# Label Clusters Chains Multi-Label Classifier

## Intro

Multi-label classification (MLC) is a type of supervised machine learning that allows instances to be categorized into two or more labels simultaneously.

- Local approach -> divides into subproblems. Might disregard correlation between labels.
- Global approach -> develops or mods algorithms to deal with the original problem. Migh miss local correalation information.

Taking correlation into account allows the model to exploit dependencies between labels. For example, Jaccar index similarity quantifies the degree of co-occurrence between labels pairs.

Classifier Chains CC aims to capture label correlation by breaking the og problem into multiple binary problems. The idea is to train a separate classifier for eahc label and then use the predictions generated as new features in the subsequent classifiers. Classifiers al linked in chain. Order matters (error propagation)

Ensemble of Classifiers Chains ECC extends CC to use multiple chains. Each chain is trained differently, and combination of different chains results make the prediction.

We hypothesize that clustering the multi-label space imporves performance by breaking the long chain of classifiers seen in ECC, which benefits datasets with high dimensionality.

## THE PROPOSAL LCC-MLC

Label Clusters Chains Multi-Label Classifier.

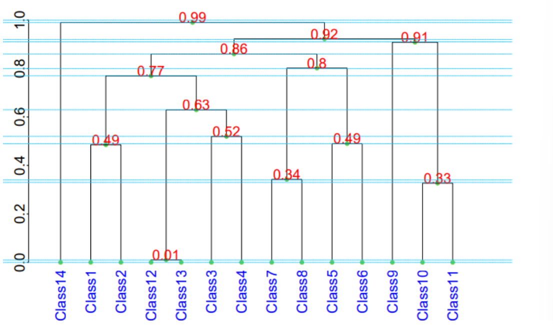
Disjoint correlated label clusters refer to clusters of correlated labels that do not overlap or share comon elementes. For example: C1 = {L1, L4, L6}, C2 = {L2, L5}, and C3 = {L3}. The key feature of LCC-ML is its chaining strategy to build chains of disjoint correlated label clusters.

**Step1**

Create a model for label correlation bassed on a similarity measure (Jaccard) aiming to cluster correlated labels into disjoint clusters.

**Step2**

We obtain the similarity matrix M = LxL from Jaccard and transform it into a dissimilarity matrix and pass it as input to tha Agglomerative Hierarchical Clustering Algorithm AHCA to organize the labels into nested clusters. It generates a dendrogram that represents the hierarchy of clusters.




Cut the dendrogram to obtain as many partitions as labels

**Step3**

To determine the best partition they use the slhoutte coefficient, a method to measure clustering quality. The quality is based on thhe proximity measure between the lables within the cluster and the distance between the these labels and their nearest cluster.

**Training**

Once the labels are grouped into disjoint clusters, the training process proceeds as follows:

**a) Sequence of Clusters**

- The clusters are ordered based on the dendrogram, following the sequence in which they were merged during the hierarchical clustering process. This order ensures that the correlations between labels within the clusters are respected.

**b) Using Labels as Features**

- **Initial Cluster**: For the first cluster, the model uses only the original features of the dataset.
- **Subsequent Clusters**: For the following clusters, the predicted labels from the previous clusters are added as new features.

**c) Individual Models per Cluster**

- For each cluster, a random forest classifier is trained using:
  - The original features of the dataset.
  - The predicted labels from previous clusters (for clusters other than the first).

**d) Training Output**

- Each cluster produces a set of predicted labels.
- These predicted labels are combined at the end to generate the final multi-label classification.

Imagine you have three clusters: **Cluster 1**, **Cluster 2**, and **Cluster 3**. The process would be as follows:

1. Train the model for **Cluster 1** using only the original features.

2. Use the predictions from **Cluster 1** as additional features to train the model for **Cluster 2**.
3. Use the combined predictions from **Cluster 1** and **Cluster 2** as additional features to train the model for **Cluster 3**.
4. Combine the final predictions from all three clusters to produce the complete multi-label classification.

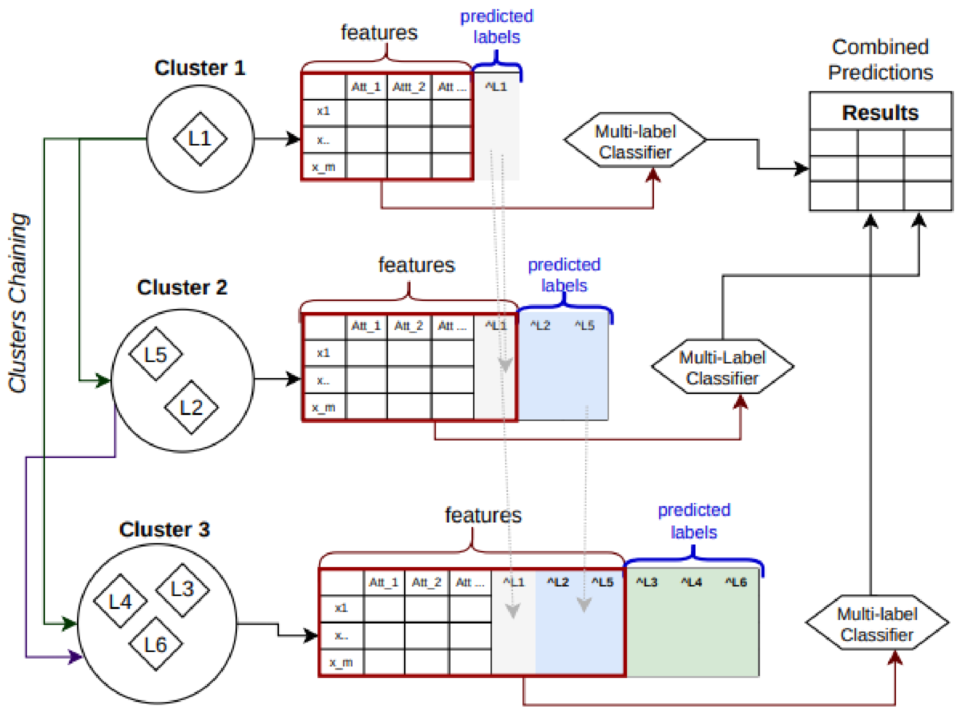



**Fig. 6**: Illustration of the LCC-ML predicting phase.