A Consensus-Based Active Learning Strategy for Multi-Label Classification

HELLO!

I am Aditya Kumar Akash

I am here to present BTP Phase 2 work.

You can find me at
adityakumarakash@gmail.com

3

"If a picture is worth a thousand words, a video says it all."

4

Contents

5

ABSTRACT

- Generating quality labeled data, while essential for automatic classification, is a time consuming task. Further accentuated when the task involves multiple labels and multiple labelers both human and machine
- Minimize human labeling effort without sacrificing the accuracy of individual labelers or inter-labeler consensus
- A novel formulation that aims to collectively optimize the cost of labeling, labeler reliability and inter-labeler consensus
- Our sampling strategy queries for instance-label pairs that maximize the expected consensus in successive labeling iterations.

6

Related work

Multi-label active learning has received a significant attention in recent and has successfully been applied to domains such as image annotation and text categorization

- These active Learning methods query for most uncertain instance, label pair from the most reliable annotato

In this work, we build upon active learning with consensus maximization.

Another major focus of our work is to show that improving the consensus amongst the users also leads to model prediction getting more aligned towards the ground truth.

7

Consensus Learning

Combining model outputs

8

Meta Learning

"We do not learn from experience...

we learn from reflecting on experience"

- John Dewey

9

Consensus Based learners

Consensus based learners use predictions from raw data, and not the raw data itself. This makes them most suitable for our use of multilabel classification.

We use "Multilabel consensus maximization for ranking" - MLCM-r from [1].

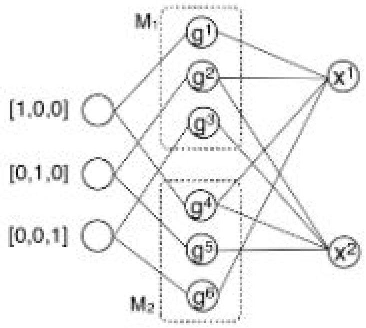
Involves bipartite graph with nodes as -

- *Instance nodes, videos are instances*
- *Group nodes* for each label from each classifiers, *tags are the labels*
- *Connection matrix*, representing *prediction of a label for an instance by a classifier*

10

MLCM-r

Example graph for 2 instances, 3 labels, 2 classifiers



Following probability distributions with each node -

- *u_i with instance nodes* - u_{il} : probability of relevance of lth label for ith instance
- *q_j with group nodes* - q_{jl} : probability of seeing label l when in node g_j, relates to how closely the labels are in opinion of the classifier to which g_j belong

11

MLCM-r optimization problem

Consensus maximized by solution of following -

$$\min_{U,Q} \sum_{i=1}^{n} \sum_{j=1}^{v} a_{ij} \|\boldsymbol{u}_i - \boldsymbol{q}_j\|^2 + \alpha \sum_{j=1}^{v} \|\boldsymbol{q}_j - \boldsymbol{b}_j\|^2$$

$$u_{ik} \geq 0, \sum_{k=1}^{l} u_{ik} = 1, i = 1, \ldots, n$$

$$q_{jk} \geq 0, \sum_{k=1}^{l} q_{jk} = 1, j = 1, \ldots, v$$

1. First term means two connected nodes should have similar probability distribution
2. Second term ensures groups do not deviate much from its initial distribution

12

Solution to MLCM-r

The solution is obtained by block coordinate descent-

$$q_j^t = \frac{\sum_{i=1}^{n} a_{ij} \boldsymbol{u}_i^{t-1} + \alpha \boldsymbol{b}_j}{\sum_{i=1}^{n} a_{ij} + \alpha}$$

$$\boldsymbol{q}_j^t = \frac{\sum_{j=1}^{v} a_{ij} \boldsymbol{q}_j^t}{\sum_{j=1}^{v} a_{ij}}$$

Solution gives u_i which is used to infer the most likely label.

The main of advantage of MLCM is that it handles the case of label-label correlations present in the data. We use MLCM-r as the consensus model in our system.

13

A Measure of consensus

The consensus of a model with other models is defined using Cohen Kappa, which captures inter-rater agreement

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

The consensus of an model is defined as the Kappa value between prediction for the model with the prediction of the MLCM-r model. The overall consensus of the system is the sum of consensus for all the models in the system. The consensus for an instance is taken as mean over Kappa between prediction of MLCM-r and each model

14

Labeler reliability

The reliability of a labeler is estimated for each label.

The reliability is updated in following manner

$$r_i^j \leftarrow r_i^j + \gamma(\kappa^j - r_i^j) \quad (7)$$

where κ^j is Kappa calculated w.r.t j^{th} label, and $\gamma < 1$ is a constant.

Accordingly, we update the MLCM-r model. For each group node, instead of α_i^j, we use $\alpha_i^j \times r_i^j$. This accounts for the user reliability in our model.

15

ACTIVE LEARNING

Learn by querying

16

Active learning by uncertainty sampling

- The underlying assumption for this sampling is that for instance in while the consensus is low indicates that there is confusion amongst the users

- For confused (uncertain) label selection is done by binarization of probability distribution. The probabilities which are near the threshold are the confusing ones.

- The confused instance, label pair is queried to the user with highest reliability on the given label

17

Active learning by Influence based sampling

- In this we sample an instance,label and user set, such that when user is queried about the selected instance, label pair the expected increase in consensus of the system is maximum.

$$\arg\max_{u,x,l} \sum_{u} \kappa_u$$

18

Active learning by Influence based sampling Experiment

For the active learning setting, we perform following experiment

- Find out instances which have consensus values in a given range, which is selected such that the consensus is neither very high nor low. For our case we choose [0.85, 0.90].
- Find out the models which have lowest consensus on the selected instance.
- The models selected add the instances into their training set. The consensus model output is used as ground truth.
- Consensus model is re-run to calculate the predictions.

19

Results

Table 3: MLCM vs MLCM with r_i^j

Model	Avg Raking loss
MLCM	0.01722
MLCM with r_i^j	0.01667

Table 4: User removal effect

Dataset	Avg Change in Consensus	Avg Change in fMeasure
medical	0.0053068	0.0006081

Table 5: Consensus change by using Active learning

Dataset	Avg Consensus change
Medical	2.22
enron	1.57

Conclusions -

We consider the problem of multi-label classification using active learning. Rather than just sampling the data based on uncertainty we try to sample the instance label pair to maximize the consensus. We provide several evidence to show that the improving the consensus moves the data towards the ground truth. We also incorporate the user-reliability into the system and show the improvement to the original framework proposed. We used this reliability to remove user with least consensus, and show that this led to improved consensus and increased F1 measure.

THANKS!

Any questions?

You can find me at

- adityakumarakash@gmail.com