Mining Advisor-Advisee Relationships in Scholarly Big Data: A Deep Learning Approach

Wei Wang, Jiaying Liu, Shuo Yu, Chenxin Zhang, Zhenzhen Xu*, Feng Xia
Mobile and Social Computing Lab: http://msclab.org/ xzz@dlut.edu.cn
School of Software, Dalian University of Technology, Dalian 116620, China

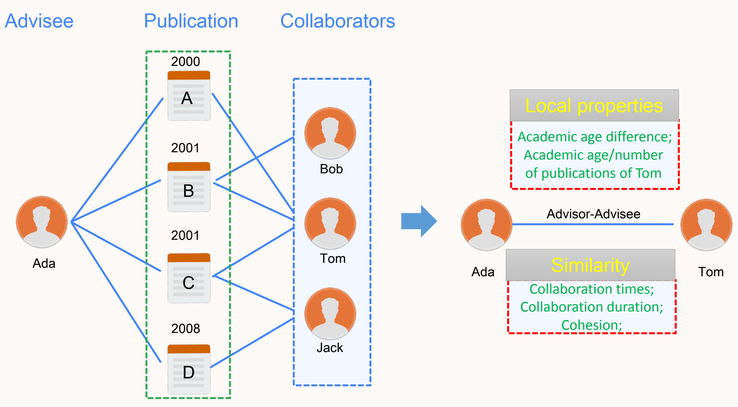
Introduction

Problem: Analyzing advisor-advisee relationships is of grate importance. However, there is no suitable large-scale mentorship data set.

Idea: Advisor-advisee relationships among researchers are hidden in scientific collaboration networks.

Solution: We proposed DAMer (Deep Advisor Miner), a deep-learning-based method which considers both the personal properties and network characteristics to automatically extract the relationships from scholarly big data.

Result: Experiments demonstrate that DAMer has better performance compared with other state-of-the-art methods in precision (91%).



Data set

The advisor-advisee pairs are collected from 16 famous universities such as Carnegie Mellon and Stanford in the field of computer science from PhDtree project. The data set contains 3,423 advisee-advisor relationships. We then gain collaborators of each advisee and their properties from DBLP digital library.

Conclusion

(1) Defining and calculating features that can measure the similarity between advisors and advisees extracted from scientific collaboration networks.
(2) Crawling advisor-advisee pairs from PhDtree project and matching them with DBLP to gain the ground truth data set.
(3) A novel method based deep learning algorithms (DAMer) is proposed and experimental results demonstrate the effectiveness (91% in precision) of DAMer.

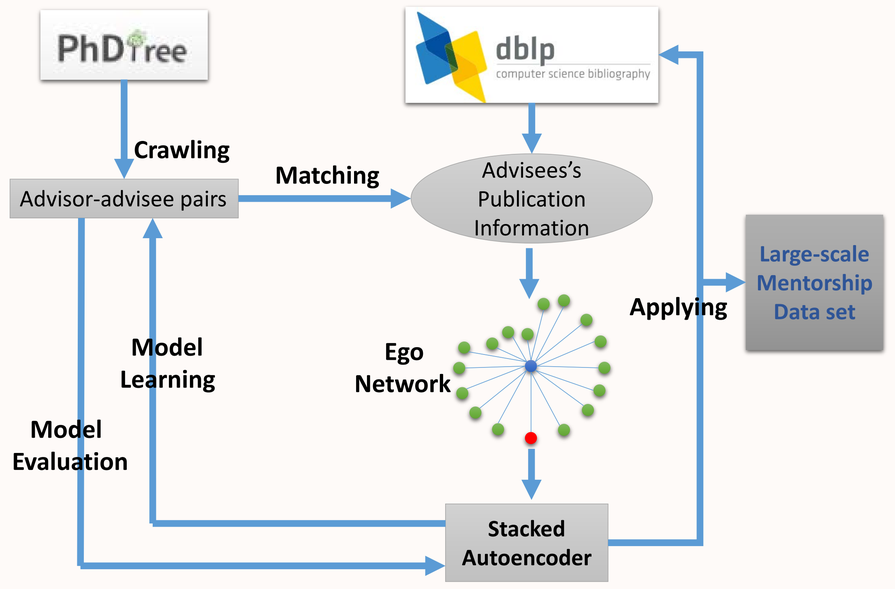
References

[1] R. D. Malmgren, J. M. Ottino, and L. A. N. Amaral. The Role of Mentorship in Protégé Performance. *Nature*, 465(7298):622–626, 2010.

[2] C. Wang, J. Han, Y. Jia, J. Tang, D. Zhang, Y. Yu, and J. Guo. Mining Advisor-Advisee Relationships From Research Publication networks. In *Proceedings of the 16th KDD*, pages 203–212. ACM, 2010.

Methods

The DAMer employs the Stacked Autoencoder (SAE) as the foundation algorithm, which is a famous deep learning model. The SAE model allows users to easily inject the personal features and network characteristics as input without the manual effort of feature selection.



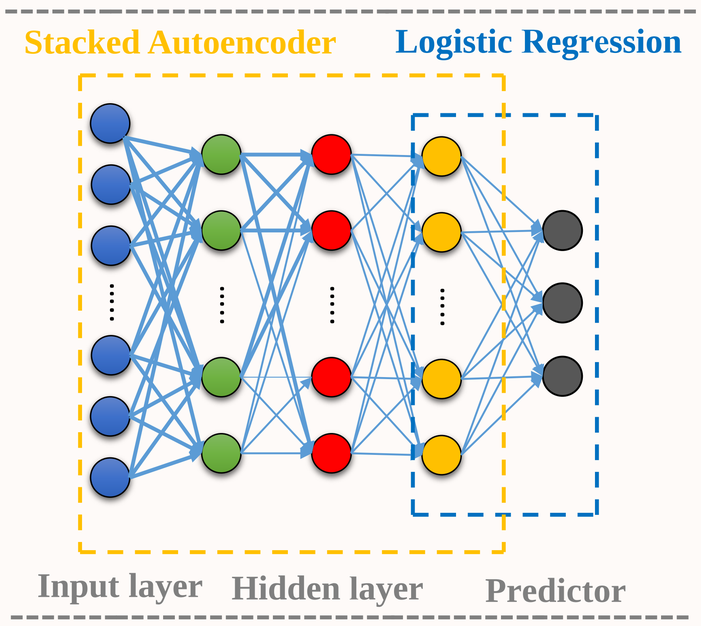


Figure 1: Overall structure of DAMer(left); Framework of stacked autoencoder (right)

Results

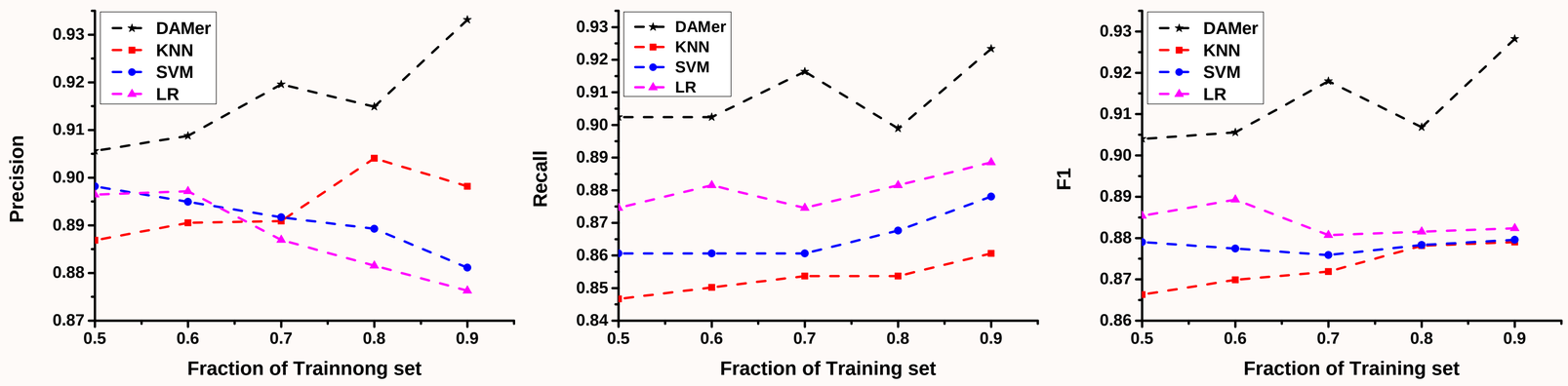


Figure 2: Comparison of precision, recall, and F1 of DAMer with/witou academic age

From Fig. 2, we can see that DAMer performs better than the-state-of-art machine learning algorithms and can reach a precision of 91%.

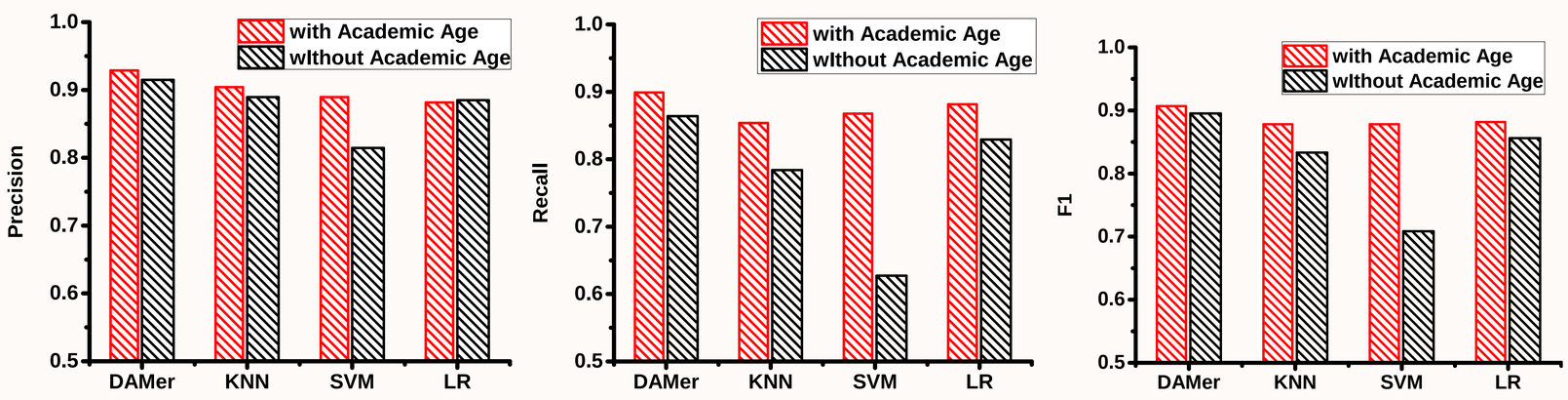


Figure 3: Comparison of precision, recall, and F1 of DAMer with other machine learning algorithms.

We can gain the conclusion from Fig. 3 that considering the factor of academic age can improve the performance of advisor-advisee identification.

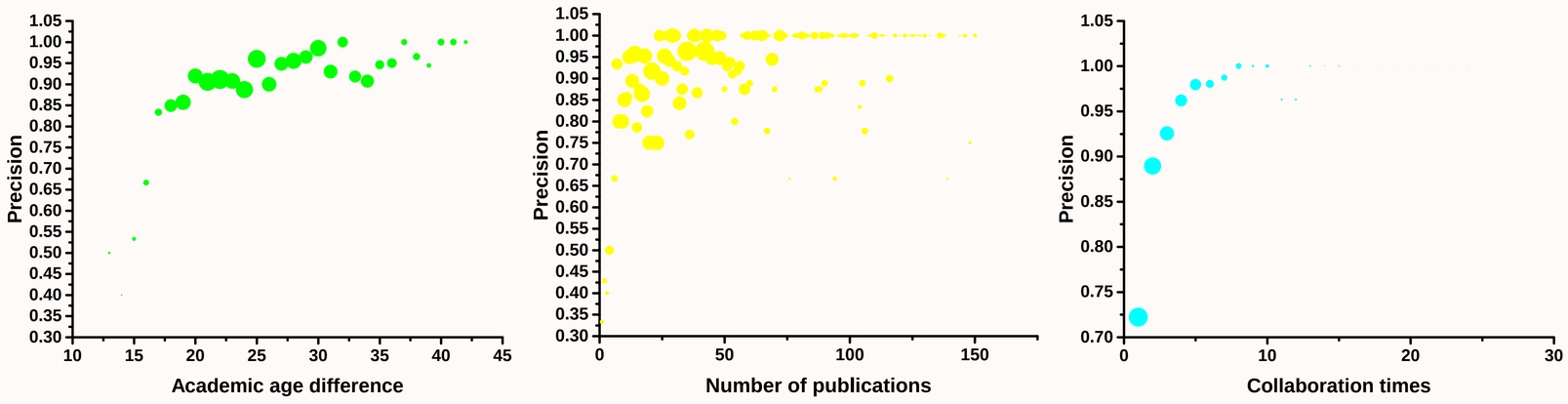


Figure 4: Effect of input features on the performance of SAE in precision.

Fig. 4 shows that DAMer has better performance in identifying advisor-advisee pairs with more academic publications, higher academic age, and more collaboration times.