桂林電子科技大學
GUILIN UNIVERSITY OF ELECTRONIC TECHNOLOGY

Drift Detection for Multi-label Data Streams Based on Label Grouping and Entropy

Zhong-wei Shi [1], **Yi-min Wen** [1], **Chao Feng** [1], **Hai Zhao** [2]

[1] Guilin University of Electronic Technology
[2] Shanghai Jiao Tong University
China

ICIaNov @ ICDM 2014, Shenzhen, China, December 14, 2014

Outline

- Introduction
- The Drift Detection Method
- Experiments and Results
- Conclusion

Introduction

In many emerging applications, each sample may be associated with more than one label and the correlation between class labels may change over time.

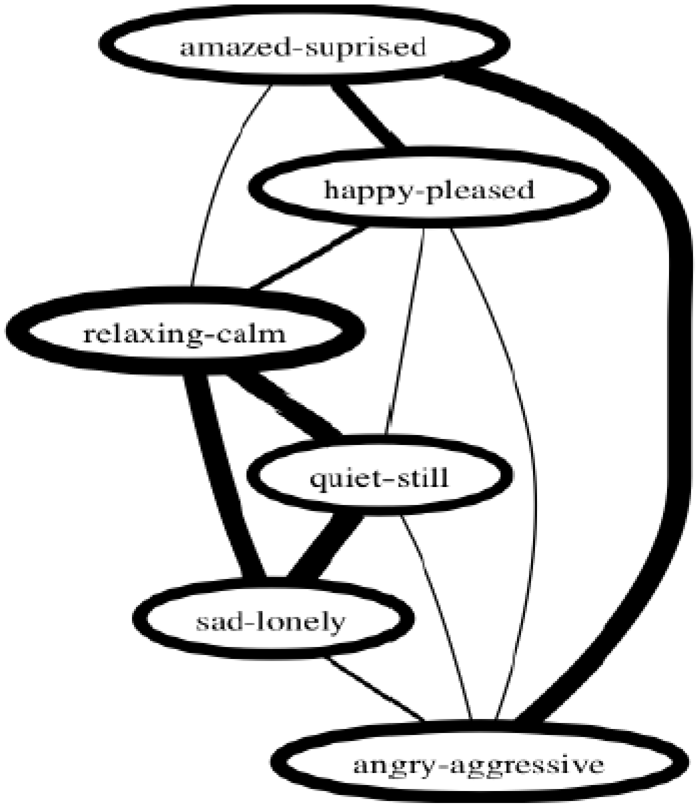
Student : IT , Literature

graduating

Worker : IT , Financial Management and Company

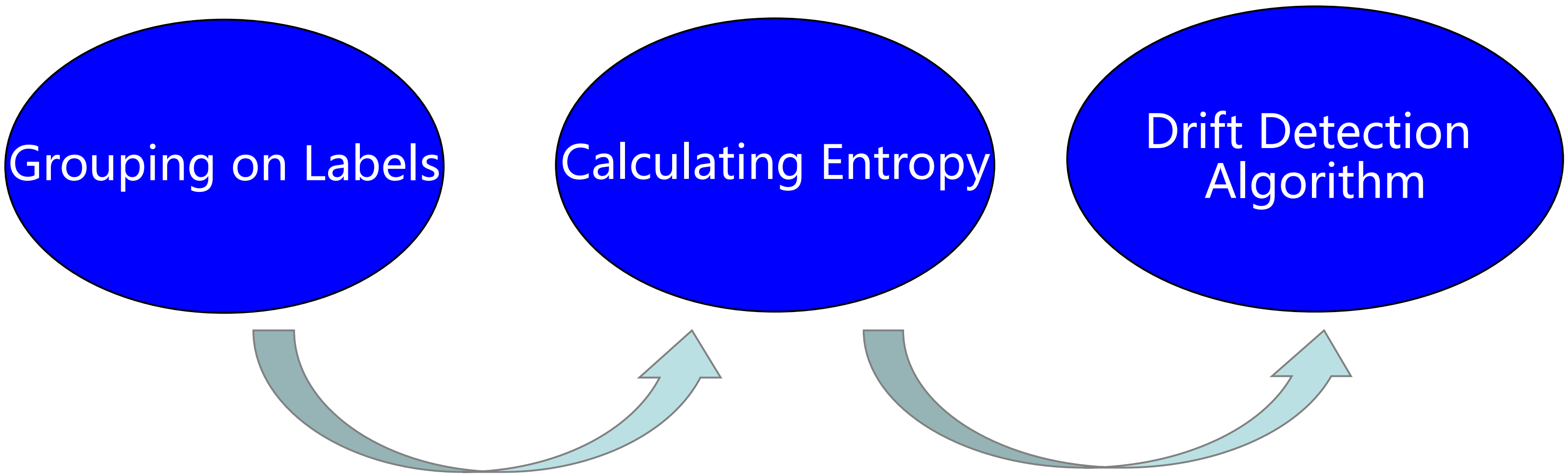
- dependencies between labels

- the correlation relationship between labels and the distribution between features and multiple labels



co-occurrences of music labeled with emotions

The Drift Detection Method



A. Grouping on Labels

a). Algorithm 1: GL1()

It employs the Apriori algorithm to mine the frequent label sets and the basic motivation is the existence of co-occurrence between labels.

	amazed	happy	relaxing	quiet	*sad*	*angry*
x_1 →	0	1	1	0	0	0
x_2 →	0	0	1	1	1	0
x_3 →	0	1	1	0	0	0
x_4 →	1	0	0	0	0	1
x_5 →	0	0	1	1	1	0
x_6 →	1	0	0	0	1	1

CO(amazed, angry) = 2
CO(happy, relaxing) = 2
CO(relaxing, quiet, sad) = 2

Call Apriori Algorithm
Minimum Support = 2

	amazed	happy	relaxing	quiet	*sad*	*angry*
x_1 →	0	1	1	0	0	0
x_2 →	0	0	1	1	1	0
x_3 →	0	1	1	0	0	0
x_4 →	1	0	0	0	0	1
x_5 →	0	0	1	1	1	0
x_6 →	1	0	0	0	1	1

THREE GROUPS:
(amazed, angry),
(happy, relaxing) and
(relaxing, quiet, sad)

	(amazed, angry)	(happy, relaxing)	(relaxing, quiet, sad)
x_1 →	0	1	1
x_2 →	0	1	1
x_3 →	0	1	1
x_4 →	1	0	0
x_5 →	0	1	1
x_6 →	1	0	1

b). Algorithm 2: GL2()

It mines the dependencies between labels by the clustering method and it can be instantiated with the k-means and EM algorithm.

Motivation: Clustering can place the similar and interdependent objects together and dissimilar and independent apart.

Algorithm 2 : GL2( )

Input : $\{(x_1, Y_1), \ldots, (x_i, Y_i), \ldots, (x_N, Y_N)\}$; $L = \{l_1, \ldots, l_m\}$; K

Output : $\bar{L} = \{\bar{l}_1, \ldots, \bar{l}_n\}$; $x_i (1 \le i \le N)$ with new labels

```
From i To m
        generate a new sample x̃_i = (y_1^i, ..., y_N^i)
generate the new label data LD = (x̃_1, ..., x̃_m)
X̄ ← call_EM(LD, K);
For Each  X̄_i ∈ X̄
        IF  x̃_j ∈ X̄_i
                Then  add l_j to l̄_i
End
Get new L̄ = {l̄_1, ..., l̄_n}
For Each x_i (1 ≤ i ≤ N)
        For Each  l̄_i ∈ L̄  (1 ≤ i ≤ n)
                IF there is  y_i^j = 1 and l_j ∈ l̄_i
                        Then  annotate x_i with new label l̄_i
End
Return  L̄
Return  each x_i with new labels l̄_i;
```

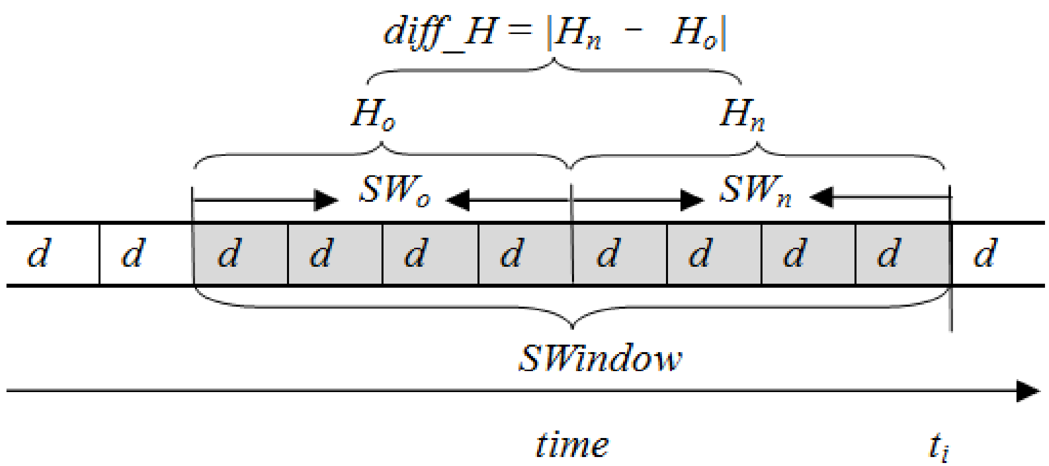
B. Calculating Entropy

$$H_i = \frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K}\sum_{b=1}^{B} H_{iskb}$$

$$H_{iskb} = -\left[P_{iskb}\log_2(P_{iskb}) + (1-P_{iskb})\log_2(1-P_{iskb})\right]$$

P_{iskb} represents the probability of a sample belongs to the k^{th} label subset, with feature domain s in b at time t_i.

C. Drift Detection Algorithm



we chose two sliding windows, respectively representing the older and the most recent sample.

Experiments and Results

- A. Data Collection

a) Synthetic Datasets

Datasets	**Z**	***ld***
Syn-one	1.8→1.8→1.8→1.8	0%→10%→0%→20%
Syn-two	1.8→3.0→2.5→4.5	0%→0%→0%→0%
Syn-three	1.8→1.8→3.5→3.5	0%→10%→0%→20%

They all contain three concept drifts and consist of 100,000 samples.

b) Real-world Datasets

Dataset	***N***	***L***	***A***	***Z***	***LDens***
tmc2007-500	28596	22	500	2.16	0.10
20NG	19300	20	1001	1.03	0.05

B. Evaluation Measures

a) Hamming-accuracy

$$Hamming-accuracy=\frac{1}{NL}\sum_{i=1}^{N}\sum_{j=1}^{L}I(y_i^{(j)}=\hat{y}_i^{(j)})$$

b) Subset Accuracy

$$SubsetAccuracy=\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{L}y_i^{(j)}\wedge\hat{y}_i^{(j)}}{\sum_{j=1}^{L}y_i^{(j)}\vee\hat{y}_i^{(j)}}$$

c) F1-macro

$$F1-macro=\frac{1}{L}\sum_{j=1}^{L}F1\left[\left(y_1^{(j)},\ldots,y_N^{(j)}\right),\left(\hat{y}_1^{(j)},\ldots,\hat{y}_N^{(j)}\right)\right]$$

C. Experiment Design

a) Verification of the label dependence’ s availability

DD with Case1: No Grouping on Labels;

DD with Case2: Call GL1();

DD with Case3: Call GL2().

b) Contrastive experiments with other methods

Method1: weight by the classification accuracy [1]

Method2: weight against the classification accuracy [2]

Method3: weight decay along with the incorrect classification [3]

D. Experimental Results and Discussion

a) The results for the verification experiments

The experimental results of Drift Detection with Case1, Case2 and Case3 over the dataset Syn-one

Measures	Methods	The Minimum Support	
		min = 0.1	min = 0.05
Subset Accuracy	DD with Case1	0.3081±0.0366	0.3081±0.0366
	DD with Case2	**0.3206±0.0372**	**0.3326±0.0410**
F1-macro	DD with Case1	0.2687±0.0419	0.2687±0.0419
	DD with Case2	**0.2855±0.0397**	**0.3031±0.0505**

Measures	Methods	The Number of Clusters		
		K = 3	K = 4	K = 5
Subset Accuracy	DD with Case1	0.3081±0.0366	0.3081±0.0366	0.3081±0.0366
	DD with Case3	**0.3263±0.0395**	**0.3257±0.0292**	**0.3203±0.0352**
F1-macro	DD with Case1	0.2687±0.0419	0.2687±0.0419	0.2687±0.0419
	DD with Case3	**0.2883±0.0509**	**0.2935±0.0383**	**0.2825±0.0411**

DD with Case2 and Case3 can get better results than Case1.

The experimental results of Drift detection with Case1, Case2 and Case3 over the dataset Syn-three

Measures	**Methods**	**The Minimum Support**	
		min = 0.1	min = 0.05
Subset Accuracy	DD with Case1	0.3256±0.0256	0.3256±0.0256
	DD with Case2	**0.3408±0.0308**	**03438±0.0320**
F1-macro	DD with Case1	0.2743±0.0364	0.2743±0.0364
	DD with Case2	**0.3007±0.0367**	**0.2956±0.0367**

DD with Case2 and Case3 can get better results than Case1.

Measures	**Methods**	**The Number of Clusters**		
		K = 3	K = 4	K = 5
Subset Accuracy	DD with Case1	0.3256±0.0256	0.3256±0.0256	0.3256±0.0256
	DD with Case3	**0.3384±0.0263**	**0.3525±0.0289**	**0.3301±0.0321**
F1-macro	DD with Case1	0.2743±0.0364	0.2743±0.0364	0.2743±0.0364
	DD with Case3	**0.2887±0.0278**	**0.3021±0.0300**	**0.2775±0.0371**

The experimental results of Drift detection with Case1, Case2 and Case3 over the dataset Syn-two

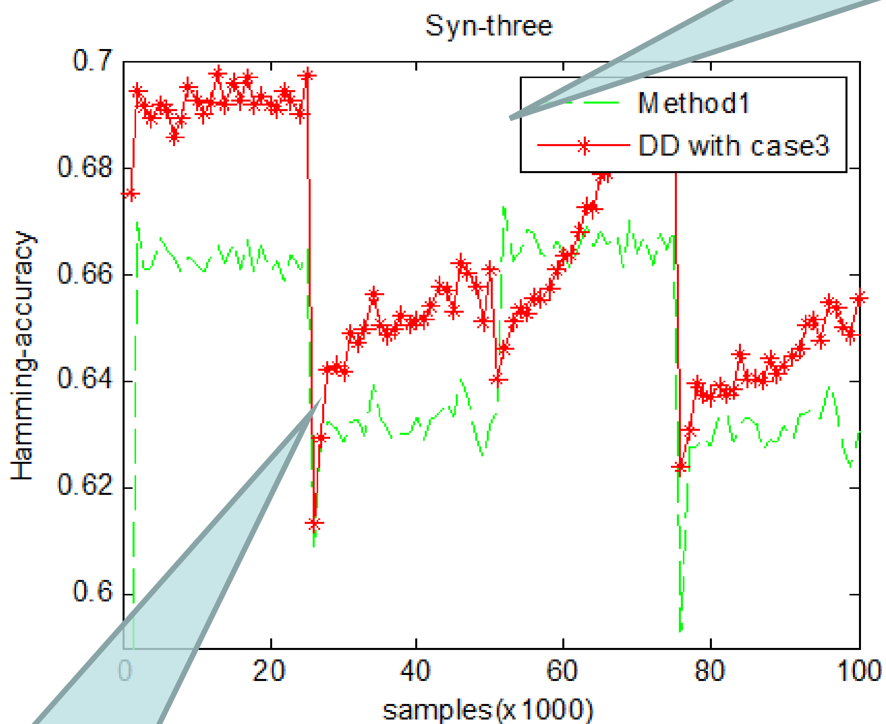
Measures	Methods	The Minimum Support	
		min = 0.1	min = 0.05
Subset Accuracy	DD with Case1	**0.3794±0.0276**	**0.3794±0.0276**
	DD with Case2	0.3743±0.0281	0.3766±0.0314
F1-macro	DD with Case1	**0.3225±0.0371**	**0.3225±0.0371**
	DD with Case2	0.3123±0.0371	0.3172±0.0334

DD with Case1 is only slightly better than DD with Case 2 and Case3.

Measures	Methods	The Number of Clusters		
		K = 3	K = 4	K = 5
Subset Accuracy	DD with Case1	**0.3794±0.0276**	**0.3794±0.0276**	**0.3794±0.0276**
	DD with Case3	0.3752±0.0395	0.3779±0.0354	0.3763±0.0345
F1-macro	DD with Case1	**0.3225±0.0371**	**0.3225±0.0371**	**0.3225±0.0371**
	DD with Case3	0.3155±0.0469	0.3211±0.0462	0.3160±0.0384

b) The results for the verification experiments

DD with case3 make reaction for these concept drifts at 25, 50 and 75



DD with case3 recover more rapidly for 1th and 3th concept drift

Methods	**Synthetic Datasets**		
	Syn-one	Syn-two	Syn-three
Method1	0.6430±0.0216	0.6484±0.0226	0.6389±0.0239
Method2	0.6528±0.0448	0.6573±0.0254	0.6551±0.0254
Method3	0.6382±0.0380	0.6464±0.0371	0.6391±0.0417
DD with case3	**0.6546±0.0222**	**0.6621±0.0212**	**0.6572±0.0226**

From the table: DD with case3 achieves high predictive performance compared with these three baseline methods.

Results on 20NG

Methods	Measures		
	Hamming-accuracy	Subset Accuracy	F1_macro
Method1	0.7878	0.2361	0.3039
Method2	0.8080	0.2580	0.3085
Method3	0.9220	0.3215	0.3728
DD with case3	**0.9447**	**0.3598**	**0.3903**

DD with case3 performs outstandingly, compared with Method1, Method2 and Method3.

Results on tmc2007-500

Methods	Real World Datasets		
	Hamming-accuracy	Subset Accuracy	F1_macro
Method1	0.8150	0.3334	0.4007
Method2	0.8224	0.3737	0.4186
Method3	0.9001	0.4087	0.4378
DD with case3	**0.9143**	**0.4303**	**0.4676**

Conclusion

- Conclusion :

  In this paper we have analyzed the unique properties of drift detection for multi-label data streams and proposed a drift detection method based on label grouping and entropy.

  We instantiate the label grouping with the *Apriori* and *EM* algorithm. The verification and contrastive experiments all show that the proposed method is promising.

- Future work :

  We will attempt to integrate the proposed method of drift detection with *Hoeffding Tree* algorithm to deal with multi-label evolving stream classification problem.

Acknowledgement

Special Thanks to :

- Prof. Yi-min Wen

Thanks to Projects:

- Guangxi Key Laboratory of Trusted Software
- Innovation Project of GUET Graduate Education
- National Natural Science Foundation of China
- Science and Technology Plan Project of Hunan Province

Thanks to Source Software:

- The open source software development kit WEKA[4] and MOA[5]

References

[1] W. Qu, Y. Zhang, J. P. Zhu and Y. Wang. “Mining Multi-label Concept-Drifting Streams Using Ensemble Classifiers,”

[2] D. Brzeziński, J. Stefanowski. “Accuracy updated ensemble for data streams with concept drift,”

[3] A. Bifet, G. Holmes, R. Kirkby and B. Pfahringer. “Moa: Massive online analysis,”

[4] http://www.cs.waikato.ac.nz/ml/index.html

[5] http://moa.cs.waikato.ac.nz/

Thanks for Your Attention

Question ？