Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Index

• Introduction
  • Multilabel Classification
  • Cross-Validation and Stratified Cross-Validation
• Methods and Experimentation
  • Genetic Algorithms
  • Mulan, Weka, Data Sets
• Results
• Conclusion
• Future Lines
• References

Juan A. Fernández del Pozo
Pedro Larrañaga
Concha Bielza

Computational Intelligence Group
Universidad Politécnica de Madrid

7-8/02/2013
TIN2010-20900-C04
Albacete
Multi-label Learning is a form of supervised learning where the classification algorithm is required to learn from a set of instances, each instance can belong to multiple classes and so after be able to predict a set of class labels for a new instance.

This is a generalized version of most popular multi-class problems where each instances is restricted to have only one class label.

There exists a wide range of applications for multi-labelled predictions, such as text categorization, semantic image labeling, gene functionality classification etc. and the scope and interest is increasing with modern applications.
Given a training set, \( S = (x_i, Y_i), 1 \leq i \leq n \), consisting \( n \) training instances, \( (x_i \in X, Y_i \in Y) \) i.i.d drawn from an unknown distribution \( D \), the goal of the multi-label learning is to produce a multi-label classifier

\[
h : X \rightarrow Y \text{ (in other words, } h X \rightarrow 2^L \text{ )}
\]

that optimizes some specific evaluation function
Introduction

- Simple Problem Transformation Methods
  - Label Powerset (LP)
  - Binary Relevance (BR)
  - Ranking by Pairwise Comparison (RPC)
  - Calibrated Label Ranking (CLR)

- Simple Algorithm Adaptation Methods
  - Tree Based Boosting
  - Lazy Learning
  - Discriminative SVM Based Methods

- Dimensionality Reduction and Subspace Based Methods
  - Shared Subspace

- Ensemble Methods
  - Random k labelsets (RakEL),
  - Pruned Sets, Random Decision Tree (RDT)

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Problem Variations

- Learning with Multiple Labels: Disjoint Case
- Multitask Learning
- Multi-Instance Multi-Label Learning (MIML)
Evaluation Metrics

- Accuracy
- Precision, Recall, F-measure, and ROC area

Prediction
- fully correct, partially correct or fully incorrect.

Target problem

- evaluating partitions,
- evaluating ranking and
- using label hierarchy
Introduction

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

**ExactMatchRatio**, \( MR = \frac{1}{n} \sum_{i=1}^{n} I(Y_i = Z_i) \)

**Precision**, \( P = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i \cap Z_i|}{|Z_i|} \)

**Accuracy**, \( A = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|} \)

**F1**, \( F_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{2|Y_i \cap Z_i|}{|Y_i| + |Z_i|} \)

**Recall**, \( R = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i \cap Z_i|}{|Y_i|} \)

**HammingLoss**, \( HL = \frac{1}{kn} \sum_{i=1}^{n} \sum_{l=1}^{k} [I(l \in Z_i \land l \notin Y_i) + I(l \notin Z_i \land l \in Y_i)] \)
**Introduction**

**Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms**

### Label-based

**Macro Averaged Measures:**

\[
\lambda - \text{Precision}, P_{macro}^\lambda = \frac{\sum_{i=1}^{n} Y_i^\lambda Z_i^\lambda}{\sum_{i=1}^{n} Z_i^\lambda}
\]

\[
\lambda - \text{Recall}, R_{macro}^\lambda = \frac{\sum_{i=1}^{n} Y_i^\lambda Z_i^\lambda}{\sum_{i=1}^{n} Y_i^\lambda}
\]

\[
\lambda - F_{1-macro}^\lambda = \frac{2 \sum_{i=1}^{n} Y_i^\lambda Z_i^\lambda}{\sum_{i=1}^{n} Y_i^\lambda + \sum_{i=1}^{n} Z_i^\lambda}
\]

\[
\text{Precision, } P_{macro} = \frac{1}{k} \sum_{i=1}^{k} P_{macro}^\lambda
\]

\[
\text{Recall, } R_{macro} = \frac{1}{k} \sum_{i=1}^{k} R_{macro}^\lambda
\]

\[
F_{1-macro} = \frac{1}{k} \sum_{i=1}^{k} F_{macro}^\lambda
\]

**Micro Averaged Measures:**

\[
\text{Precision, } P_{micro} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k} \sum_{i=1}^{n} Z_i^j}
\]

\[
\text{Recall, } R_{micro} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k} \sum_{i=1}^{n} Y_i^j}
\]
**Introduction**

*Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms*

---

**One-error, $O = \frac{1}{n} \sum_{i=1}^{n} I(\text{arg min}_{\lambda \in \mathcal{L}} r_i(\lambda) \notin Y_i^l)***

**Coverage, $C = \frac{1}{n} \sum_{i=1}^{n} \max_{\lambda \in Y_i} r_i(\lambda) - 1***

**Ranking Loss, $RL = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|Y_i^l|} |(\lambda_a, \lambda_b) : r_i(\lambda_a) > r_i(\lambda_b), (\lambda_a, \lambda_b) \in Y_i^l \times \bar{Y}_i^l|***

**Average Precision, $AP = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|Y_i^l|} \sum_{\lambda \in Y_i^l} \frac{|\{\lambda' \in Y_i^l : r_i(\lambda') \leq r_i(\lambda)\}|}{r_i(\lambda)}***

**Hierarchical Loss, $H-Loss = \frac{1}{m} \sum_{i=1}^{n} |\lambda| \lambda \in (Y_i \bowtie Z_i) \land \text{Ancestor}(\lambda) \cap (Y_i \bowtie Z_i) = \emptyset|***
Multi-label Datasets and Statistics

- Distinct Label Set (DL)
  \[ DL = |\{Y\}| \exists x : (x, Y) \in S \]

- Proportion of Distinct Label Set (PDL)
  \[ PDL = \frac{DL}{|S|} \]

- Label Cardinality (Lcard)
  \[ LCard = \frac{1}{n} \sum_{i=1}^{n} |Y_i| \]

- Label Density (LDen)
  \[ LDen = \frac{LCard}{k} \]
Stratified cross-validation reduces the variance of the estimates and improves the estimation of the generalization performance of classifier algorithms.

However, how to stratify a data set in a multi-label supervised classification setting is a hard problem, since each fold should try to mimic the joint probability distribution of the whole set of class variables.
In this work we propose to solve the problem with a genetic algorithm.

Several experiments with state-of-the-art multi-label algorithms are carried out to show how our method leads to a variance reduction in the k-fold cross-validated classification performance measures, compared with other non-stratified schemes.
• The multi-label classification associates a subset of labels \( S \subseteq L \) with each instance.

• Each label can be considered a class variable with a binary sample space (the absence/presence of the label), therefore having \( |S| \) class variables.

• In order to honestly estimate a performance measure (typically the classification accuracy) of a multi-label classification algorithm we need a partition of the dataset for training and testing.
The k-fold cross-validation method allows us to estimate the measure and its variance by using the average of the corresponding k training-and-testing schemes.

A good k-fold partition of the data set must keep the statistical properties of the original data.

In particular, a stratified partition would keep the joint probability distribution (jpd) of the \(|S|\) class variables, hopefully leading to reduce the variance as compared with other non-stratified partitions.
• We first reduce the (usually high) dimension $|S|$ by selecting a subset of labels by means of the Partition Around Medoids (PAM) algorithm, the most common realisation of k-medoids clustering.

• This differs from the HOMER algorithm, which reduces the subset of labels with a hierarchically clustering and uses this smaller subset in the learning and classification stages.
• Instead, we use a small subset $S' \subseteq S$ of labels, with $|S'| = 2 \times \log N$ labels, where $N$ is the cardinal of the data set, to compute the jpd of the $|S'|$ class variables, and then we use all the labels to learn the multi-label classification model and classify new instances.

• We formulate the search for the stratified partition as an evolutionary optimization problem, solved by means of a genetic algorithm.

• Representation of data set partition ($\{1,2,3,4\}: \{1,1,2,2\}, \{1,2,1,2\},...$), Kullback-Leibler (KL) divergence based fitness (min, random, max) between the data set jpd and each fold jpd, $X$, $M$ and $S$ operators, pop.size, initialization, stop.policy
The fitness function to evaluate candidate partitions of the data set is based on the KL divergence that measures how different two distributions are.

Since a partition consists of k samples, we obtain k divergences, between both the distribution of the $|S'|$ class variables and that in the whole data set (D).
The objective function is to minimize the maximum KL divergence found in the k folds of the partition.

We also use other two fitness functions: a random partition, which is the most used procedure in machine learning, and a worst case situation, given by the maximization of the minimum KL divergence found in the k folds.

Min: \[ \min \{ \max \{ \text{KL}(D, \text{jpd}(D), \text{jpd}(F_i)) \}, i=1:K \} \]
Random: \[ \text{mean} \{ \text{KL}(D, \text{jpd}(D), \text{jpd}(F_i)) \}, \text{random } F_i, i=1:K \]
Max: \[ \max \{ \min \{ \text{KL}(D, \text{jpd}(D), \text{jpd}(F_i)) \}, i=1:K \} \]
We test the proposal over several multi-labeled data sets available in Mulan, a Java library for multi-label learning: "bibtex", "yeast", "enron", "medical", "delicious", "bookmarks", "tmc2007-500" and "genbase".

We use the recent classification algorithms "MulanMLkNN", "MulanIBLR_ML", "MulanLabelPowersetJ48", "MulanLabelPowersetBayesNet" and perform the stratified k-fold cross-validation estimation.

We evaluate the methodology against the usual simple k-fold cross-validation (K=5,10).

The experiments have been implemented in R and have been run on Magerit (CesViMa).


• BayesNet is a Weka classifier.

Methods and Experimentation

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

“MulanMLkNN”

“MulanIBLR_ML”

“MulanLabel PowersetJ48”

“MulanLabel PowersetBayesNet”
Methods and Experimentation

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

<table>
<thead>
<tr>
<th>attributes</th>
<th>name</th>
<th>domain</th>
<th>instances</th>
<th>nominal</th>
<th>numeric</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>bibtex</td>
<td>text</td>
<td>7395</td>
<td>1836</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>bookmarks</td>
<td>text</td>
<td>87856</td>
<td>2150</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>delicious</td>
<td>text</td>
<td>16105</td>
<td>500</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>enron</td>
<td>text</td>
<td>1702</td>
<td>1001</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>genbase</td>
<td>biology</td>
<td>662</td>
<td>1186</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>medical</td>
<td>text</td>
<td>978</td>
<td>1449</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>tmc2007</td>
<td>text</td>
<td>28596</td>
<td>49060</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>yeast</td>
<td>biology</td>
<td>2417</td>
<td>0</td>
<td>103</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>attributes</th>
<th>name</th>
<th>labels</th>
<th>cardinality</th>
<th>density</th>
<th>distinct</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>Bibtex</td>
<td>159</td>
<td>2.402</td>
<td>0.015</td>
<td>2856</td>
</tr>
<tr>
<td></td>
<td>bookmarks</td>
<td>208</td>
<td>2.028</td>
<td>0.010</td>
<td>18716</td>
</tr>
<tr>
<td></td>
<td>delicious</td>
<td>983</td>
<td>19.020</td>
<td>0.019</td>
<td>15806</td>
</tr>
<tr>
<td></td>
<td>enron</td>
<td>53</td>
<td>3.378</td>
<td>0.064</td>
<td>753</td>
</tr>
<tr>
<td></td>
<td>genbase</td>
<td>27</td>
<td>1.252</td>
<td>0.046</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>medical</td>
<td>45</td>
<td>1.245</td>
<td>0.028</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>tmc2007</td>
<td>22</td>
<td>2.158</td>
<td>0.098</td>
<td>1341</td>
</tr>
<tr>
<td></td>
<td>yeast</td>
<td>14</td>
<td>4.237</td>
<td>0.303</td>
<td>198</td>
</tr>
</tbody>
</table>

Mulan: A Java Library for Multi-Label Learning Datasets
import java.util.*; import java.lang.*; import java.io.*;
import weka.core.Utils;
import weka.core.Instances;
import weka.core.TechnicalInformation;
import mulan.classifier.lazy.MLkNN;
import mulan.data.MultiLabelInstances;
import mulan.evaluation.Evaluator;
import mulan.evaluation.Evaluation;
import mulan.classifier.MultiLabelOutput;
import mulan.evaluation.MultipleEvaluation;

public class MulanMLkNN {
    public static void main(String[] args) throws Exception {
        String arffTrainFilename = Utils.getOption("arffTrain", args);
        String arffTestFilename = Utils.getOption("arffTest", args);
        String szdat = Utils.getOption("szdat", args);
        String numclass = Utils.getOption("numclass", args);
        MultiLabelInstances dataTrain =
            new MultiLabelInstances( arffTrainFilename,
                new Integer( numclass).intValue());
        MultiLabelInstances dataTest =
            new MultiLabelInstances( arffTestFilename,
                new Integer( numclass).intValue());
        MLkNN learnerMLkNN = new MLkNN();
        Evaluator eval = new Evaluator( 2712);// seed
        Evaluation result;        MultipleEvaluation results;
        learnerMLkNN.build( dataTrain); //train
        result  = eval.evaluate( learnerMLkNN, dataTrain);  //test
        int INSTANCES = new Integer( szdat).intValue();
        Instances iTest = dataTest.getDataSet();
        String output_predict = new String("");
        for( int i=0; i < INSTANCES; i++) {
            //System.out.println( i + "   " + iTest.get( i));
            String op = new String(
                learnerMLkNN.makePrediction( iTest.get( i)).toString());
            int i0 = op.indexOf( "]");
            int i1 = op.indexOf( "]");
            op = " " + op.substring( i0+1, i1);//\[]
            output_predict = output_predict + op + "\n";
        }
    }
}
• We have performed seven repetitions of every scenario and have summarized the results.

• We have run a high number of experiments by varying:

  7 repetitions x
  8 data sets x
  2 k-fold schemes x
  3 fitness functions x
  4 PAM-cluster configurations x
  4 classification models.
• We also estimate six multi-label performance measures categorized according to the required type of output from a multi-label model:

  Hamming Loss (Bipartition),
  Subset Accuracy (Bipartition),
  Coverage (Ranking),
  Ranking Loss (Ranking),
  Mean Average Precision (Probabilities) and
  Micro-Averaged AUC (Probabilities).
## Results

### Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

<table>
<thead>
<tr>
<th>Output</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;reply&quot;</td>
<td>&quot;K&quot;</td>
<td>&quot;fitness&quot;</td>
<td>&quot;dataset&quot;</td>
<td>&quot;model&quot;</td>
<td>&quot;kld&quot;</td>
<td>&quot;Mean&quot;</td>
<td>&quot;Vari&quot;</td>
<td>&quot;am&quot;</td>
<td></td>
</tr>
<tr>
<td>&quot;1&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;genbase&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.039830678</td>
<td>0</td>
<td>0</td>
<td>&quot;hl&quot;</td>
</tr>
<tr>
<td>&quot;13&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;genbase&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.035472034</td>
<td>0.037500000</td>
<td>0.003645833</td>
<td>&quot;hl&quot;</td>
</tr>
<tr>
<td>&quot;229&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;medical&quot;</td>
<td>&quot;MulanIBLR_ML&quot;</td>
<td>0.046185681</td>
<td>0.917857142</td>
<td>0.000795699</td>
<td>&quot;ga&quot;</td>
</tr>
<tr>
<td>&quot;148&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;medical&quot;</td>
<td>&quot;MulanIBLR_ML&quot;</td>
<td>0.084760586</td>
<td>0.920535714</td>
<td>0.000283003</td>
<td>&quot;ga&quot;</td>
</tr>
<tr>
<td>&quot;8324&quot;</td>
<td>7</td>
<td>5</td>
<td>&quot;ran&quot;</td>
<td>&quot;yeast&quot;</td>
<td>&quot;MulanLabelPowersetJ48&quot;</td>
<td>0.557441975</td>
<td>0.555555555</td>
<td>0.228395061</td>
<td>&quot;rl&quot;</td>
</tr>
<tr>
<td>&quot;790&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;max&quot;</td>
<td>&quot;yeast&quot;</td>
<td>&quot;MulanLabelPowersetJ48&quot;</td>
<td>0.670284611</td>
<td>0.133333333</td>
<td>0.047222222</td>
<td>&quot;rl&quot;</td>
</tr>
<tr>
<td>&quot;4140&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;enron&quot;</td>
<td>&quot;MulanLabelPowersetBayesNet&quot;</td>
<td>0.435892104</td>
<td>1.096000000</td>
<td>2.767715555</td>
<td>&quot;rc&quot;</td>
</tr>
<tr>
<td>&quot;1631&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;enron&quot;</td>
<td>&quot;MulanLabelPowersetBayesNet&quot;</td>
<td>0.395055705</td>
<td>1.512500000</td>
<td>3.876562500</td>
<td>&quot;rc&quot;</td>
</tr>
<tr>
<td>&quot;1152&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;bibtex&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.048509519</td>
<td>0.975000000</td>
<td>0.003125000</td>
<td>&quot;map&quot;</td>
</tr>
<tr>
<td>&quot;1338&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;bibtex&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.053564006</td>
<td>1</td>
<td>0</td>
<td>&quot;map&quot;</td>
</tr>
<tr>
<td>&quot;2167&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;delicious&quot;</td>
<td>&quot;MulanIBLR_ML&quot;</td>
<td>0.578744569</td>
<td>0.746153846</td>
<td>0.036501479</td>
<td>&quot;mab&quot;</td>
</tr>
<tr>
<td>&quot;1445&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;delicious&quot;</td>
<td>&quot;MulanIBLR_ML&quot;</td>
<td>0.627057111</td>
<td>0.500000000</td>
<td>0</td>
<td>&quot;mab&quot;</td>
</tr>
<tr>
<td>&quot;1168&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;bookmarks&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.031957106</td>
<td>0</td>
<td>0</td>
<td>&quot;hl&quot;</td>
</tr>
<tr>
<td>&quot;1346&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;bookmarks&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.033203022</td>
<td>0.025000000</td>
<td>0.003125000</td>
<td>&quot;hl&quot;</td>
</tr>
<tr>
<td>&quot;1182&quot;</td>
<td>1</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;tmc2007-500&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.298353251</td>
<td>0.872916666</td>
<td>0.000347222</td>
<td>&quot;ga&quot;</td>
</tr>
<tr>
<td>&quot;1353&quot;</td>
<td>2</td>
<td>5</td>
<td>&quot;min&quot;</td>
<td>&quot;tmc2007-500&quot;</td>
<td>&quot;MulanMLkNN&quot;</td>
<td>0.374339402</td>
<td>0.762053571</td>
<td>0.000284664</td>
<td>&quot;ga&quot;</td>
</tr>
</tbody>
</table>
Results

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula =
  trf(DF.hl[DF.hl$model == MulanMODEL[iM], ]$Vari) ~
  DF.hl[DF.hl$model == MulanMODEL[iM], ]$fitness)

$`DF.hl[DF.hl$model == MulanMODEL[iM], ]$fitness`

MulanMLkNN
<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>min-max</td>
<td>-0.0150344</td>
<td>-0.0215530</td>
<td>-0.0085157</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0105317</td>
<td>-0.0170504</td>
<td>-0.0040130</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.0045026</td>
<td>-0.0020160</td>
<td>0.0110213</td>
</tr>
</tbody>
</table>

MulanIBLR_ML
<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>min-max</td>
<td>-0.0127776</td>
<td>-0.0203942</td>
<td>-0.00516107</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0073495</td>
<td>-0.0149661</td>
<td>0.00026702</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.0054280</td>
<td>-0.0021884</td>
<td>0.01304468</td>
</tr>
</tbody>
</table>

MulanLabelPowersetJ48
<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>min-max</td>
<td>-0.0146596</td>
<td>-0.0221309</td>
<td>-0.0071882</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0108203</td>
<td>-0.0182917</td>
<td>-0.0033490</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.0038392</td>
<td>-0.0036320</td>
<td>0.0113105</td>
</tr>
</tbody>
</table>

MulanLabelPowersetBayesNet
<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>min-max</td>
<td>-0.0144732</td>
<td>-0.0209432</td>
<td>-0.0080033</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0104391</td>
<td>-0.0169091</td>
<td>-0.0039691</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.0040341</td>
<td>-0.0024358</td>
<td>0.0105041</td>
</tr>
</tbody>
</table>
### Results

**Tukey multiple comparisons of means**

95% family-wise confidence level

Fit: \texttt{aov(formula =}
\texttt{trf(DF.rl[DF.rl$model == MulanMODEL[iM],]$Vari) ~}
\texttt{DF.rl[DF.rl$model == MulanMODEL[iM],]$fitness)}}

\texttt{`DF.rl[DF.rl$model == MulanMODEL[iM],]$fitness`}

<table>
<thead>
<tr>
<th>Model</th>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>MulanMLkNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min-max</td>
<td>-0.00261802</td>
<td>-0.015286</td>
<td>0.010050</td>
<td>0.85900</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.00239891</td>
<td>-0.015067</td>
<td>0.010269</td>
<td>0.88003</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.00021910</td>
<td>-0.012449</td>
<td>0.012887</td>
<td>0.99892</td>
</tr>
<tr>
<td>MulanIBLR_ML</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min-max</td>
<td>-0.0063262</td>
<td>-0.0187302</td>
<td>0.0060777</td>
<td>0.41240</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0022075</td>
<td>-0.0146115</td>
<td>0.0101965</td>
<td>0.89317</td>
</tr>
<tr>
<td>Ran-min</td>
<td>0.0041187</td>
<td>-0.0082852</td>
<td>0.0165227</td>
<td>0.67917</td>
</tr>
<tr>
<td>MulanLabelPowersetJ48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min-max</td>
<td>-0.0020904</td>
<td>-0.017882</td>
<td>0.013701</td>
<td>0.93923</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0037771</td>
<td>-0.019568</td>
<td>0.012014</td>
<td>0.81631</td>
</tr>
<tr>
<td>Ran-min</td>
<td>-0.0016867</td>
<td>-0.017478</td>
<td>0.014105</td>
<td>0.95995</td>
</tr>
<tr>
<td>MulanLabelPowersetBayesNet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min-max</td>
<td>0.0026603</td>
<td>-0.0075241</td>
<td>0.0128449</td>
<td>0.78543</td>
</tr>
<tr>
<td>ran-max</td>
<td>-0.0013027</td>
<td>-0.0114873</td>
<td>0.0088817</td>
<td>0.94312</td>
</tr>
<tr>
<td>Ran-min</td>
<td>-0.0039631</td>
<td>-0.0141477</td>
<td>0.0062213</td>
<td>0.59044</td>
</tr>
</tbody>
</table>

![Mean of ranking loss](image1)
![Variance of ranking loss](image2)
## Results

### Tukey multiple comparisons of means

95% family-wise confidence level

```r
Fit: aov(formula =
  trf(DF.map[, model == MulanMODEL[iM], ]$Vari) ~
  DF.map[, model == MulanMODEL[iM], ]$fitness)

$`DF.map[, model == MulanMODEL[iM], ]$fitness

MulanMLkNN
  diff  lwr  upr  p adj
  min-max -0.0067258 -0.0113132 -0.00213852 0.00405
  ran-max -0.0053966 -0.0099840 -0.00080932 0.01989
  ran-min  0.0013291 -0.0032581  0.00591655 0.74364

MulanIBLR_ML
  diff  lwr  upr  p adj
  min-max -0.0076973 -0.0130452 -0.00234953 0.00470
  ran-max -0.0047286 -0.0100764  0.00061921 0.08845
  Ran-min  0.0029687 -0.0023790  0.00831660 0.35355

MulanLabelPowersetJ48
  diff  lwr  upr  p adj
  min-max -0.0076973 -0.0130452 -0.00234953 0.00470
  ran-max -0.0047286 -0.0100764  0.00061921 0.08845
  Ran-min  0.0029687 -0.0023790  0.00831660 0.35355

MulanLabelPowersetBayesNet
  diff  lwr  upr  p adj
  min-max -0.00600728 -0.0097236 -0.0022909 0.00174
  ran-max -0.00535589 -0.0090722 -0.0016395 0.00465
  ran-min  0.00065138 -0.0030649  0.0043677 0.89618
```
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Mean of ranking coverage

Variance of ranking coverage

Mean of global accuracy

Variance of global accuracy
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Results

Mean of ranking coverage

Variance of ranking coverage

Mean of global accuracy

Variance of global accuracy

Mean

Variance

Min  Ran  Max

bookmarks
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Mean of hamming loss

Variance of hamming loss

Mean of ranking coverage

Variance of ranking coverage

delicious

delicious

delicious

delicious
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Mean of ranking coverage

Variance of ranking coverage

Mean of global accuracy

Variance of global accuracy
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Kullback-Leibler divergence based fitness (min, random, max)

Kullback-Leibler divergence based fitness (min, random, max)
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Kullback-Leibler divergence based fitness (min, random, max)

Kullback-Leibler divergence based fitness (min, random, max)
Results

Stratified Cross-Validation in Multi-Label Classification Using Genetic Algorithms

Kullback-Leibler divergence based fitness (min, random, max)

Kullback-Leibler divergence based fitness (min, random, max)
Conclusions

• The genetic algorithm based on a smaller subset of labels allows us to compute the jpd of the labels in high multidimensional data sets.

• It also allows to find a partition that behaves well in the whole data set, with all the labels.

• The resulting stratified partition proposed here to be used in the k-fold cross-validation estimation reduces the variance of the performance measures.
References

[Ph. D Qualifying Review Paper]


CesViMa, Magerit, http://www.cesvima.upm.es/
Index

• Introduction
  • Multilabel Classification
  • Cross-Validation and Stratified Cross-Validation
• Methods and Experimentation
  • Genetic Algorithms
  • Mulan, Weka, Data Sets
• Results
• Conclusion
• Future Lines
• References