$	Jazz
$f_3: X \rightarrow \{\text{Bossa}, \sim\text{Bossa}\}$	$\sim\text{Bossa}$
$f_4: X \rightarrow \{\text{Pop}, \sim\text{Pop}\}$	Pop
Conjunto de Rótulos Predito	$\{\text{Jazz}, \text{Pop}\} \subseteq L$

O conjunto de rótulos de um novo objeto é predito através da simples combinação das saídas produzidas por cada um dos classificadores binários

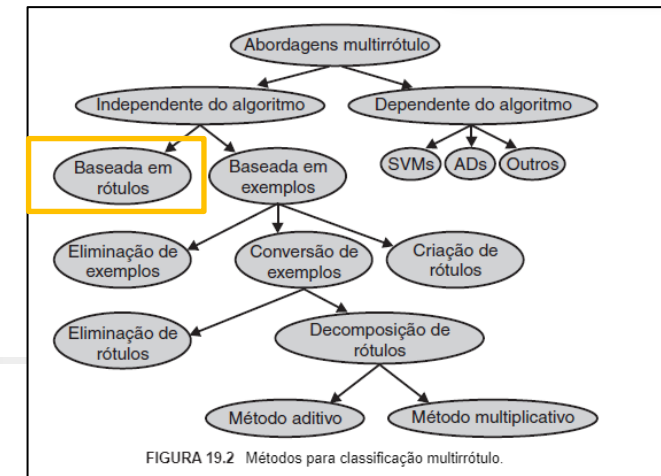
Transformação Baseada nos Rótulos das Classes (**Binary Relevance (BR)**)



■ Vantagens:

- Assim como a abordagem LP, trata-se de um método simples e intuitivo
- Assim como a abordagem RPC, é capaz de prever conjuntos de rótulos que não estão presentes na base de dados multirrótulo original
- Ao contrário do LP e RPC, o método BR possui complexidade linear em k (número de classes)

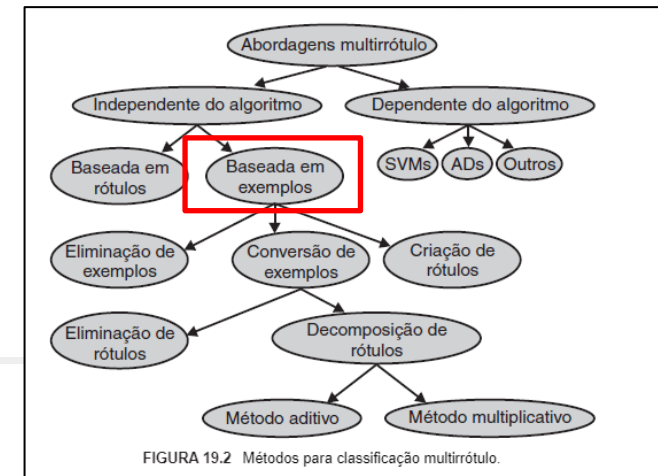
Transformação Baseada nos Rótulos das Classes (**Binary Relevance (BR)**)



- Desvantagem:

- Um classificador BR ignora as possíveis correlações entre os rótulos de classe, uma vez que os classificadores binários tomam decisões independentes
 - Apesar disso, o método costuma apresentar desempenho satisfatório
- O processo de transformação é reversível, i.e., é possível recuperar as classes do problema original a partir do novo problema criado

Transformação Baseada nos Exemplos

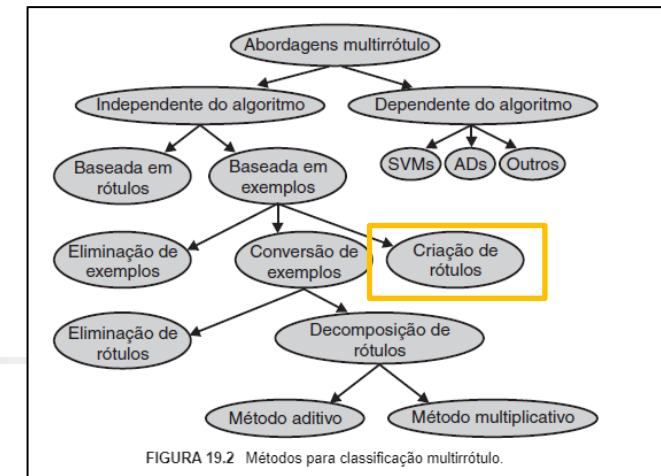


- Ao contrário do método anterior, esse método não produz apenas problemas de classificação binária, podendo também produzir problemas multiclasse
- Três diferentes estratégias são propostas para esse tipo de transformação

Transformação Baseada nos Exemplos

Criação de Novos Rótulos para os Exemplos

Multirrótulo Existentes (**Label Powerset (LP)**)



- Para cada exemplo, todas as classes atribuídas àquele exemplo são combinadas em uma nova e única classe
- Com a criação de novas classes, as classes do problema original não são perdidas
- O número de classes envolvidas no problema pode aumentar consideravelmente, e algumas classes podem terminar com poucos exemplos que as representem
 - Se forem utilizados classificadores multiclasse a quantidade desses classificadores se mantém a mesma
 - Se forem utilizados classificadores binários, o número de classificadores necessários aumenta

Transformação Baseada nos Exemplos

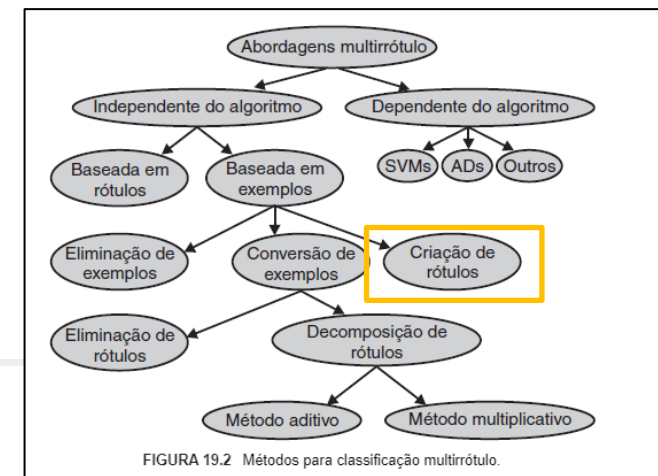
Criação de Novos Rótulos para os Exemplos

Multirrótulo Existentes (**Label Powerset (LP)**)

X	Metal	Jazz	Bossa	Pop
X ₁	•			•
X ₂		•	•	
X ₃		•		
X ₄	•			
X ₅		•	•	•



X	Y
X ₁	Metal-Pop
X ₂	Jazz-Bossa
X ₃	Jazz
X ₄	Metal
X ₅	Jazz-Bossa-Pop



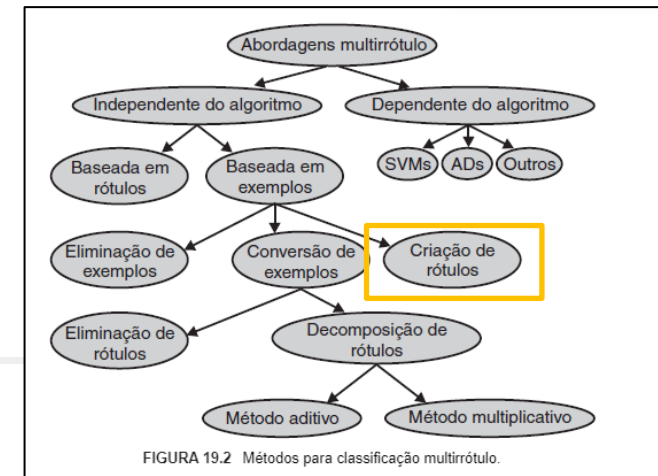
Transformação Baseada nos Exemplos

Criação de Novos Rótulos para os Exemplos

Multirrótulo Existentes (**Label Powerset (LP)**)

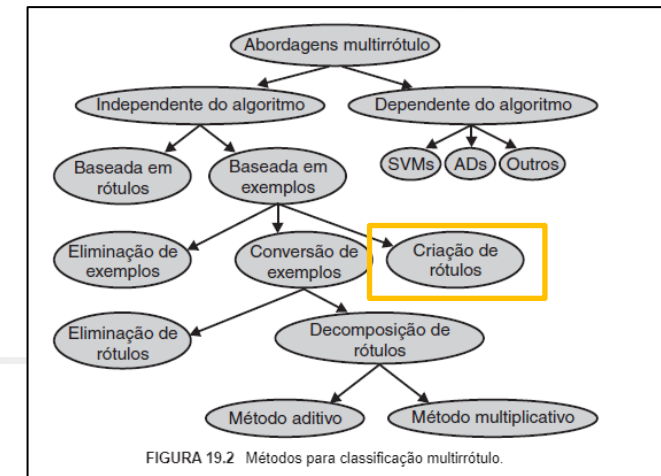
- Vantagem:

- É simples e oferece a vantagem de implicitamente levar a correlação entre rótulos em consideração



Transformação Baseada nos Exemplos

Criação de Novos Rótulos para os Exemplos Multirrótulo Existentes (**Label Powerset (LP)**)

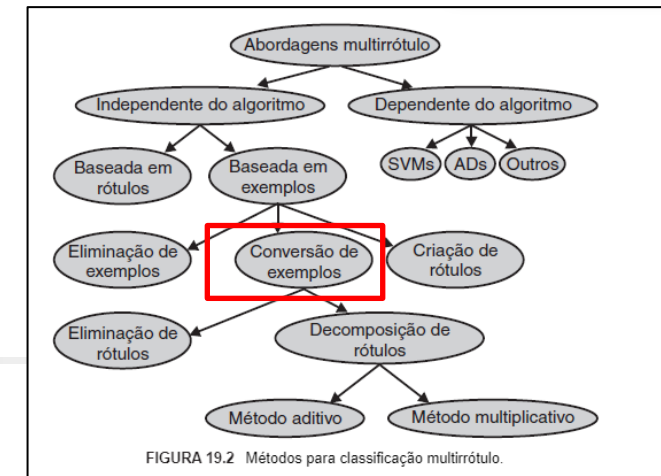


■ Desvantagens:

- Não é capaz de prever labelsets que não estejam presentes na base de treinamento original
 - O modelo do exemplo não será capaz de classificar uma nova música como “Jazz-Pop” ou somente como “Bossa” porque essas combinações não existem a partir da base original
- A transformação pode gerar um número exponencial de classes compostas, algumas delas associadas a um número reduzido de objetos
 - O número máximo de classes compostas é dado por $\min(N, 2^q)$, onde N corresponde ao número de objetos da base de treinamento e q representa o número de rótulos
 - Costuma ser inviável para muitos problemas reais

Transformação Baseada nos Exemplos

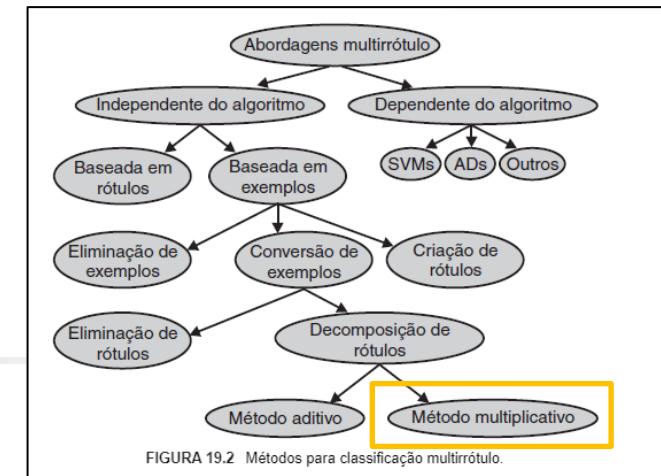
Conversão de Exemplos Multirrótulo em Exemplos Monorrótulo



- Existem duas variações
 - Simplificação ou eliminação de rótulos
 - Todos os exemplos multirrótulo são convertidos em exemplos monorrótulo
 - **Decomposição de rótulos**
 - Decompõe todos os exemplos multirrótulo em um conjunto de exemplos monorrótulo

Transformação Baseada nos Exemplos

Decomposição de rótulos (**Ranking by Pairwise Comparision (RPC)**): Método multiplicativo



- Utiliza-se uma combinação de todos os possíveis problemas monorrótulo
 - É similar ao método todos-contra-todos, utilizado para dividir um problema multiclasse em um conjunto de problemas binários
- O método é reversível, permitindo a recuperação do problema multirrótulo original

Transformação Baseada nos Exemplos

Decomposição de rótulos (**Ranking by Pairwise Comparision (RPC)**): Método multiplicativo



X	Y: Metal-vs-Jazz
x ₁	Metal(1)-Jazz(0)
x ₂	Metal(0)-Jazz(1)
x ₃	Metal(0)-Jazz(1)
x ₄	Metal(1)-Jazz(0)
x ₅	Metal(0)-Jazz(1)

X	Y: Metal-vs-Bossa
x ₁	Metal(1)-Bossa(0)
x ₂	Metal(0)-Bossa(1)
x ₄	Metal(1)-Bossa(0)
x ₅	Metal(0)-Bossa(1)

X	Y: Metal-vs-Pop
x ₄	Metal(1)-Pop(0)
x ₅	Metal(0)-Pop(1)

X	Y: Jazz-vs-Bossa
x ₃	Jazz(1)-Bossa(0)

X	Y: Jazz-vs-Pop
x ₁	Jazz(0)-Pop(1)
x ₂	Jazz(1)-Pop(0)
x ₃	Jazz(1)-Pop(0)

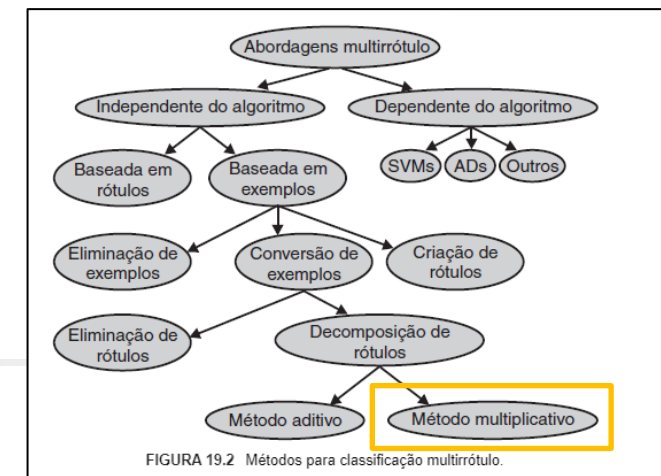
X	Y: Bossa-vs-Pop
x ₁	Bossa(0)-Pop(1)
x ₂	Bossa(1)-Pop(0)

X	Metal	Jazz	Bossa	Pop
x ₁	•			•
x ₂		•	•	
x ₃		•		
x ₄	•			
x ₅		•	•	•

Cada base derivada deve conter os objetos da base de treinamento que estejam associados ao rótulo l_i ou l_j , mas nunca objetos associados a ambos os rótulos

Transformação Baseada nos Exemplos

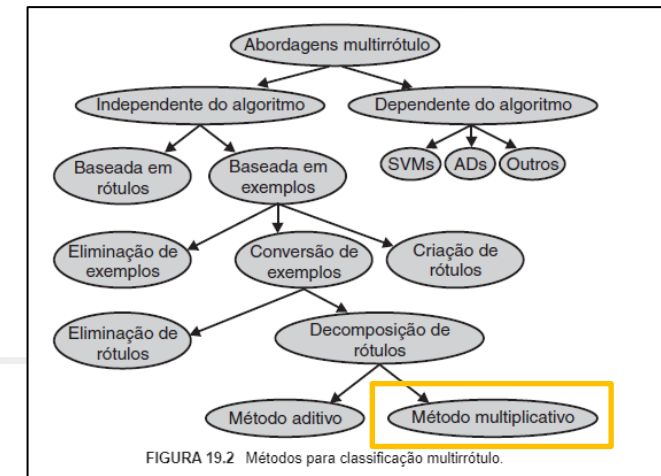
Decomposição de rótulos (**Ranking by Pairwise Comparision (RPC)**): Método multiplicativo



- Um novo objeto a ser classificado deve ser primeiro submetido a todos os classificadores binários
- O conjunto de rótulos final predito para o objeto será obtido através da aplicação de uma função de limiar capaz de separar os rótulos com maior número de vitórias daqueles com menor número

Transformação Baseada nos Exemplos

Decomposição de rótulos (**Ranking by Pairwise Comparision (RPC)**): Método multiplicativo



■ Vantagens:

- Consegue prever combinações de rótulos que não estão presentes na base de dados original
- Consegue incorporar implicitamente as correlações entre os rótulos no processo de construção do modelo de classificação

■ Desvantagens

- É dependente de uma função de limiar que, na prática, precisa ser calibrada de acordo com as diferentes bases de dados
- Pode atingir complexidade quadrática em termos de espaço e tempo, uma vez que, no pior caso, um total de $k(k-1)/2$ classificadores binários precisarão ser treinados e mantidos em memória
 - Todos precisam ser consultados no momento da classificação
 - Acaba sendo intratável em alguns problemas reais onde k é alto (k =nro. de classes)



Densidade e Cardinalidade de Rótulo

- O número de exemplos multirrótulo e classes de uma base de dados pode influenciar o desempenho dos métodos
- Os conjuntos de dados não são todos igualmente multirrotulados
- Em alguns casos, o número de classes de cada exemplo pode ser pequeno se comparado ao número total de exemplos, enquanto em outros, pode ser grande



Densidade e Cardinalidade de Rótulo

X = um conjunto de dados multirrótulo com n exemplos

$$CR(X) = \frac{1}{n} \sum_{i=1}^n sum(y_i).$$

A **cardinalidade** é dada pelo número médio de rótulos dos exemplos de X ($CR(X) = k \times DR(X)$) = frequência absoluta (exemplo: média de 3 rótulos por amostra)

Quantifica o número de rótulos alternativos que caracterizam os exemplos

$$DR(X) = \frac{1}{n} \sum_{i=1}^n \frac{sum(y_i)}{k}.$$

A **densidade** é dada pelo número médio de rótulos dos exemplos de X dividido por k (número total de classes) ($DR(X) = CR(X)/k$) = frequência relativa (exemplo: na média, 10% do total de rótulos por amostra)



Densidade e Cardinalidade de Rótulo


- Dois conjuntos com a mesma cardinalidade, mas com uma grande diferença no número de rótulos (diferentes densidades), podem apresentar propriedades diferentes, que podem afetar o desempenho dos algoritmos
- Por exemplo, dois conjuntos com a mesma cardinalidade de dois rótulos por objeto, mas com diferentes densidades:
 - (1) dois rótulos por objeto dentre quatro possíveis classes
 - (2) dois rótulos por objeto dentre 40 possíveis classes
- O número de possíveis combinações no segundo caso é bem maior que no primeiro



Propriedades das Bases de Dados Multirrótulo

- O desempenho dos diferentes métodos pode ser afetado pelas características da base de dados
- A tabela a seguir apresenta uma série de indicadores utilizados para fornecer uma razoável indicação das “propriedades multirrótulo” de uma base de dados


Propriedades das Bases de Dados Multirrótulo



Base de Dados	<i>N</i>	<i>q</i>	<i>LCard</i>	<i>LDens</i>	<i>NC</i>	<i>NU</i>	<i>NU/NC</i>	<i>NP</i>
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Número de objetos

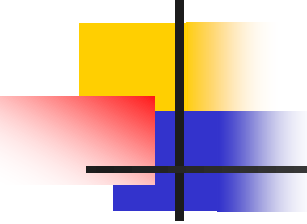
Propriedades das Bases de Dados Multirrótulo



Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Rótulos

Propriedades das Bases de Dados Multirrótulo




Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Maioria dos objetos está associada a apenas um rótulo

Cardinalidade = número médio de rótulos por objeto


Propriedades das Bases de Dados Multirrótulo



Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Densidade = $LCard/q$

Propriedades das Bases de Dados Multirrótulo



Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028


NC = Número de combinações
de rótulos distintas

Propriedades das Bases de Dados Multirrótulo

Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

NU = Número de combinações únicas,
i.e., o número de combinações de rótulos
que possuem frequência igual a 1

Propriedades das Bases de Dados Multirrótulo



Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

NU/NC = Proporção de labelsets únicos em relação ao número total de labelsets distintos

Propriedades das Bases de Dados Multirrótulo

Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

NP = número de pares RPC = fornece o número de classificadores binários que precisariam ser treinados para construir um classificador multirrótulo através do método RPC

Propriedades das Bases de Dados Multirrótulo

Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Método LP

X	Y
x_1	Metal-Pop
x_2	Jazz-Bossa
x_3	Jazz
x_4	Metal
x_5	Jazz-Bossa-Pop

Inviável para a maioria das bases de dados, pois teria que lidar com um grande número de classes compostas que ocorrem apenas uma vez, i.e., teria que lidar com apenas um objeto por classe composta

O número de combinações distintas de rótulos é superior a 30 na maioria das bases, o que pode ser considerado um número de rótulos de classes consideravelmente alto para um problema de classificação multiclasse

Muitas bases possuem um número alto de combinações que estão associadas a apenas um objeto

Algumas bases com valor superior a 50%, o que significa que entre o total de combinações de rótulos distintas, mais de 50% estão associadas a um único objeto

Propriedades das Bases de Dados Multirrótulo

Base de Dados	N	q	$LCard$	$LDens$	NC	NU	NU/NC	NP
Emotions	593	6	1,87	0,31	27	4	0,15	15
Scene	2.407	6	1,07	0,18	15	3	0,20	15
Flags	194	7	3,39	0,49	54	24	0,44	21
Yeast	2.417	14	4,24	0,30	198	77	0,43	91
POF-16	904	16	3,17	0,20	475	358	0,75	120
Birds	645	19	1,01	0,05	133	73	0,55	171
Genbase	662	27	1,25	0,05	32	10	0,31	350
Medical	978	45	1,25	0,03	94	33	0,35	984
Enron	1.702	53	3,38	0,06	753	573	0,76	1.378
Cal500	502	174	26,04	0,15	502	502	1,00	15.028

Métodos RPC x BR

Costuma ser computacionalmente viável para qualquer problema real de classificação

Indica que o número de classificadores binários que precisariam ser treinados e mantidos em memória para o método RPC é igual ou muito próximo ao limite superior de $q(q-1)/2$ para todas as bases de dados

Por outro lado, o método BR requer o treinamento de apenas q classificadores



Medidas de Avaliação

- Na classificação multirrótulo, um exemplo pode ser classificado de maneira parcialmente errada ou parcialmente correta. Situações:
 - Quando um classificador atribui corretamente a um exemplo pelo menos uma das classes a que ele pertence, mas também não atribui ao exemplo uma ou mais classes às quais ele pertence
 - Quando um classificador atribui a um exemplo uma ou mais classes às quais ele não pertence



Medidas de Avaliação

- Por exemplo, se os rótulos reais de uma música são “Pop” e “Jazz” e o classificador rotula essa música como “Pop” e “Bossa”, este resultado está parcialmente correto



Medidas de Avaliação

$$\text{acurácia}(\hat{f}, X) = \frac{1}{n} \sum_{i=1}^n \frac{y_i \text{ AND } z_i}{y_i \text{ OR } z_i}$$

Mede a proporção de rótulos classificados de maneira correta, levando em consideração resultados parcialmente corretos

Desvantagem: ela não é muito severa para penalizar predições binárias erradas (sejam falsos positivos ou falsos negativos)

	Y_i	Z_i	<i>Acurácia Multirrótulo</i>	<i>Comentários</i>
E ₁	l_1, l_2, l_3, l_4	l_1, l_2, l_3, l_4	1,00	Classificação perfeita
E ₂	l_1	l_2, l_3, l_4	0,00	Pior caso
E ₃	l_1	l_2	0,00	Pior caso
E ₄	l_1, l_3	l_1, l_2	0,33	
E ₅	l_1	l_1, l_2	0,50	
E ₆	l_1, l_2	l_1, l_2, l_3, l_4	0,50	Número maior de falsos positivos em relação à E ₅ , mas o valor continua sendo 0.50

Medidas de Avaliação

$$HL = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \Delta Z_i|}{q}$$

A expressão $|Y_i \Delta Z_i|$ representa a diferença simétrica entre Y_i e Z_i , a qual é equivalente à operação booleana XOR

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

Quanto menor for o valor do **Hamming Loss**, melhor é a classificação, sendo que a predição perfeita ocorre quando o seu valor é igual a zero

Mede o número médio de predições binárias incorretas por objeto

Embora mais adequada para penalizar predições binárias incorretas, a medida não diferencia o caso com um número maior de classificações corretas do caso que possui um menor número

	Y_i	Z_i	Acurácia Multirrótulo	Comentários
E ₁	l_1, l_2, l_3, l_4	l_1, l_2, l_3, l_4	0,00	Classificação perfeita $Y_i = [l_1, _, _, _]$
E ₂	l_1	l_2, l_3, l_4	1,00	Pior caso $\longrightarrow Z_i = [_, l_2, l_3, l_4]$
E ₃	l_1	l_2	0,50	
E ₄	l_1, l_3	l_1, l_2	0,50	Esta classificação é melhor do que a de E ₃ , mas o valor continua sendo 0.50
E ₅	l_1	l_1, l_2	0,25	
E ₆	l_1, l_2, l_3	l_1, l_2, l_3, l_4	0,25	Número maior de predições corretas em relação à E ₅ , mas o valor continua sendo 0.25

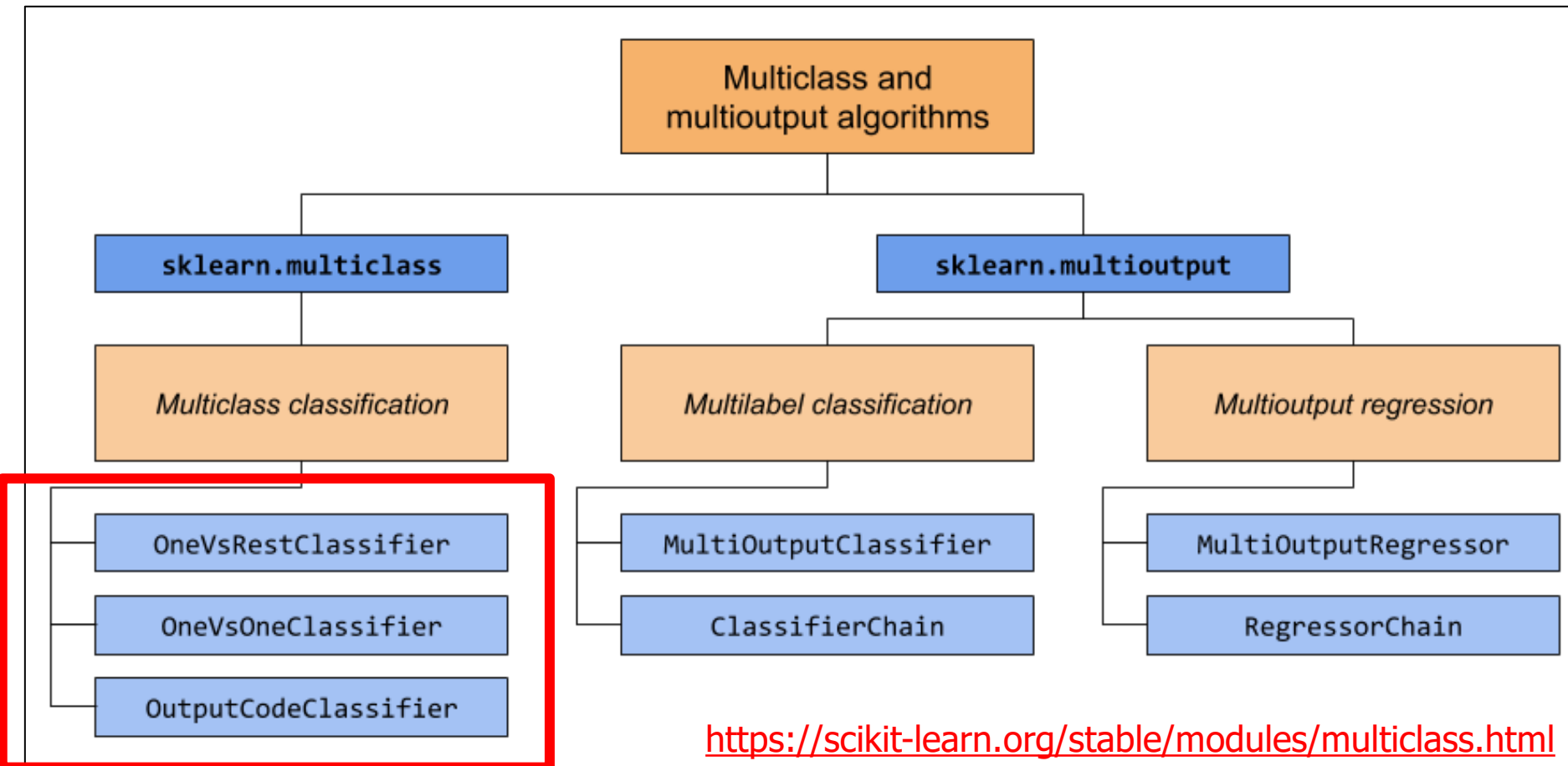


Medidas de Avaliação

- Independente de suas limitações, ambas as medidas são importantes, uma vez que fornecem informações complementares sobre o desempenho de um classificador multirrótulo
- Várias outras medidas são possíveis
 - As plataformas Mulan e Meka disponibilizam mais de vinte diferentes métricas aos seus usuários

SKLearn

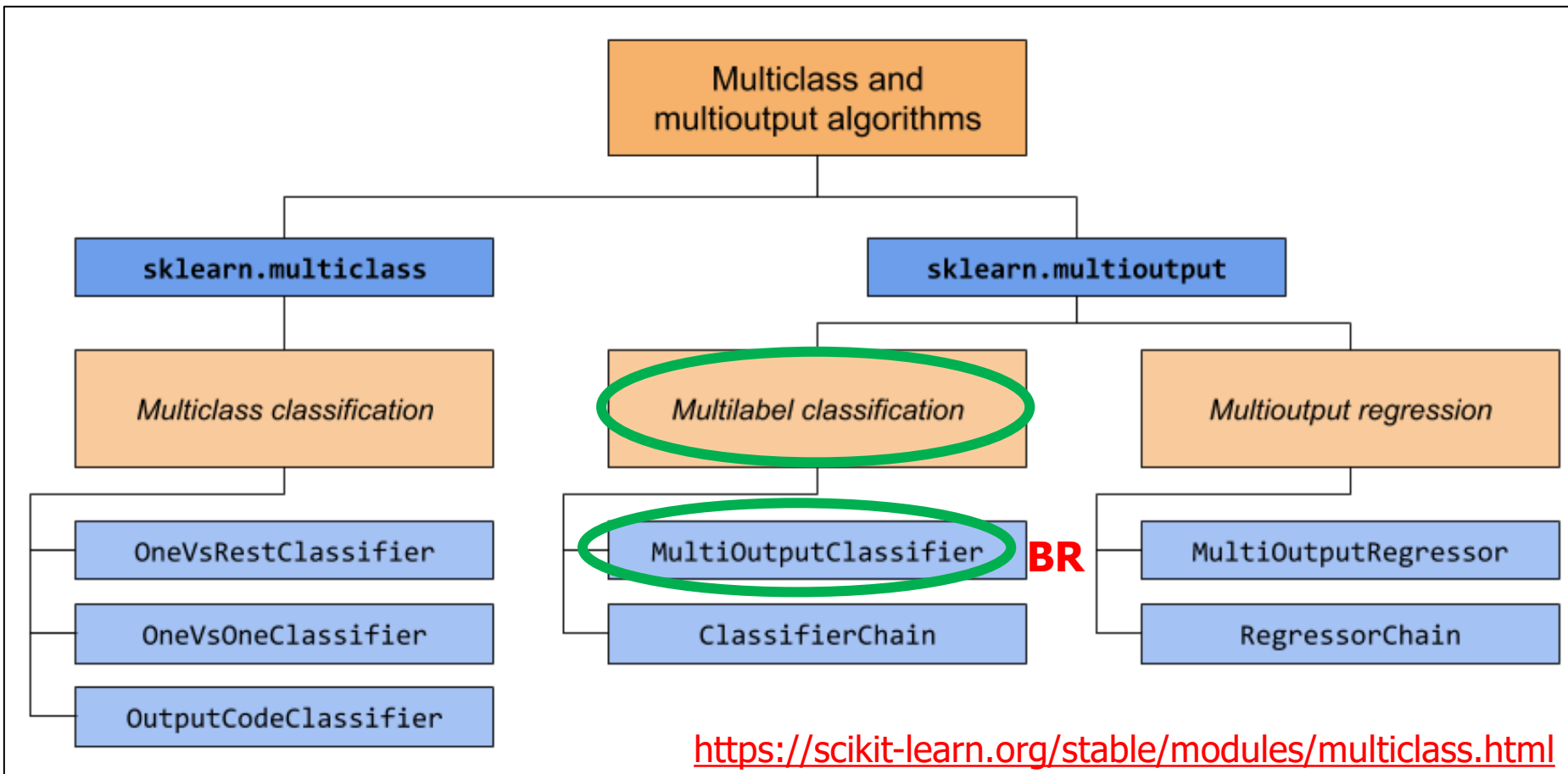
	Number of targets	Target cardinality	Valid type of target
Multiclass classification	1	>2	'multiclass'
Multilabel classification	>1	2 (0 or 1)	'multilabel-indicator'
Multiclass-multioutput classification	>1	>2	'multiclass-multioutput'
Multioutput regression	>1	Continuous	'continuous-multioutput'



<https://scikit-learn.org/stable/modules/multiclass.html>

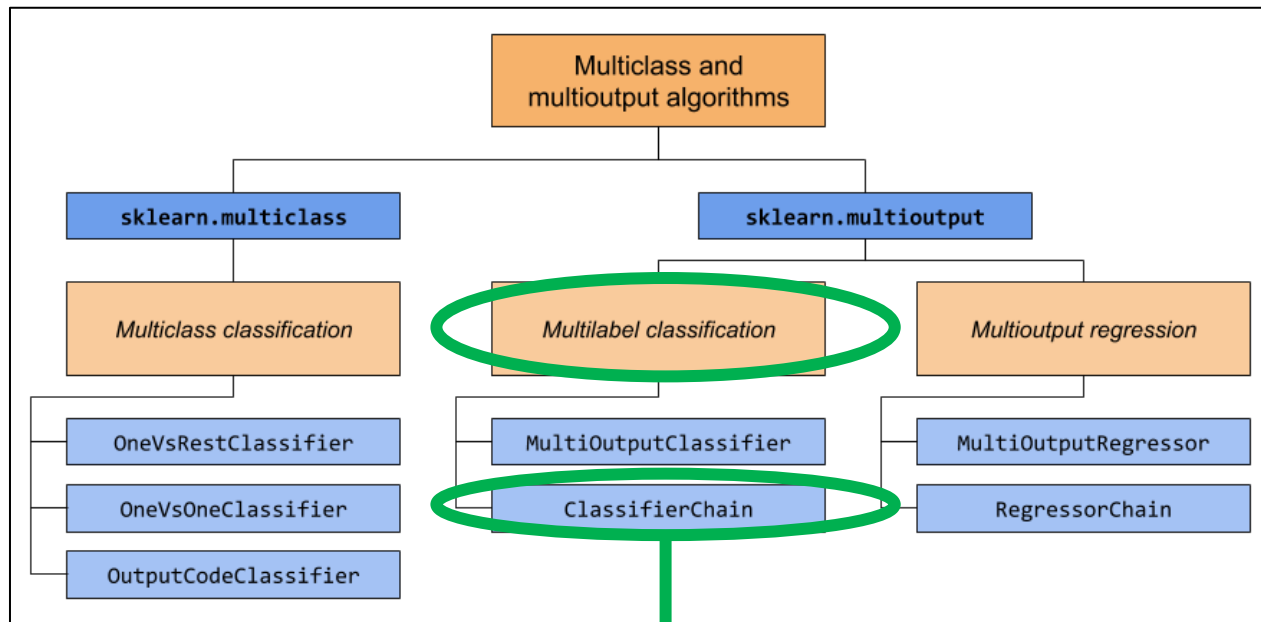
SKLearn

	Number of targets	Target cardinality	Valid type_of_target
Multiclass classification	1	>2	'multiclass'
Multilabel classification	>1	2 (0 or 1)	'multilabel-indicator'
Multiclass-multioutput classification	>1	>2	'multiclass-multioutput'
Multioutput regression	>1	Continuous	'continuous-multioutput'



SKLearn

	Number of targets	Target cardinality	Valid type_of_target
Multiclass classification	1	>2	'multiclass'
Multilabel classification	>1	2 (0 or 1)	'multilabel-indicator'
Multiclass-multioutput classification	>1	>2	'multiclass-multioutput'
Multioutput regression	>1	Continuous	'continuous-multioutput'

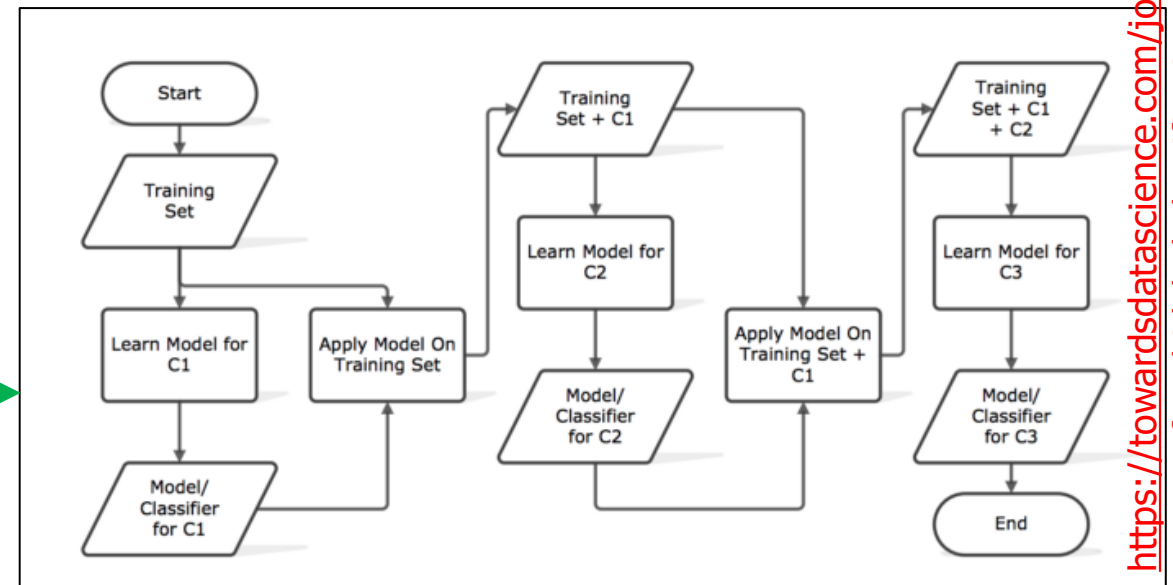


<https://scikit-learn.org/stable/modules/multiclass.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.ClassifierChain.html#sklearn.multioutput.ClassifierChain>

https://link.springer.com/content/pdf/10.1007/978-3-642-04174-7_17.pdf [Classifier Chains for Multi-label Classification]

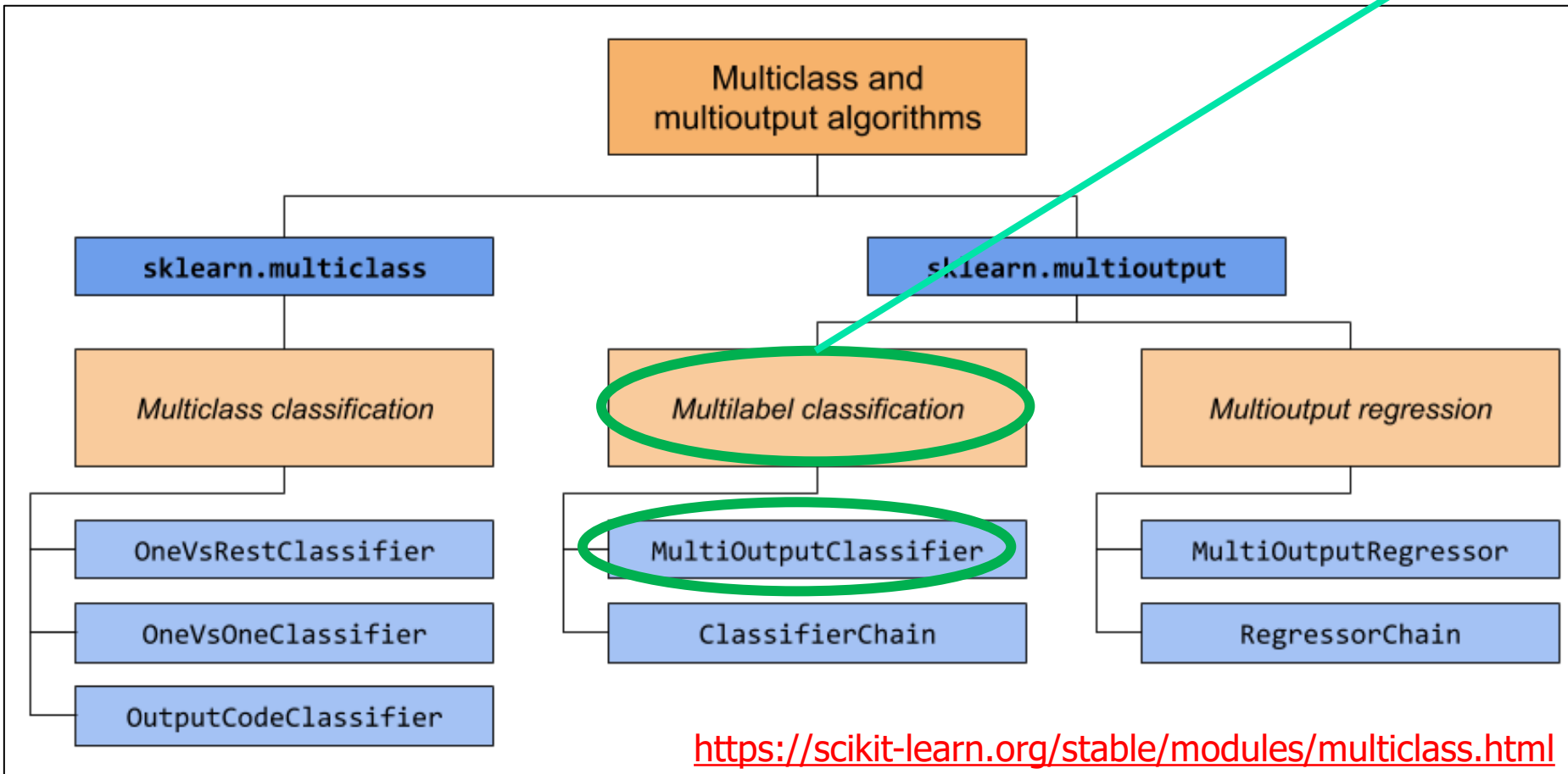
Classifier Chains



<https://towardsdatascience.com/journey-to-the-center-of-multi-label-classification-384c40229bff>

SKLearn

	Number of targets	Target cardinality	Valid type_of_target
Multiclass classification	1	>2	'multiclass'
Multilabel classification	>1	2 (0 or 1)	'multilabel-indicator'
Multiclass-multioutput classification	>1	>2	'multiclass-multioutput'
Multioutput regression	>1	Continuous	'continuous-multioutput'



<https://scikit-learn.org/stable/modules/multiclass.html>

Multitask classification

Example: properties “type of fruit” (“apple”, “pear” and “orange”) and “colour” (“green”, “red”, “yellow” and “orange”) for a set of images of fruit

Each sample is an image of a fruit, a label is output for both properties and each label is one of the possible classes of the corresponding property



Links e referências

- Meka
 - <http://waikato.github.io/meke/>
- Mulan
 - <http://mulan.sourceforge.net/>
- Multi-Label Classification in Python
 - <http://scikit.ml/>
- Introdução à Classificação Multirrotulo
 - <https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/7/11/36-1?inline=1>

