Consensus Learning Model

Aditya Kumar Akash

IIT Bombay

adityakumarakash@gmail.com

October 13, 2015

Overview

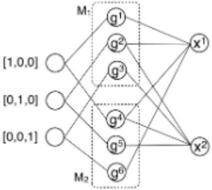
1. Problem
   - Problem as graph
2. Solution
   - Solution of optimization problem

Multilabel classification

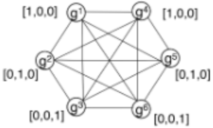
Data form - (x, z), x is the feature vector of an instance
Let L be set of all l possible labels,
z is a vector of length l, $z_l \in \{0, 1\}$, denoting relevance of label to the instance

Model

A bipartite graph modelling relations between instances and predictions is used for prediction combination.



(a) MLCM-r



(b) Graph of group nodes encoding label relationships

Graph has group nodes g and instance nodes x. Dashed rectangle surround group nodes from one classifier.

Bipartite Graph

In the graph -

- Group Node represents a label depicted by connection of group node to left most nodes
- Instance nodes connected to more than one group node thus representing multilabel prediction
- Relation between labels given by figure b

Formulation

Notations -

- Each node associated with probability distribution over l labels
- $u_{il} \rightarrow$ prob of i-th instance taking label class 1 on label l, thus prob of relevance
- $q_{jl} \rightarrow$ prob of seeing l-th label given the j-th label
- $v = m * l \rightarrow$ Number of group nodes = classifiers * labels

TABLE II: Notations for MLCM-r

Symbol	Meaning
A	$a_{i,j}$ is the prediction of label (j mod l) on $\mathbf{x}_i$ by the $\lfloor j/l \rfloor$
B	Label node class distribution
U	$u_{i\ell}$ is the probability that label ℓ is relevant to $\mathbf{x}_i$
Q	$q_{j\ell}$ is the probability of seeing label ℓ given label j
I_k	k dimensional identity matrix

Formulation

Optimization problem formulated $\rightarrow$

$$\min_{U,Q} \sum_{i=1}^{n} \sum_{j=1}^{v} a_{ij} \|u_i - q_j\|^2 + \alpha \sum_{j=1}^{v} \|q_j - b_j\|^2$$
$$u_{il} \geq 0, \sum_l u_{il} = 1, i = 1, ..., n$$
$$q_{jl} \geq 0, \sum_l q_{jl} = 1, j = 1, ..., v$$

- First term ensures that if x_i is assigned g_j then their probability distribution must be close
- Second term puts constraints that a group g_j 's distribution is not much deviating from the initial prediction

Algorithm

Input: group-object affinity matrix A, initial labeling matrix Y; parameters α and ϵ;
Output: consensus matrix U;
Algorithm:

Initialize U^0,U^1 randomly
$t \leftarrow 1$
while $||U^t - U^{t-1}|| > \epsilon$ **do**
  $Q^t = (D_v + \alpha K_v)^{-1}(A^T U^{t-1} + \alpha K_v Y)$
  $U^t = D_n^{-1} A Q^t$
return U^t

$$\vec{q}_{j\cdot}^{\,(t)} = \frac{\sum_{i=1}^{n} a_{ij}\vec{u}_{i\cdot}^{\,(t-1)} + \alpha k_j \vec{y}_{j\cdot}}{\sum_{i=1}^{n} a_{ij} + \alpha k_j} \qquad \vec{u}_{i\cdot}^{\,(t)} = \frac{\sum_{j=1}^{v} a_{ij}\vec{q}_{j\cdot}^{\,(t)}}{\sum_{j=1}^{v} a_{ij}}$$

Here, $D_v = diag(\sum_{i=1}^{n} a_{ij})_{v \times v}$, $D_n diag(\sum_{j=1}^{v} a_{ij})_{n \times n}$,
$K_v = diag(\sum_{z=1}^{l} y_{jz})_{v \times v}$

Closed Solution

$$Q^* = (I_v - D_\lambda D_v^{-1} A' D_n^{-1} A)^{-1} D_{1-\lambda} B \qquad (3)$$

where $D_v = \text{diag}(\mathbb{1}'A)$, $D_n = \text{diag}(\mathbb{1}'A')$, $\mathbb{1}$ is a column vector with all entries being 1. $D_\lambda = D_v(\alpha I + D_v)^{-1}$ and $D_{1-\lambda} = \alpha(\alpha I + D_v)^{-1}$. After Q^* is obtained, U^* is obtained using

$$U^* = D_n^{-1} A Q^* \qquad (4)$$

where the $v \times v$ diagonal matrices D_λ and $D_{1-\lambda}$ have (j, j) entries as $\frac{w_j}{w_j + \alpha k_j}$ and $\frac{\alpha k_j}{w_j + \alpha k_j}$.

Current Work

- Working on update equation when new edges or instances are added
- Model hierarchical tree structure on labels
- Implementing the above algorithm to find out the effectiveness

References

Sihong Xie, Xiangnan Kong, Jing Gao, Wei Fan, Philip S. Yu
Multilabel Consensus Classification

Jing Gaoy, Feng Liangy, Wei Fanz, Yizhou Suny, Jiawei Han
Graph-based Consensus Maximization among Multiple Supervised and Unsupervised Models

The End