Faculty of Engineering and Information Technology
University of Technology, Sydney

Learning label dependency for multi-label classification

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Bin Fu

February 2018

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor Associate Prof. Guandong Xu for the continuous support of my Ph.D study and research, for his patience, motivation, and immense knowledge. His guidance helped me during the whole time of research and writing of this thesis. I could not have imagined having a better advisor for my Ph.D study.

I also would like to appreciate my co-supervisor Prof. Longbing Cao for providing me with continuous support throughout my Ph.D study and research. Without his professional guidance and persistent help, this thesis would not have been possible.

I thank my fellow labmates in Advanced Analytics Institute: Fangfang Li, Junfu Yin. xueping Peng, Xiao zhu for the inspiring discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in these four years.

Last but not the least, I would like to thank my family: my wife and my parents, for their unconditional support, both financially and emotionally throughout the whole PhD studying.

Bin Fu

January 2018 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Published

- **Bin Fu**, Guandong Xu, and Longbing Cao, et al. Coupling multiple views of relations for recommendation. Lecture Notes in Artificial Intelligence, 2015, 9078. pp. 732-743.

- **Bin Fu**, Zhihai Wang, and Guandong Xu et al. Multi-label learning based on iterative label propagation over graph. Pattern Recognition Letters, 2014, 42. pp. 85-90.

- **Bin Fu**, Guandong Xu, and Zhihai Wang, et al. Leveraging supervised label dependency propagation for multi-label learning. In: Proceedings of the 13th IEEE International Conference on Data Mining (ICDM2013), Dallas, USA, 2013. pp. 1061-1066.

- **Bin Fu**, Zhihai Wang, and R. Pan et al. Learning tree structure of labels dependency for multi-label learning. In: Proceedings of 16 Pacific-Asia Knowledge Discovery and Data Mining (PAKDD 2012). Kuala Lumpur, Malaysia, 2012. pp. 159-170.

- Guandong Xu, **Bin Fu**, and Yanhui Gu. Point-of-interest recommendations via a supervised random walk algorithm, IEEE Intelligent Systems, 31(1), 2016. pp. 15-23.

- Charles Chu, Guandong Xu, James Brownlow, **Bin Fu**. Deployment of churn prediction model in financial services industry. In: proceedings of 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), 2016, pp. 1-2.

Papers to be Submitted/Under Review

- **Bin Fu**, Guandong xu, and Zhihai Wang. Label ranking by learning preference between positive and missing labels

- **Bin Fu**, Guandong xu, and Zhihai Wang. Multi-label learning with missing labels using boosted matrix factorization.

Research Reports of Industry Projects

- Discovering Deep Insights Into SG Contribution. Colonial First State, Dec 2016.

- Discovering Deep Insights Into Customer Retention. Colonial First State, Dec 2015.

Abstract

Multi-label classification is an important topic in the field of machine learning. In many real applications, there exist potential dependencies or correlations between labels, and exploiting the underlying knowledge could effectively improve the learning performance. Therefore, how to learn and utilize the dependencies between labels has become one of the key issues of multi-label classification.

This thesis firstly summarizes existing works and analyses their advantages and disadvantages. Several effective methods for multi-label classification are then proposed, focusing on ways of exploiting various types of label dependencies. The contributions of this thesis mainly include:

(1) A method that uses a tree-structured restricted Bayesian network to represent the dependency structure of labels is proposed. This work is inspired by the ClassifierChain method. Compared with ClassifierChain, our method's advantage is that the dependencies between labels are represented using a Bayesian network rather than a randomly selected chain, so more appropriate label dependencies could be determined. Furthermore, ensemble learning technique is used to construct and combine multiple tree-structured Bayesian networks, thus the mutual dependencies between labels could be fully exploited and the final model could be more robust. The experimental results verify the effectiveness of these methods. Compared with other baselines, the show better performance due to more appropriate label dependencies are captured.

(2) A common strategy of exploiting label dependencies is, for every la-

bel, to the labels it depends on and use these labels as auxiliary features in the training phase. The issues of this strategy are that the influence of label dependencies could be depressed by existing features and indirect label dependencies could not be taken into consideration. Therefore, a new learning paradigm that separates the influence of existing features and labels is introduced, and the impact of label dependencies could be well intensified in this way. Moreover, a method that models the propagation of label dependencies as a RWR process (Random Walk with Restart) is proposed. In this method, label dependencies are encoded as a graph, and the dynamic and indirect dependencies between labels are utilized through the RWR process over the label graph. The experimental results validate this method, showing that it outperforms other baselines in terms of learning a label ranking.

(3) Based on above method, a method that takes multiple factors into consideration when learning label dependencies is proposed. In this method, dependency between two labels is characterized from different perspectives, and is determined by learning a linear combination of multiple measures. A particular loss function is designed, and thus the optimal label dependencies, i.e., the dependency matrix in RWR process, can be obtained by minimizing the loss function. The advantage of this method include: a) label dependencies are measures and combined from different perspectives, and b) label dependences that are optimal to a particular loss function now are obtained. The experimental results indicate that this method could further learn a better label ranking compared with the previous one, given an explicit loss function.

(4) A novel method that learns label ranking by exploiting preferences between true labels and other labels is proposed. In this method, the original instance space and label space are mapped into a low-dimensional space using matrix factorization technique. Therefore, one advantage of the method is that the number of label is reduced greatly, and problem with massive labels now can be handle efficiently. Moreover, a loss function is formulated based on the assumption that an instance's true labels which have been

given explicitly should be ranked before other labels which are not provided explicitly. It is then used to guide the process of matrix factorization and label ranking learning. The advantage of this novel assumption is that it alleviate issue in traditional assumption that if a label is not given explicitly, it should not be a true label. Therefore, this method is also applicable to data that are partially labelled. Its effectiveness is validated by the experimental result which shows that it could rank explicitly given label well before other labels for a given instance.

In summary, this thesis has proposed several effective methods that exploit label dependencies from different perspectives, and their effectiveness have been validated by experiments. These achievements lay a good foundation for further research and applications.