

Big Data with Ensemble Selection

Yanan Xiao, *Student Member, IEEE*, Aziza Al Sawafi

Abstract—We present our approach to the big data problem put forward in KDD Cup 2009. We carry out a more technical analysis of our initial plan, and then make some modification to it. Later we run our process in Amazon cloud servers. We obtain a quite satisfactory result, the average score is 75%.

Keywords—Customer relationship prediction, ensemble selection, classification, data mining

Aziza Al Sawafi First year Computing and Information Science student at Masdar Inst., got a bachelor degree in Network Engineering (United Arab Emirates University). Sport, drawing, designing, blogging, reading poems, photography, and riding horse/bicycle are my interests beside all things that are related to networking and computer science.

I. INTRODUCTION

WE proposed an approach to tackle the KDD Cup 2009 in [1]. We planned to carry out a meta classification method, i.e. ensemble selection with a reduced set of base classifiers when compared with [2].

II. INITIAL PROPOSAL

In this section we describe what we proposed as our initial plan.

III. REVISED PROPOSAL

IV. RESULT AND CONCLUSION

Well, the result shall come out. Eventually.

ACKNOWLEDGMENT

The authors would like to thank Dr. Wei Lee for giving high quality data mining lectures and selecting this challenging but rewarding topic as this semester's project. They would like to give more gratitude to Masdar Institute for the studying and researching environment.

REFERENCES

- [1] Y. Xiao and A. Sawafi, "Getting the max out of ensemble selection," *Journal of Masdar Institute*, vol. 1, pp. 323–330, Dec. 2013.
- [2] A. Niculescu-Mizil *et al.*, "Winning the KDD Cup orange challenge with ensemble selection," in *Proceedings of KDD-Cup 2009 competition*, Paris, France, Jun. 2009, pp. 23–34.

Yanan Xiao A first year master student as well as IEEE student member in CIS program, Masdar Institute. He loves programming when all the coursework is finished. When he feels tired of programming, he would read some books.

Y. Xiao, A. Sawafi are 1st year master students with the Department of Electrical Engineering and Computer Science, Masdar Institute of Science and Technology, Masdar City, Abu Dhabi, UAE. P.O. 54224 Email: {yxiao,aalsawafi@masdar.ac.ae}.

One team member quit Masdar Institute not soon after the project. But we managed to get things done by ourselves.