



UFOP

Python HRSVM Manual

Guilherme Oliveira Santos

17 de outubro de 2023

icea
Instituto de Ciências
Exatas e Aplicadas

DECSI
DEPARTAMENTO DE
COMPUTAÇÃO E SISTEMAS



UFOP

Sumário

1 Hierarchical Multi-Label Classification HRSVM

► Hierarchical Multi-Label Classification HRSVM

► Bibliography

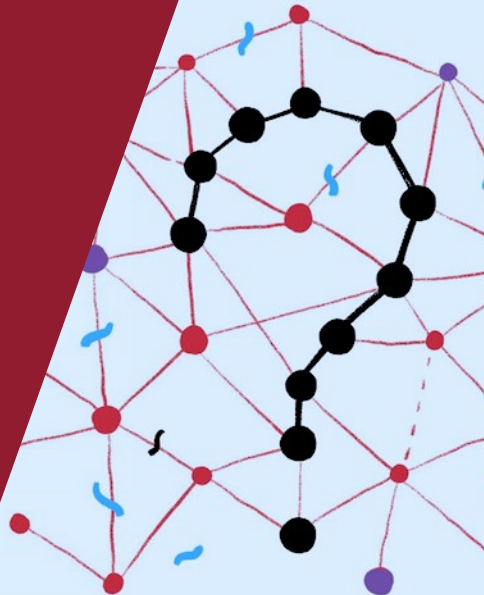


UFOP

Conteúdo

1 Hierarchical Multi-Label Classification HRSVM

- Dataset Preparation
- Parameter Configuration
- System Execution in Python





Binary and Multi-class Dataset

1 Hierarchical Multi-Label Classification HRSVM

- Cada linha contém uma parte de classe e uma parte de características, seguindo a sintaxe:

- `<class> <feature>:<value> <feature>:<value> ...`

Cada linha consiste em 2 partes:

- A parte da classe, que descreve as classes às quais o exemplo pertence.
- A parte de característica, que descreve o valor de cada característica no formato "X:Y" separado por espaço, onde X é o identificador da característica e Y é o valor da característica. Características não especificadas serão tratadas como zero.



Binary and Multi-class Dataset

1 Hierarchical Multi-Label Classification HRSVM

- Exemplos:

spam 1:0.0024 5:0.2141

non-spam 1:1.234 2:0.1221 3:0.11 4:-0.12 5:0.7903

- O primeiro exemplo pertence à classe 'spam', onde a característica 1 tem um valor de 0.0024 e a característica 5 tem um valor de 0.2141. As características não especificadas são 2, 3 e 4, que todas têm um valor de 0.
- O segundo exemplo pertence à classe 'não spam', onde as características 1, 2, 3, 4 e 5 têm valores de 1.234, 0.1221, 0.11, -0.12 e 0.7903, respectivamente.



Multi-Label Dataset

1 Hierarchical Multi-Label Classification HRSVM

- Para um conjunto de dados multirrótulo, o formato é semelhante ao conjunto de dados binário, com a diferença de que a parte da classe pode ter várias classes, separadas por vírgula, seguindo a sintaxe:
 - `<class>,<class>,... <feature>:<value> <feature>:<value> ...`

Cada linha descreve um exemplo de treinamento. A parte da classe pode conter várias classes separadas por vírgula. A sintaxe da parte de características é idêntica à do conjunto de dados de classificação binária.



Multi-Label Dataset

1 Hierarchical Multi-Label Classification HRSVM

- Exemplo:

`action,comedy,romantic 1:0.0024 5:0.2141`

- A figura acima indica que este exemplo pertence às classes "ação", "comédia" e "romântico". A característica 1 tem um valor de 0.0024 e a característica 5 tem um valor de 0.2141. As características não especificadas tem 0.



UFOP

Hierarchical Multi-Label Dataset

1 Hierarchical Multi-Label Classification HRSVM

Para a classificação hierárquica multirrótulo, são necessários dois arquivos:

- um arquivo de dados e uma descrição da hierarquia
- O arquivo de dados/características, que possui o mesmo formato da classificação multirrótulo

A descrição da hierarquia descreve a estrutura hierárquica das classes e segue um formato semelhante ao mostrado a seguir:

• `<parent>` `<child>`



UFOP

Parameter Configuration

1 Hierarchical Multi-Label Classification HRSVM

A interface deste sistema é via linha de comando, é facilmente configurável usando várias opções e parâmetros correspondentes.

Option	Parameter	Description
		num_features is the number of features of dataset.
-r*	coef0	Set coef0 in kernel function (default 0)
-c*	cost	Set the parameter C of C-SVC (default 1)
-m	cache_size	Set cache memory size in MB (default 100)
-e	epsilon	Set tolerance of termination criterion (default 0.001)
-v	n	n-fold cross validation mode The system randomly splits the data into n parts and calculates cross validation accuracy on them.
-X	n	specify random seed in k-fold cross validation mode (default: random)

Option	Parameter	Description
-t*	kernel_type	Set type of kernel function (default 2) 0 linear 1 polynomial 2 radial basis function 3 sigmoid
-k*	classification_type	Set the classification type (default binary) binary binary classification multiclass multi-class classification multilabel multi-label classification hierarchical <hierarchical file> hierarchical classification Note that in case of hierarchical mode, you need to specify hierarchical file next to this option.
-a	enable_R-SVM	Enable/disable R-SVM algorithm (default 0) 0 disable 1 enable
-f	enable_feature_selection	Enable/disable feature selection algorithm (default 0) 0 disable 1 enable
-l	minimum_example_in_class	Set the threshold which program will remove classes that have number of examples lower than threshold (default 0)
-d*	degree	Set degree in kernel function (default 3)
-g*	gamma	Set gamma in kernel function (default 1/num_features)

'*' significa será automaticamente configurada no modo de classificação hierárquica



System Execution in Python

1 Hierarchical Multi-Label Classification HRSVM

Após concluir a preparação do conjunto de dados e a definição dos parâmetros, o próximo passo é a execução do sistema.

- Por meio da linha de comando (Windows);
- Terminal (Linux);
- Ou pelas funções implementadas em python.

A figura abaixo fornece uma visão geral do processo de treinamento em python:

```
#Hierarchical train example:

parameters = "-k hierarchical ""HRSVM-PY/hierarchicalExample/train.hf"" -a 1 -l 50"
trainPath = "HRSVM-PY/hierarchicalExample/train.dat"
predictModelPath = "HRSVM-PY/hierarchicalExample/predictModel"
#svmTrain(parameters, trainPath, predictModelPath)
```

Primeiro, o usuário deve fornecer os parâmetros de treinamento, o caminho do arquivo de treinamento, e a saída do modelo de classificação.



System Execution in Python

1 Hierarchical Multi-Label Classification HRSVM

Se você deseja executar o sistema no modo hierárquico, também deve fornecer o caminho para a descrição da hierarquia:

```
#Hierarchical train example:  
  
parameters = "-k hierarchical \"HRSVM-PY/hierarchicalExample/train.hf\" -a 1 -l 50"  
trainPath = "HRSVM-PY/hierarchicalExample/train.dat"  
predictModelPath = "HRSVM-PY/hierarchicalExample/predictModel"  
#svmTrain(parameters, trainPath, predictModelPath)
```

O caminho da descrição da hierarquia é "HRSVM-PY/hierarchicalExample/train.hf". O R-SVM está habilitado. As classes que contêm exemplos inferiores a 50 serão removidas. O caminho dos dados de treinamento esta em 'trainPath'. A pasta do classificador será criada em 'predictModelPath'.



System Execution in Python

1 Hierarchical Multi-Label Classification HRSVM

Após a conclusão do processo de treinamento, o sistema produzirá um classificador na forma de uma pasta. A pasta do classificador contém um classificador para cada classe no conjunto de treinamento.

- Para o processo de previsão, o usuário deve fornecer o caminho dos dados a serem previstos.
- O caminho do classificador.
- O caminho de saída da previsão.

```
PredictData = "HRSVM-PY/hierarchicalExample/test.dat"  
classifierPath = "HRSVM-PY/hierarchicalExample/predictModel"  
#svmPredict(PredictData, classifierPath, "HRSVM-PY/hierarchicalExample/predict")
```



System Execution in Python

1 Hierarchical Multi-Label Classification HRSVM

O exemplo acima indica que "svm-predict()", fará previsões no 'PredictData', usando o classificador em 'classifierPath', com output no path "HRSVM-PY/hierarchicalExample/predict".

O sistema fornecerá três saídas:

- O arquivo "SVM-score", que contém os escores SVM de cada classe em formato CSV.
- O arquivo "resultado da previsão", que contém os rótulos de classe que o sistema previu para cada exemplo.
- O arquivo "relatório", que contém as medições de desempenho de cada classe e um resumo de todas as medições. Observe que este arquivo pode variar dependendo do tipo de classificação.



UFOP

Sumário

2 Bibliography

► Hierarchical Multi-Label Classification HRSVM




► Bibliography



UFOP

Referências Bibliográficas

2 Bibliography

-  Manish Thapliyal. (2021). A Beginner's Guide to Hierarchical Classification. Medium.
<https://medium.com/@manish54.thapliyal>
-  Noa Weiss. (2021). The Hitchhiker's Guide to Hierarchical Classification. Towards Data Science.
<https://towardsdatascience.com/>
-  Jump Thanawut. (Jan 15, 2016). HRSVM-Research GitHub Repository. GitHub.
<https://github.com/JumpThanawut/HRSVM-Research/tree/master>



UFOP

Python HRSVM Manual *Obrigado!*