Multi-label Classification
1 Temmuz 19

Multi-label classification is a type of classification where the instances can have a set of labels. These multiple labels for an instance can be inter-dependent, which makes it hard to make reductions in the solution approaches. We can define multi-label classification more formally such that $X = \{x_i, y_i\}$, the set of instances where x_i is the feature vector and y_i is the set of labels for the ith instance for $i=1..N$; $L = \{l_j\}$ is the set of possible labels for $j=1..M$. The sizes of y_i's are not necessarily equal; one instance might have three labels while the other can have one. Y_i's can also be represented as vectors with 0 or 1 values, of size M. The classifier trained on X can predict a set of labels for a given test example based on the training examples. We should see that training set should contain multi-labelled instances, otherwise the classifier may not work properly.
An example for a set of multi-labelled instances can be given as X = { ("I am hungry", [negative, personal]), ("I am happy", [positive, personal]), ("She is smiling", [positive, not-personal], ("They are far away", [negative, not-personal])}. The classifier should predict two separate and not quite correlated classes in this example. One classifier for sentiment classification and one other classifier for personality detection can easily be combined and work very well for further predictions. Suppose we have a set of documents {(doc1, [scientific, serious]), (doc2, [comics, loose], (doc3, [ultimatom, serious])))}. Training two separate classifiers for the dimensions domain and severity level would result in losing the clues that the two give about each other. Adding one of the dimensions (classes) in the feature vectors for the classification of the other again might cause cancelling this information and indeed it may not be the case for some datasets to have equal numbers of labels for all instances.

It is a need in some text classification applications to make use of multiple labels for textual units; document classification, keyphrase extraction, topic detection can be example cases. It is also widely used in image classification tasks such as image annotation and scene classification.

We would also like to indicate that the sum of label assignment probabilities for an instance may not be exactly 1 if label ranking is applied.

There have been many efforts to find good classifiers for multi-labelled datasets in the literature. We touch upon some of them below.

Approaches

Classical and neural models are different in their approach to the problem. In the classical learning case, the approaches that work on dataset are more common while in neural models the algorithms are more central to adapt to the problem of multi-label classification.

We mention some of the popular methods in the classical domain:

- Binary relevance: The labels for all instances are treated separately and |L| many binary classifiers are trained for L as the set of labels, proposed in [1] and [3] and later widely used. With a good binary classifier like SVM it gives very good results though label correlations are given away.
- Label reduction: One label is selected out of the multiple labels for each instance and single label classification is performed. Thus, multi-label classification is removed completely. Label selection is done either randomly or based on the contribution / probability / complexity [2] of the assigned classes.
- Instance copying: New single label datasets are generated as many as the combinations of multiple labels in the original dataset and the results of single-label classification on each dataset are combined [3].
- Label combination: The multiple labels are named to be a new class and single-label classification is performed [4].
- Classifier chain: A chain of M binary classifiers is constructed, with each classifier making use of the label information from the previous classifier in connection [5]. The labels are used as features in a chain.

As for neural approaches, we can say that many new models with special parameters, optimization methods have been proposed. The initial models are mostly centered on the problem of using the output layer of the neural networks as a multi-label classifier. One of the widely known efficient models proposed, called BP-MLL, is a single-layer feed forward neural network with a modified error function such that it consider all the units in the output layer where each unit is for one label [6]. This way not a single label but multiple labels are considered in learning along with their interdependencies. An extension to this model is proposed with an additional classifier after its output unit to select the most probable set of labels and is shown to perform better than the previous one [7].
Label embeddings have been widely studied with an emphasis on reduced computational cost [8, 9, 10]. This method is said to be more suitable for datasets with large sizes of label sets, comparable to instance set size. Representing multiple labels for each instance as vector, the model compresses this output vector or label vector so that both the interdependencies in labels and their relation to instances are represented in a shrinked subspace which helps performing more efficient training and prediction in terms of complexity and correctness.

There are also multi-label adapted deep learning models with modifications to CNN and RNN architectures [11,12, 13, 14]. Deep models are suitable for large datasets with large numbers of labels.

Datasets

A number of datasets can be found for experimental or real world use on

- http://mulan.sourceforge.net/datasets-mlc.html

Evaluation measures

For multi-label classification, evaluation is considered in two distinct approaches; ranking-based measures evaluates real valued label scores as well as their ranks and bipartition methods evaluates exact fits, i.e., classical f-score measures. Rank loss, one-error, coverage, average precision are the most commonly used metrics within the frame of ranking-based evaluation. Micro or macro averaged precision, recall and f1-score are the most widely used ones in bipartition evaluation methods.

Implementations available

Considering possible tasks and dataset scope, using one of the classical algorithms like binary relevance with SVM or BP-MLL will give good results.

- Scikit-multilearn: A library based on scikit-learn; presents implementations of a number of multi-label classifiers, mostly classical available at http://scikit.ml/
- BP-MLL: Two different implementations using tensorflow and one other in matlab are available
  - https://github.com/hinanmu/BPMLL
  - https://github.com/vanHavel/bp-mll-tensorflow
  - (matlab code) http://lamda.nju.edu.cn/code_BPMLL.ashx?AspxAutoDetectCookieSupport=1

References

[1] Karalič, A., and V. Pirnat. "Significance level based multiple tree classification." *Informatica* 15.5 (1991): 12.
[2] Godbole, Shantanu, and Sunita Sarawagi. "Discriminative methods for multi-labeled classification." *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, Heidelberg, 2004.
[3] Shen, Xipeng, et al. "Multilabel machine learning and its application to semantic scene classification." *Storage and Retrieval Methods and Applications for Multimedia 2004*. Vol. 5307. International Society for Optics and Photonics, 2003.
[4] Cheng, Weiwei, and Eyke Hüllermeier. "Combining instance-based learning and logistic regression for multilabel classification." *Machine Learning* 76.2-3 (2009): 211-225.
[5] Cheng, Weiwei, Eyke Hüllermeier, and Krzysztof J. Dembczynski. "Bayes optimal multilabel classification via probabilistic classifier chains." *Proceedings of the 27th international conference on machine learning (ICML-10)*. 2010.
[6] Zhang, Min-Ling, and Zhi-Hua Zhou. "Multilabel neural networks with applications to functional genomics and text categorization." *IEEE transactions on Knowledge and Data Engineering* 18.10 (2006): 1338-1351.
[7] Nam, Jinseok, et al. "Large-scale multi-label text classification—revisiting neural networks." *Joint european conference on machine learning and knowledge discovery in databases*. Springer, Berlin, Heidelberg, 2014.
[8] Hsu, Daniel J., et al. "Multi-label prediction via compressed sensing." *Advances in neural information processing systems*. 2009.
[9] Lin, Zijia, et al. "Multi-label classification via feature-aware implicit label space encoding." *International conference on machine learning*. 2014.
[10] Bhatia, Kush, et al. "Sparse local embeddings for extreme multi-label classification." *Advances in neural information processing systems*. 2015.
[11] Wei, Yunchao, et al. "Cnn: Single-label to multi-label." *arXiv preprint arXiv:1406.5726* (2014).
[12] Wang, Jiang, et al. "Cnn-rnn: A unified framework for multi-label image classification." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[13] Chen, Guibin, et al. "Ensemble application of convolutional and recurrent neural networks for multi-label text categorization." *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017.
[14] Gargiulo, Francesco, Stefano Silvestri, and Mario Ciampi. "Deep Convolution Neural Network for Extreme Multi-label Text Classification." *HEALTHINF*. 2018.