USING DATA MINING TO SUPPORT RESEARCH ON FRAUD

VINICIUS ALMENDRA[1]

Fraud is a recurring phenomenon at online actions sites like eBay. The enormous amount of transaction data publically available offers a good opportunity for fraud prevention based on learning methods. However, online auction sites usually neither confirm nor deny fraudulent behavior: they simply suspend seller accounts and publicize feedback information supplied by buyers. While some cases receive some media attention, most of them are hidden in the site's database. This limits the possibility of developing and testing new learning methods for fraud prevention, due to the scarcity of fraud samples. In order to overcome this limitation, we designed a data mining system to extract live information from online auction sites and incrementally build a dataset useful for fraud research, comprising information about sellers, products and feedbacks. The system uses supervised learning to recognize in the textual comments evidences of non-delivery fraud, which are valuable information for fraud elicitation. We implemented a prototype of the system and evaluated it using data extracted from a major online auction site.

Keywords: fraud elicitation, online auction sites, multinomial kernel, Support Vector Machine, cross-validation, bootstrap interval, data mining.

[1] Faculty of Mathematics and Informatics, University of Bucharest, str. Academiei 14, 010014 Bucharest, Romania (vinicius.almendra@gmail.com)