OnlineCM: Real-time Consensus Classification with Missing Values

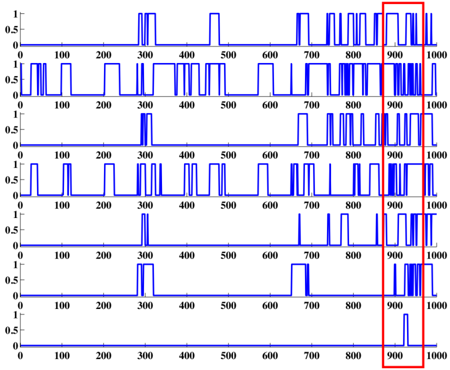
Bowen Dong[1], Sihong Xie[1], Jing Gao[2], Wei Fan[3], Philip S. Yu[1]

April 30, 2015

[1]Deparment of Computer Science, University of Illinois at Chicago, IL, USA
[2]Department of Computer Science, University at Buffalo, NY, USA
[3]Baidu Big Data Lab, Sunnyvale, CA, USA

Motivations

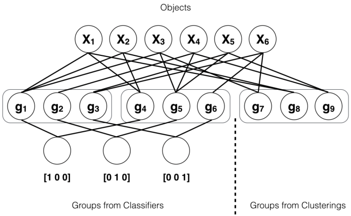
Combining predictions from multiple sources or models has been shown to be a useful technique in data mining. (e.g. anomaly detection)



Challenges

1. The **high velocity** and **high volume** of the data render existing batch mode prediction aggregation algorithms infeasible.
2. Due to the **heterogeneity**, predictions from multiple models might not be perfectly synchronized, leading to abundant **missing values** in the prediction stream.

Algorithms



Symbol	Definition
$A_{n\times v} = [a_{ij}]$	indicator of object i in group j
$U_{n\times c} = [u_{ij}]$	probability of object i w.r.t. class j
$Q_{v\times c} = [q_{ij}]$	probability of group i w.r.t. class j
$Y_{v\times c} = [y_{ij}]$	indicator of object i predicted as class j

$$\min_{Q,U} \sum_{i=1}^{n}\sum_{j=1}^{v} a_{ij}\|u_{i\cdot} - q_{j\cdot}\|^2 + \alpha\sum_{j=1}^{v} k_j\|q_{j\cdot} - y_{j\cdot}\|^2$$

$$\text{s.t.} \quad u_{i\cdot} \geq 0, \quad |u_{i\cdot}| = 1, \quad i = 1, \dots, n$$

$$q_{j\cdot} \geq 0, \quad |q_{j\cdot}| = 1, \quad j = 1, \dots, v$$

$$(4.1) \qquad Q^{(t)} = (D_v + \alpha K_v)^{-1}(A'U^{(t-1)} + \alpha K_v Y)$$

$$(4.2) \qquad U^{(t)} = D_n^{-1}AQ^{(t)}$$

$$(4.3) \quad Q^{(t)} = (D_v + \alpha K_v)^{-1}(A'D_n^{-1}AQ^{(t-1)} + \alpha K_v Y)$$

As $t \to \infty$, $Q^{(t)}$ converges to

$$(4.4) \qquad Q^* = (I - D_\lambda S)^{-1}D_{1-\lambda}Y$$

$$(4.5) \qquad U = D_n^{-1}AQ^* = \frac{1}{m}AQ^*$$

$$(4.6) \qquad u_{n+1,\cdot} = \frac{1}{m}\sum_{j=1}^{v} a_{n+1,j}Q^*_{j\cdot}$$

- D_v: D_v is diagonal with the j-th diagonal elements being $\sum_{i=1}^{n} a_{ij}$, as we add one more row to A, D_v is updated by $\sum_{i=1}^{n} a_{ij} + a_{n+1,j}$.
- D_λ: $D_\lambda = (D_v + \alpha K_v)^{-1}D_v$, therefore, once we updated D_v, D_λ can be easily derived from D_v.
- $D_{1-\lambda}$: $D_{1-\lambda} = (D_v + \alpha K_v)^{-1}(\alpha K_v)$, which is similar to D_λ.
- $S = D_v^{-1}A'D_n^{-1}A$: since we know how to update D_v, we just need to update $A'D_n^{-1}A$. The new ij-th entry of $A'D_n^{-1}A$ is given by $\frac{1}{m}\sum_{k=1}^{n+1} a_{ik}a_{kj} = \frac{1}{m}(\sum_{k=1}^{n} a_{ik}a_{kj} + a_{i,n+1}a_{n+1,j})$.

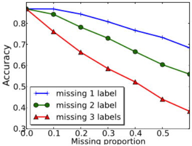
Experiments

Table 2: Data Sets Description

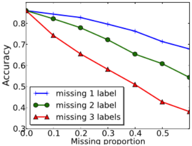
Data Set	# of training	# test	# classes	# features
rcv1	15564	518571	53	47236
mnist	60000	10000	10	778
SensIT	25010	1000000	10	10
covtype	20000	561012	7	54

Table 3: Classification Accuracy

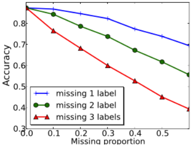
Methods	rcv1	mnist	SensIT	covtype
BGCM	0.8954	0.9187	0.8843	0.8901
OnlineCM	**0.8772**	**0.8822**	**0.8692**	**0.8758**
IncrementalCM	0.8240	0.8415	0.8375	0.8176



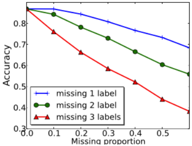
(a) rcv1



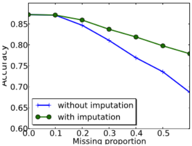
(b) ensIT



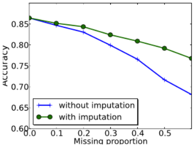
(c) mnist



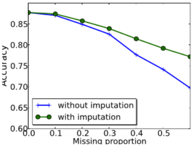
(d) covtype



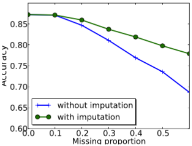
(a) rcv1



(b) ensIT



(c) mnist



(d) covtype