Frontier and Future Directions

12

12.1 Overview

In this book, we have introduced the current research works on broad learning and its applications on online social networks. This book has covered 4 main parts, where the first 3 parts include 6 main research directions about broad learning based social network mining problems, including (1) *network alignment*, (2) *link prediction*, (3) *community detection*, (4) *information diffusion*, (5) *viral marketing*, and (6) *network embedding*. These problems introduced in this book are all very important for many concrete real-world social network applications and services. A number of state-of-the-art algorithms have been proposed to solve these problems, which are introduced in great detail in this book. *Broad learning* is a very promising research area, and several potential future development directions about broad learning will be illustrated in the following sections.

12.2 Large-Scale Broad Learning

Data generated nowadays is usually of very large scale, and fusion of such big data from multiple sources together will render the scalability problem much more challenging. For instance, the online social networks (like Facebook) usually involve millions even billions of active users, and the social data generated by these users in each day will consume more than 600 TB storage space (in Facebook). One of the major future developments about *broad learning* is to develop scalable fusion and mining algorithms that can handle such a large volume challenge (of big data). One tentative approach is to develop information fusion and mining algorithms based on distributed platforms, like Spark and Hadoop [1], and handle the data with a large distributed computing cluster. Another method to resolve the scalability challenge is from the model optimization perspective. Optimizing the existing learning algorithms and proposing new approximated learning algorithms with a much lower time complexity are desirable in the future research projects. In addition, applications of the latest deep learning models to learn the low-dimensional latent representations of such large-scale datasets can be another alternative approach to achieve the scalable *broad learning* objective.

J. Zhang, P. S. Yu, *Broad Learning Through Fusions*,
https://doi.org/10.1007/978-3-030-12528-8_12

12.3 Multi-Source Broad Learning

Current research works on multiple source data fusion and mining mainly focus on aligning entities in one single pair of data sources (i.e., two sources), where information exchange between the sources mainly relies on the anchor links between these aligned entities. Meanwhile, when it comes to fusion and mining of multiple (more than two) sources, the problem setting will be quite different and become more challenging. For example, in the alignment of more networks, the transitivity property of the inferred anchor links needs to be preserved [3]. Meanwhile, in the information transfer from multiple external aligned sources to the target source, the information sources should be weighted differently according to their importance. How to determine the different weight parameters of these sources is still an open problem by this context so far. Therefore, the diverse variety of the multiple sources will lead to more research challenges and opportunities. New information fusion and mining algorithms for the multi-source scenarios can be another great opportunity to explore *broad learning* in the future.

12.4 Broad Learning Applications

Besides the research works on online social networks, the third potential future development direction of broad learning lies its broader applications on various categories of datasets, like enterprise internal data [4, 5, 8, 9], geo-spatial data [2, 6, 7], knowledge graph base data, financial time series data, images/video data, as well as pure text data. Some prior research works on fusing enterprise context information sources, like enterprise social networks, organizational chart, and employee profile information have been done already [4, 5, 8, 9]. Several interesting problems, like organizational chart inference [5], enterprise link prediction [4], information diffusion at workplace [8], and enterprise employee training [9], have been studied based on the fused enterprise internal information. In addition, analysis of the correlation of different traveling modalities (like shared bicycles [2, 6, 7], bus and metro train) with the city zonings in smart city; and fusing multiple knowledge bases, like Douban and IMDB, for knowledge discovery and truth finding can all be good application scenarios for broad learning.

12.5 Summary

We would like to conclude this chapter and this book with the acknowledgement for the readers, who have finished reading the whole book. We really hope the readers can learn useful knowledge from the book, and use the knowledge in your courses, your research projects, and your work.

12.6 Exercises

1. (Easy) Please briefly describe the broad learning problems we have covered in this textbook.
2. (Hard) Please think about some potential problems we can study based on broad learning in the future.

References

1. S. Jin, J. Zhang, P. Yu, S. Yang, A. Li, Synergistic partitioning in multiple large scale social networks, in *2014 IEEE International Conference on Big Data (Big Data)* (2014)
2. Q. Zhan, J. Zhang, X. Pan, P. Yu, Discover tipping users for cross network influencing, in *2016 IEEE 17th International Conference on Information Reuse and Integration (IRI)* (2016)
3. J. Zhang, P. Yu, Multiple anonymized social networks alignment, in *2015 IEEE International Conference on Data Mining* (2015)
4. J. Zhang, Y. Lv, P. Yu, Enterprise social link prediction, in *Conference on Information and Knowledge Management* (2015)
5. J. Zhang, P. Yu, Y. Lv, Organizational chart inference, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)* (2015)
6. J. Zhang, X. Pan, M. Li, P. Yu, Bicycle-sharing system analysis and trip prediction, in *2016 17th IEEE International Conference on Mobile Data Management (MDM)* (2016)
7. J. Zhang, X. Pan, M. Li, P. Yu, Bicycle-sharing systems expansion: station re-deployment through crowd planning, in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPACIAL '16)* (2016)
8. J. Zhang, P. Yu, Y. Lv, Q. Zhan, Information diffusion at workplace, in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)* (2016)
9. J. Zhang, P. Yu, Y. Lv, Enterprise employee training via project team formation, in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)* (2017)