A Survey on Online Feature Selection

S. Nivetha[1], S. Nithya Roopa[2]

[1]M.E, Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049

[2]Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore-641049

Abstract: ***Feature selection is one of the main techniques used in data mining. In defiance of its consequence, most learning of feature selection is limited to batch learning. Dissimilar to existing batch learning methods, online learning can be elected by an encouraging family of well-organized and scalable machine learning algorithms for large-scale approach. The greatest quantity of online learning need to retrieve all the attributes/features of occurrence. Such a simple surroundings are not invariable for real-world applications when statistics illustration is of high-dimensionality. The problem of Online Feature Selection (OFS) is that online learner is allowed to maintain a classifier which involved only a small and fixed number of features.Online feature selection is to make accurate prediction for an object using a small number of active features. This article addresses two different tasks of online feature selection: 1)learning with full input 2) learning with partial input. The proposed system presents novel algorithm to solve each of the two problems and give their performance analysis.***

Keywords: feature selection, classification

1. Introduction

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems.

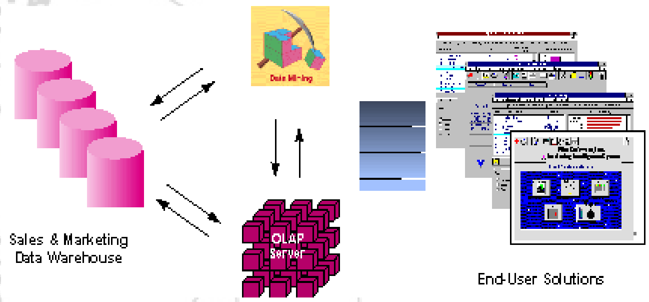
1.1 The foundation of Data Mining

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

1.2 Architecture of Data Mining

To best apply these advanced techniques, they must be fully integrated with a data warehouse as well as flexible interactive business analysis tools. Many data mining tools currently operate outside of the warehouse, requiring extra steps for extracting, importing, and analyzing the data. Furthermore, when new insights require operational implementation, integration with the warehouse simplifies the application of results from data mining. The resulting analytic data warehouse can be applied to improve business processes throughout the organization.



An OLAP (On-Line Analytical Processing) server enables a more sophisticated end-user business model to be applied when navigating the data warehouse. The multidimensional structures allow the user to analyze the data as they want to view their business – summarizing by product line, region, and other key perspectives of their business. The Data Mining Server must be integrated with the data warehouse and the OLAP server to embed ROI-focused business analysis directly into this infrastructure.

1.3 Data Mining Tasks

- Classification
- Clustering
- Association Rule Discovery
- Sequential Pattern Discovery
- Regression
- Deviation Detection

1.4 Challenges of Data Mining

- Scalability
- Dimensionality
- Complex and Heterogeneous Data
- Data Quality
- Data Ownership and Distribution
- Privacy Preservation
- Streaming Data

2. Related Work

2.1 OTL A Framework of Online Transfer Learning

A new machine learning called OTL (Online Transfer Learning) is implemented in this paper. OTL transfers the knowledge from source domain to the target domain. Data can be in both domains[1].Data can be present in different class and feature distribution. The proposed technique addresses two kinds of OTL: one is to perform OTL in a homogeneous domain and the other is to perform OTL across heterogeneous domains. In proposed system OTL algorithm is used. There are two problems in this paper, 1) homogeneous OTL in the target domain shares the same features as the source / old one.2) Addresses the challenge of heterogeneous OTL where the feature space of the target domain is different from the source domain.

The goal of Transfer Learning (TL) is to extract the knowledge from one or more source domain and then apply them to the target. Methods are roughly classified into inductive, transductive and unsupervised. **Inductive TL** is used to induce the target domain where the knowledge is transferred from the source domain. **Transductive TL** is used to extract the knowledge from source domain to improve the prediction task in the target domain without labeled data in the target. **Unsupervised TL** is used to resolve the unsupervised learning task in target domain. TL is classified as homogeneous and heterogeneous.

w1,t and w2,t can be adjusted dynamically. We suggest the following updating scheme for adjusting the weights:

$$w_{1,t+1} = \frac{w*s(h)}{w*s(h)+w*s(f)}$$

When using the square loss ℓ_ (z, y) = (z − y)2 *for* z ∈ [0, 1] *and* y ∈ {0, 1} *and the above exponentially weighting update method and setting* η = 1/2, *we have the bound of the ensemble algorithm as:*

$$\sum_{t=1}^{T} l^{*}(w_{1,t}\prod(h(x_{2t}))+w_{2,t}\prod(f_t(x_{2t})),\prod(y_{2t}))\leq 2\ln 2+ \min\{\sum_{t=1}^{T} l^{*}(\prod(h(x_{2t})),\prod(y_{2t})), \sum_{t=1}^{T} l^{*}(\prod(f_t(x_{2t})),\prod(y_{2t}))\}$$

2.2 Online Feature Selection for Mining Big Data

Online learning requires all the attributes/features of training instance. The problem of Online Feature Selection (OFS) is that it the online learner is allowed only to maintain a classifier which involved a small and fixed number of features and to solve[2] the feature selection problem by an online learning method. The goal of OFS is to develop online classifiers that only fixed a small number of features. The goal of Feature Selection (FS) is to select the most relevant features in the whole feature space to improve the prediction performance of the predictors. FS is divided into 3 categories: filter, wrapper and embedded.

Advantage
- More scalable.
- Efficiency.
- Efficacy.

Disadvantage
- Large number of online learning.

2.3 Efficient Learning with partially observed attributes

Three variants of budget learning is used. The learner is allowed to access a limited number of training set[3].
LOCAL BUDGET: A constraint is imposed on the number of available attributes per training.
GLOBAL BUDGET: Overall accessible training attributes are constrained.
PREDICTION ON A BUDGET: Constraints on the number of available attributes.

The result is complemented for general lower bound. This paper uses local budget constraints: A baseline algorithm. It describes the straightforward adaption to the local budget setting. The adaption is based on a direct non adaptive estimation of the loss function. A popular approach of learning a linear regression is minimizing the empirical loss on the training set. In this conclusion, learning algorithm finds a ε- good predictor using k attributes.

Advantage
Efficiency

Drawback
The EM (Expectation Maximization) approach is that it might find sub optimal solutions.

2.4 Online Multiple Kernel Learning

Online Multiple Kernel Learning (OMKL) is to learn a kernel based prediction to pool of predefined kernels in an online learner. [4]It challenges the online learning because both the kernel classifiers and their linear combination weights must be learned.

Two setups for OMKL: Combining binary prediction or real-valued outputs from multiple kernel classifiers. Proposed algorithm uses both deterministic and stochastic approaches in OMKL. Stochastic approach randomly chooses a classifier for updating some sampling strategies. Mistake bound is derived for all by proposed OMKL algorithm. The combination of multiple kernels order is to optimize the kernel based learning method. Online learning framework for multiple kernels is based on the two types of online learning technique:
- Perceptron algorithm - learn a classifier for a given kernel.
- Hedge algorithm - linearly combines multiple classifier.

The proof essentially combines the proofs of the Perceptron algorithm and the Hedge algorithm. First, following the analysis in, we can easily have

$$\sum_{t=1}^{T} q_t \cdot z_t \leq \frac{\ln(\frac{1}{\beta})}{1-\beta} \min \sum_{t=1}^{T} z_i^{t} + \frac{\ln m}{1-\beta}$$
$1 \leq i \leq m$

The result is
$$M = \sum_{t=1}^{T} I(q_t \cdot z_t \geq 0.5) \leq 2 \sum_{t=1}^{T} q_t$$

Advantage:
- To design their own kernel matrices.
- Video features.
- Audio features.
- Text features.

Drawback:
- Less interpretable.
- Computationally expensive.

2.5 LIBOL: A Library for Online Learning Algorithms

It is an open source to use library for large scale online learning. It is efficient and scalable among[5] state-of-the-art online learning algorithms for online large-scale classification task.It is easy-to-use command line tools. It is not only a machine learning tool box, but also a comprehensive platform for conducting online learning research.

Advantage:
- Efficiency
- Scalability
- Parallelization
- Adaptability

The **drawback** is that potentially informative higher order interactions may be overlooked.

S. No	*Algorithm*	*Merits*	*Demerits*
1	Online learning algorithm	Attractive efficiency Scalability	A disadvantage of these approaches is that they rely on training datasets.
2	Online learning algorithm	More scalable. Efficiency. Efficacy	Large number of online learning
3	Baseline algorithm	Efficiency	The EM (Expectation Maximization) approach is that it might find sub optimal solutions.
4	OMKL-p algorithm	To design their own kernel matrices. Video features. Audio features. Text features.	Less interpretable. Computationally expensive.
5	Online learning algorithm for linear classifier.	Efficiency. Scalability. Parallelization. Adaptability.	The **drawback** is that potentially informative higher order interactions may be overlooked.

3. Conclusion

Online Transfer Learning is used to transfer the knowledge from source to the target domain. It presented the two novel OTL algorithms which can be used in the proposed system. It extensively examined their empirical performance. Online Feature Selection for mining big data aims to select the fixed number of features for prediction and compare the proposed Online Feature Selection technique with state-of-the-art batch feature selection algorithm. It is more efficient and scalable than some state-of-the-art batch feature selection.

The goal of the LIBOL is easy learning with massive data streams to tackle the grand challenge of big data analytics. Online Multiple Kernel Learning proposed the framework of OMKL by learning the combination of multiple kernels from a pool.

References

[1] P. Zhao and S. C. H. Hoi. OTL: A framework of online transfer learning. In *ICML*,pages 1231–1238, 2010.R. Caves, Multinational Enterprise and Economic Analysis, Cambridge University Press, Cambridge, 1982. (book style)

[2] S. C. H. Hoi, J. Wang, P. Zhao, and R. Jin. Online feature selection for mining big data. In *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, Big Mine '12, pages 93–100, New York, NY, USA, 2012. ACM.

[3] N. Cesa-Bianchi, S. Shalev-Shwartz, and O. Shamir. Efficient learning with partially observed attributes. *Journal of Machine Learning Research*, pages 2857–2878, 2011.

[4] S. C. H. Hoi, R. Jin, P. Zhao, and T. Yang. Online multiple kernel classification. *Machine Learning*, 90(2):289–316, 2013.

[5] S. C. H. Hoi, J. Wang, and P. Zhao. *LIBOL: A Library for Online Learning Algorithms*. Nan yang Technological University, 2012.

[6] J. Wang, P. Zhao, and S. C. H. Hoi. Cost-sensitive Online classification. In *ICDM*, pages 1140–1145, 2012

[7] A. Rostamizadeh, A. Agarwal, and P. L. Bartlett. Learning with missing features. In *UAI*, pages 635–642, 2011.

[8] Yang, Haiqin, Xu, Zenglin, King, Irwin, and Lyu, Michael. Online learning for group lasso. In *27th Int'l Conf. on Machine Learning (ICML2010)*, Haifa, Israel, 2010.

Author Profile

S. Nivetha ME (Computer Science & Engineering) Kumaraguru College Of Technology, Coimbatore.

S. Nithya Roopa, Assistant Proffessor (Department of Computer Science & Engineering) Kumaraguru College of Technology, Coimbatore.