

# Detecting Fraudsters in Online Auction Using Variations of Neighbor Diversity

Laksamee Khomnotai<sup>1,3,\*</sup>, Jun-Lin Lin<sup>1,2</sup>

<sup>1</sup>Department of Information Management, Yuan Ze University, Taoyuan, Taiwan, ROC.

<sup>2</sup>Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Taiwan, ROC.

<sup>3</sup>Faculty of Management Science, Nakhon Ratchasima Rajabhat University, Nakhon Ratchasima, Thailand.

Received 13 March 2015; received in revised form 24 June 2015; accepted 01 July 2015

## Abstract

Inflated reputation fraud is a serious problem in online auction. Recent work suggested that neighbor diversity is an effective feature for discerning fraudsters from normal users. However, there exist many different methods to quantify diversity in the literature. This raises the problem of finding the most suitable method to calculate neighbor diversity for detecting fraudsters. We collect four different methods to quantify diversity, and apply them to calculate neighbor diversity. We then use these various neighbor diversities for fraudster detection. Experimental results on a real-world dataset demonstrate that, although these diversities were calculated differently, their performances on fraudster detection are similar. This finding reflects the robustness of neighbor diversity, regardless of how the diversity is calculated.

**Keyword:** online auction, fraudster detection, neighbor diversity, entropy

## References

- [1] D. Houser and J. Wooders, "Reputation in Auctions: Theory, and Evidence from eBay," *Journal of Economics & Management Strategy*, vol. 15, pp. 353-369, 2006.
- [2] J. C. Wang and C. Q. Chiu, "Detecting online auction inflated-reputation behaviors using social network analysis," *Proc. North American Association for Computational Social and Organizational Science (NAACSOS 05)*, June 2005, pp. 26-28.
- [3] D. Chau, S. Pandit, and C. Faloutsos, "Detecting Fraudulent Personalities in Networks of Online Auctioneers," *Knowledge Discovery in Databases: PKDD 2006, LNAI 4213*, pp. 103-114, 2006.
- [4] S. Pandit, D. H. Chau, S. Wang, and C. Faloutsos, "Netprobe: a fast and scalable system for fraud detection in online auction networks," *Proceedings of the Sixteenth International Conference on World Wide Web (WWW 07)*, May 2007, pp. 201-210.
- [5] J. C. Wang and C. C. Chiu, "Recommending trusted online auction sellers using social network analysis," *Expert Systems with Applications*, vol. 34, pp. 1666-1679, April 2008.
- [6] C. C. Chiu, Y. C. Ku, T. Lie, and Y. C. Chen, "Internet Auction Fraud Detection Using Social Network Analysis and Classification Tree Approaches," *International Journal of Electronic Commerce*, vol. 15, pp. 123-147, Spr. 2011.
- [7] J. L. Lin and L. Khomnotai, "Using Neighbor Diversity to Detect Fraudsters in Online Auctions," *Entropy*, vol. 16, pp. 2629-2641, May 2014.
- [8] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.
- [9] J. L. Lin, "On the Diversity Constraints for Portfolio Optimization," *Entropy*, vol. 15, pp. 4607-4621, October 2013.
- [10] D. H. Chau and C. Faloutsos, "Fraud detection in electronic auction," *European Web Mining Forum at ECML/PKDD*, pp. 87-97, October 2005.

\* Corresponding author. E-mail address: k.laksamee@hotmail.com

Tel.: +66-86-6507843, +66-44-354936; Fax: +66-44-354936

- [11] Z. Bin, Z. Yi, and C. Faloutsos, "Toward a Comprehensive Model in Internet Auction Fraud Detection," Proceedings of the 41st Annual Hawaii International Conference on System Sciences, IEEE Press, Jan. 2008, p. 79.
- [12] W. You, L. Liu, M. Xia, and C. Lv, "Reputation Inflation Detection in a Chinese C2C Market," Electronic Commerce Research and Applications, vol. 10, pp. 510-519, September-October 2011.
- [13] W. H. Chang and J. S. Chang, "An Effective Early Fraud Detection Method for Online auctions," Electronic Commerce Research and Applications, vol. 11, pp. 346-360, July-August 2012.
- [14] M. Morzy, "New Algorithms for Mining the Reputation of Participants of Online Auctions," Algorithmica, vol. 52, pp. 95-112, 2005.
- [15] M. Morzy, "Cluster-Based Analysis and Recommendation of Sellers in Online Auction," Computer Systems Science and Engineering, vol. 22, pp. 279-287, 2007.
- [16] S. J. Lin, Y. Y. Jheng, and C. H. Yu, "Combining Ranking Concept and Social Network Analysis to Detect Collusive Groups in Online Auctions," Expert Systems with Applications, vol. 39, pp. 9079-9086, 2012.
- [17] C. H. Yu and S. J. Lin, "Web Crawling and Filtering for On-line Auctions from a Social Network Perspective," Information Systems and e-Business Management, vol. 10, pp. 201-218, 2012.
- [18] V. Batagelj and M. Zaveršnik, "Fast Algorithms for Determining (Generalized) Core Groups in Social Networks," Advances in Data Analysis and Classification, vol. 5, pp. 129-145, July 2011.
- [19] V. Batagelj and M. Zaveršnik, "An  $O(m)$  Algorithm for Cores Decomposition of Networks," arXiv preprint cs/0310049, pp. 1-10, October 2003.
- [20] V. Batagelj and A. Mrvar, "Pajek— Analysis and Visualization of Large Networks," Graph Drawing, vol. 2265, pp. 477-478, 2002.
- [21] M. O. Hill, "Diversity and Evenness: A Unifying Notation and Its Consequences," Ecology, vol. 54, pp. 427-432, 1973.
- [22] H. Tuomisto, "A Consistent Terminology for Quantifying Species Diversity? Yes, It Does Exist," Oecologia, vol. 164, pp. 853-860, December 2010.
- [23] L. Jost, "Entropy and Diversity," Oikos, vol. 113, pp. 363-375, 2006.
- [24] H. Tuomisto, "A Diversity of Beta Diversities: Straightening Up a Concept Gone Awry. Part 1. Defining Beta Diversity as a Function of Alpha and Gamma Diversity," Ecography, vol. 33, pp. 2-22, 2010.
- [25] I. Usta and Y. M. Kantar, "Mean-Variance-Skewness-Entropy Measures: A Multi-Objective Approach for Portfolio Selection," Entropy, vol. 13, pp. 117-133, 2011.
- [26] X. X. Huang, "An entropy method for diversified fuzzy portfolio selection," International Journal of Fuzzy Systems, vol. 14, pp. 160-165, March 2012.
- [27] A. E. Magurran, Ecological diversity and its measurement, United Kingdom: Croom Helm Ltd, 1988.
- [28] P. A. Frost and J. E. Savarino, "For better performance: constrain portfolio weights," The Journal of Portfolio Management, vol. 15, pp. 29-34, 1988.
- [29] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," Management Science, vol. 55, pp. 798-812, May 2009.
- [30] R. Jagannathan and T. Ma, "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," The Journal of Finance, vol. 58, pp. 1651-1684, 2003.
- [31] C. H. Yu and S. J. Lin, "Fuzzy rule optimization for online auction frauds detection based on genetic algorithm," Electronic Commerce Research, vol. 13, pp. 169-182, May 2013.
- [32] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed, California: Morgan Kaufmann Publishers, 2005.

## List of Symbols and Nomenclature

SNA	Social network analysis
J48	Classification tree J48 algorithm
NN	Neural network algorithm
SVM	Support vector machine
CW	Center weight
$D_s^r$	The neighbor diversity on the number of received ratings based on Shannon entropy
$D_{max}^r$	The neighbor diversity on the number of received ratings based on max weight diversity
$D_{min}^r$	The neighbor diversity on the number of received ratings based on min weight diversity
$D_2^r$	The neighbor diversity on the number of received ratings based on canonical $L^p$ -norm diversity with the value of $pow=2$
$D_3^r$	The neighbor diversity on the number of received ratings based on canonical $L^p$ -norm diversity with the value of $pow=3$

- $D_{CS}^r$  The neighbor diversity on the number of received ratings based on canonical Shannon entropy
- $D_S^c$  The neighbor diversity on the number of cancelled transactions based on Shannon entropy
- $D_{max}^c$  The neighbor diversity on the number of cancelled transactions based on max weight diversity
- $D_{min}^c$  The neighbor diversity on the number of cancelled transactions based on min weight diversity
- $D_2^c$  The neighbor diversity on the number of cancelled transactions based on canonical  $L^p$ -norm diversity with the value of  $pow=2$
- $D_3^c$  The neighbor diversity on the number of cancelled transactions based on canonical  $L^p$ -norm diversity with the value of  $pow=3$
- $D_{CS}^c$  The neighbor diversity on the number of cancelled transactions based on canonical Shannon entropy

