Multi label / multi classification learning

Menu

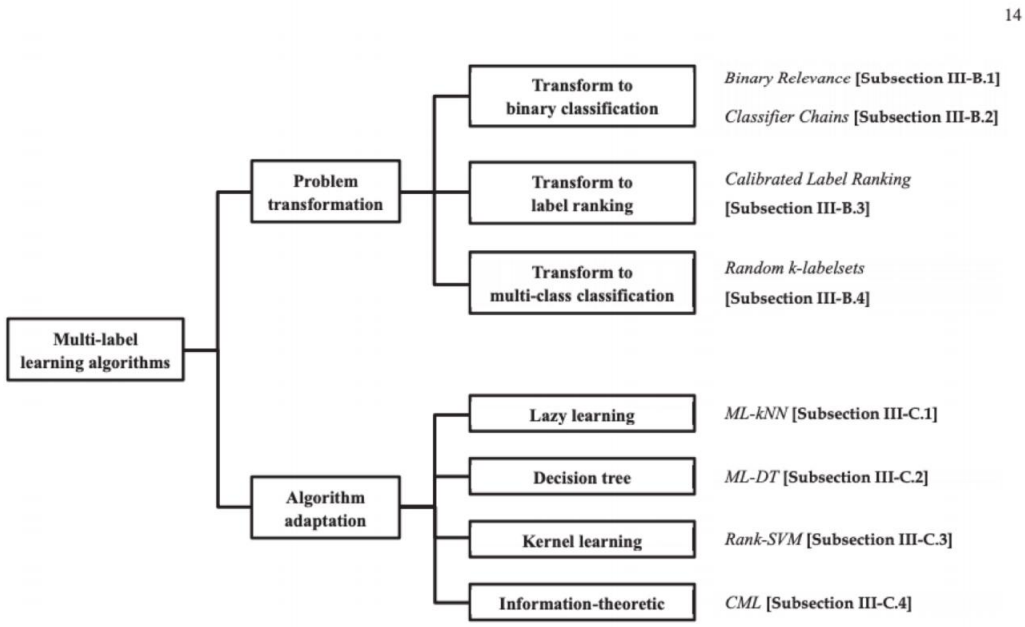
1 definition

Multi label problem: each label is not mutually exclusive (eg: movie labels are comedies, action films and Mandarin films)

Multi classification problem: each category is mutually exclusive (eg: education background is primary school or junior high school or high school or university).

It can be seen that multi classification problem is contained in multi label problem.

2 framework



Reference： A review on multi-label learning algorithms --2014 TKDE(IEEE Transactions on Knowledge and Data Engineering)
https://www.researchgate.net/publication/263813673_A_Review_On_Multi-Label_Learning_Algorithms?enrichId=rgreq945f5f0a8f6325d0f93a02d440de1366-XXX&enrichSource=Y292ZXJQYWdlOzI2MzgxMzY3MztBUzozMDE4MzQzMDEzOTQ5NTFAMTQ0ODk3NDMxMzMyMw%3D%3D&el=1_x_3&_esc=publi

3 Algorithmic thinking

Regardless of multi label or multi classification, the algorithm ideas can be divided into three categories:

3.1 First-order strategy (First level strategy)

Method: one vs rest:

(ignore the correlation with other tags, such as decomposing multiple tags into multiple independent binary classification problems)
Advantages: simple and efficient.
Disadvantage: ignoring the correlation between tags, the result may not be good

3.2 Second-order strategy (Second level strategy)

Method: one vs one:
(consider the pairwise association between tags. For example: 1. Between relevant label and independent label; 2. Any label pair)
Advantages: considering the relevance of tags, good generalization ability.
Disadvantage: only one vs one correlation can be represented

3.3 High-order strategy (Advanced strategy)

Method: multi vs multi
(consider the relationship between multiple tags, such as the influence of all other tags on each tag)
Advantage: the model considers the association between tags
Disadvantages: large amount of calculation, not scalable

4 Multi label / multi classification general solutions

4.1 problem transformation

4.1.1 Binary classification

1. Multiple binary classification
2. Chain of classification

In this case, the first classifier only trains on the input data, and then each classifier trains on all previous classifiers in the input space and chain.
Advantages: consider the association between multiple tags, so is **Advanced strategy**
Disadvantages: the quality of the algorithm is affected by the order of the chain, which can be solved by random sequence; the parallel calculation is missing, because the chain call is needed

4.1.2 Ranking method:

4.1.2.1 Calibrated Label Ranking

The basic idea of the algorithm is to transform the multi label learning problem into the label sorting problem;

Compare two tags to get a sort list

Advantages: solve the problem of imbalance between classes, consider the relationship between two labels, **two-level strategy**

Disadvantages: high complexity, more classifiers

4.1.3 Multi classification method:

4.1.3.1 LP (label powerest)

Map back to the label set according to the natural number of output

Map 2q2 ^ q2q, the set of possible labels, to 2q2 ^ q2q natural numbers.

Disadvantages: the generalization ability is low, and only the seen categories can be solved; the categories are too large and inefficient

4.1.3.2 Random k-Labelsets

Randomly divide subsets of length k, shrink sample space, construct n subsets, and integrate n trainers (voting method)

Advantages: three level strategy

4.1.3.3 Error Correcting Output Codes (ECOC)

Advantages: Advanced strategy

Disadvantages: too complex model and too many classifiers

4.2 Algorithm adaptation

4.2.1 Multi-Label k-Nearest Neighbor (ML-KNN)

Advantages: simple and fast; robust noise loss data through K selection

Disadvantages: large storage cost and large sample uneven influence; first level strategy

4.2.2 Multi-Label Decision Tree (ML-DT)

First, calculate the information gain IG of each feature, select the feature with the largest IG to divide the sample into left and right subsets, and recurse until the stop condition is met (for example, the number of subsets in the leaf node is 100)

At the end, for unknown samples, a path is traversed along the root node to the leaf node, and the probability of each label 0 and 1 in the leaf node sample subset is calculated. The label with a

probability of more than 0.5 is defined as the unknown sample label.
Disadvantages: first level strategy

4.2.3 Ranking Support Vector Machine (Rank-SVM)

Advantages: the hyperplane of "correlation-uncorrelation" label pair is defined, the relationship between two labels considered, the second level strategy

5. Existing multi classification / multi label classifiers

Reference: https://sklearn.apachecn.org/docs/0.21.3/13.html

5.1 Classifier only for multi classification problems

5.1.1 One vs one multi class classifier

(Problem transformation) (Second level strategy)
sklearn.svm.NuSVC
sklearn.svm.SVC.
sklearn.gaussian_process.GaussianProcessClassifier (setting multi_class = “one_vs_one”)

5.1.2 One vs rest multi class classifier

(Problem transformation) (Second level strategy)
sklearn.ensemble.GradientBoostingClassifier
sklearn.gaussian_process.GaussianProcessClassifier (setting multi_class = “one_vs_rest”)
sklearn.svm.LinearSVC (setting multi_class=”ovr”)
sklearn.linear_model.LogisticRegression (setting multi_class=”ovr”)
sklearn.linear_model.LogisticRegressionCV (setting multi_class=”ovr”)
sklearn.linear_model.SGDClassifier
sklearn.linear_model.Perceptron
sklearn.linear_model.PassiveAggressiveClassifier
XGBOOST

5.2 General classifiers for multi label / multi classification problems

(Algorithm adaptation)
sklearn.tree.DecisionTreeClassifier
sklearn.tree.ExtraTreeClassifier

sklearn.ensemble.ExtraTreesClassifier
sklearn.neighbors.KNeighborsClassifier
sklearn.neural_network.MLPClassifier
sklearn.neighbors.RadiusNeighborsClassifier
sklearn.ensemble.RandomForestClassifier
sklearn.linear_model.RidgeClassifierCV