Postprint: Fraudulent Transaction Detection Using Deep Belief Networks and Fuzzy Sets

Authors: Zhang Liyi, Liu Chang

Date: 2017-10-11T00:00:00+00:00

Abstract

Purpose: To address the problem of fake transactions in e-commerce platforms.

Method: Based on consumers' historical purchase and review behavior data, this study proposes a fake transaction identification method that combines deep belief networks and fuzzy sets, which identifies fake transactions by recognizing users who conduct fake transactions (brushing participants).

Results: The identification accuracy reaches 89%, and compared with experimental results of shallow machine learning models, its overall performance demonstrates significant improvement.

Limitations: The experimental dataset is relatively small compared to the massive number of brushing participants on Taobao. Only Taobao data is used as validation data, without involving other e-commerce platforms.

Conclusion: This method can effectively identify brushing participants and mitigate the problem of fake transactions in e-commerce.

Full Text

Combining Deep Belief Networks and Fuzzy Sets for Fraud Transaction Recognition

Zhang Liyi, Liu Chang
(School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract

[**Objective**] This study addresses the pervasive problem of fraud transactions in e-commerce platforms. [**Methods**] We propose a novel fraud detection method that integrates deep belief networks (DBN) and fuzzy set theory, leveraging consumers' historical purchase and review behavior data to identify fraudulent transaction participants (known as "brushing cheaters"). By detecting these cheaters, we can effectively identify fraudulent transactions themselves. [**Results**] Experimental validation using data from Taobao.com demonstrates that the method achieves 89% accuracy, showing significant performance improvements compared to shallow machine learning models. [**Limitations**] The experimental dataset is relatively small compared to the vast number of cheaters on Taobao, and validation is limited to Taobao data without extending to other e-commerce platforms. **Conclusion** The proposed method effectively identifies brushing cheaters and can help reduce fraud transactions in e-commerce.

Keywords: Fraud transaction; Cheater recognition; Product reviews; Deep learning; Fuzzy set
Classification: G202

Introduction

E-commerce in China has experienced rapid development with continuously expanding transaction volumes. According to statistics from the China E-Commerce Research Center [1], by the end of 2014, China's online retail transaction scale reached 2.82 trillion RMB, surpassing the United States to become the world' s largest online retail market. This enormous market attracts increasing participation. As reported by CCIDNet [2], by the end of 2013, the number of Taobao stores reached 9 million, with daily online product listings exceeding 800 million. Such vast product and merchant quantities entail fierce competition. To secure top search rankings and attract consumer attention and purchases, fraudulent practices involving fake transactions to boost store credibility and product sales have emerged. In China, transaction fraud for brushing purposes has formed a massive profit chain, with specialized companies providing brushing services to sellers. To increase sales volume and enhance merchant/product reputation, cheaters must provide positive reviews, which significantly influence purchasing decisions. Moreover, the volume of product reviews determines how long users stay on product detail pages. When shopping on Taobao, consumers primarily rely on product sales volume and reviews for decision-making. However, fake sales and reviews severely mislead consumers and damage their interests. Therefore, identifying fraudulent transactions and fake reviews is crucial for the healthy development of e-commerce.

Fake reviews constitute a critical component of fraudulent transactions, defined as positive reviews given without factual basis to promote product sales, or negative reviews to damage product reputation. Current research on fake review identification follows two main directions: (1) identifying fake reviews directly from the review content itself, and (2) identifying fake reviews by detecting fake review posters. Common detection methods include supervised learning algorithms such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). While these achieve decent classification performance, they require large labeled datasets for training, incurring substantial manual annotation costs, and feature selection significantly impacts results.

Related Work

In content-based fake review detection, Jindal et al. [3-4] categorized spam reviews into three types: fake reviews, irrelevant reviews, and non-review information, proposing methods to identify suspicious reviews through unexpected rule detection and rule clustering. Ott et al. [5] employed standard word and part-of-speech N-gram features for supervised learning on fake reviews from Amazon Mechanical Turk and genuine reviews from TripAdvisor.com, using SVM for classification. Ren et al. [6] argued that fake and genuine reviews differ in linguistic structure and sentiment polarity, proposing a genetic algorithm-based approach to optimize feature selection for linguistic structure and sentiment polarity, then applying hard and soft clustering algorithms for detection. Feng et al. [7] utilized probabilistic context-free syntactic stylometric rules for fake review identification, employing SVM classifiers on standard datasets.

Beyond review content analysis, researchers have also exploited reviewer behavioral features. Fei et al. [8] identified fake reviewers by analyzing review bursts, arguing that reviews appearing within the same burst period share common characteristics—either all from fake reviewers or all from genuine ones—enabling cheater inference. Lim et al. [9] detected fake reviewers through anomalous rating behaviors, where continuous extreme ratings (either excessively high or low) increase the likelihood of being a fake reviewer. Jiang et al. [10] summarized two behavioral patterns of spam reviewers: continuous reviews of a specific product within short timeframes, and significant discrepancies between actual purchase volume and positive reviews. They identified spam reviewers through behavioral analysis and rating deviation assessment.

These studies primarily analyze all review information about reviewed subjects rather than examining whether a user is a fake reviewer based on their entire historical review data on a platform. Furthermore, previous research employed shallow machine learning models like SVM and KNN—supervised models requiring extensive labeled samples and significant annotation time costs. Shallow models rely heavily on manually engineered features, making feature quality the performance bottleneck when models operate correctly [11]. Unlike shallow learning, deep learning transforms sample features from the original space to a new feature space through layer-wise transformation, making classification or prediction easier and demonstrating powerful capability to learn dataset essential features from limited samples [12]. Deep learning research has primarily focused on speech recognition [13], natural language processing [14], and image processing [15].

Deep Belief Networks (DBN) represent a deep neural network comprising multiple layers of unsupervised Restricted Boltzmann Machines (RBM) and a supervised Backpropagation (BP) network, constituting a machine learning model in deep learning [16]. Compared to earlier supervised deep learning models like Convolutional Neural Networks (CNNs) and traditional supervised shallow models, DBN as a semi-supervised deep learning model can leverage large-scale unlabeled samples for training, saving annotation time while its unsupervised learning process discovers more accurate features and overcomes local optima limitations. As a deep architecture, DBN can learn abstract features, weaken erroneous features from shallow layers, improve classification performance, and mitigate overfitting. While CNN [17] is specially designed for 2D image recognition with advantages in image processing, it is unsuitable for this scenario.

Deep learning simulates human brain mechanisms for data interpretation, with its hierarchical theory grounded in neuroscience: lower layers process primary inputs, feeding results to higher layers that learn advanced features. In Zeki' s research [18], knowledge originates from either innate inheritance or acquired learning. While inherited knowledge is immutable, the precursors and assumptions of acquired knowledge are useful to humans and can be refined through experience and unconscious thinking. This motivates using fuzzy set theory to simulate inherited knowledge and DBN to simulate acquired knowledge. Fuzzy set theory describes and processes uncertain phenomena mathematically, transforming traditional binary logic into continuous logic on the [0,1] interval, widely applicable in pattern recognition [19]. In pattern recognition, fuzzy sets can describe the degree to which an object belongs to a category [20]. For cheater identification, we transform the binary logic of "is a cheater" or "is not a cheater" into membership degrees. Introducing fuzzy set concepts into deep learning can effectively improve prediction accuracy. Fu et al. [21] incorporated fuzzy set theory in Chinese sentence-level sentiment classification, modeling the inherent fuzziness of sentiment polarity classification directly. We propose a fraud transaction identification method combining DBN and fuzzy sets, comparing its performance with shallow machine learning models KNN and SVM.

Feature Engineering

To identify cheaters, we distinguished between fraudulent and normal users from Taobao review data and extracted features from their three-month review and purchase histories. Taobao employs similar credibility metrics for sellers and buyers: seller credit scores and buyer credit scores. However, these simple metrics cannot accurately differentiate normal buyers from cheaters, necessitating additional derived attributes to characterize user behaviors.

Whitrow et al. [22] noted that different statistical time periods critically impact model performance: overly short periods fail to capture sufficient consumer history, while overly long periods introduce excessive noise and may hide identifiable features. Therefore, appropriate time windows significantly affect cheater identification.

F1: Registration Duration
Users with shorter registration times are more likely to be cheaters. Since cheaters are also consumers who shop on Taobao, they often register separate accounts for brushing activities to avoid penalties affecting their normal consumption, thereby reducing punishment costs. We use the time interval between

user registration and their last collected review as a feature.

F2: Real-Name Authentication Status
Unauthenticated users are more likely to be cheaters. Real-name authentication better protects user funds and interests during disputes, so normal consumers typically complete it. Cheaters, not aiming for actual consumption and seeking to protect personal information, often remain unauthenticated, which also helps conceal their identity if detected. We use 0 for unauthenticated and 1 for authenticated users as a feature.

F3: Total Product Categories Purchased
Users purchasing more product categories are more likely to be cheaters. Cheaters buy products as requested by clients rather than based on genuine needs, resulting in higher category counts than normal users within a given period. We count total purchased product categories up to the last collected review time.

F4: Daily Product Categories Purchased
We calculate the average number of product categories purchased per buying day in the past month up to the last review time.

F5: Review Length
Users with large variance in review lengths may be cheaters. We use the ratio of total review characters to total review count for each user.

F6: Daily Review Count
Users with more daily reviews are more likely to be cheaters. To assist merchants in deceiving consumers and make fake transactions appear authentic, cheaters review every completed fake transaction, yielding higher average daily review counts than normal users. We calculate the average daily review count in the past month up to the last review time.

F7: Monthly Review Count
We calculate the average monthly review count over the past three months up to the last review time.

F8: Duplicate Review Rate
Higher duplicate review rates indicate cheaters. Cheaters use comments provided by brushing agencies and are compensated based on completed fake transaction volume, resulting in higher duplicate rates than normal users. We compute the ratio of duplicate reviews to total reviews up to the last review time.

F9: Content Review Rate
Higher content review rates suggest cheaters. Reviewing is mandatory in fake transactions, so cheaters' content review rates typically exceed normal users. We calculate the ratio of total reviews to reviewer credit points (excluding anonymous and system-default positive reviews) up to the last review time.

F10: Repeated Merchant Rate
Users frequently purchasing from the same merchant are more likely to be

cheaters. New merchants hire cheaters for multiple fake transactions to boost credibility, resulting in high merchant repetition in cheaters' purchase histories. We compute the ratio of repeated merchants to total merchants up to the last review time.

F11: Consumer Credit Score Daily Growth Rate
Faster credit score growth indicates cheaters. Buyer credit scores accumulate per order item, with different ratings contributing varying points. Cheaters, profiting from brushing, conduct numerous fake transactions, yielding higher credit score growth than normal consumers. We use average daily credit score (credit score divided by registration days) as a feature.

F12: Consumer Credit Score Monthly Growth Rate
We use average monthly credit score (credit score divided by total purchase months) as a feature.

[Figure 1: see original paper] shows statistical descriptions of the dataset, revealing differences between fraudulent and normal transaction behaviors through consumer purchase features. For the derived feature "daily review count,"fraudulent users' mean daily review count is 5.764382, approximately 2.45 times higher than normal users' mean of 2.350858. Similar quantitative differences appear across other behavioral features.

[Figure 2: see original paper] presents independent samples t-test results for consumer behavioral features. For "registration days," Levene' s test shows F=0.009 and Sig.=0.925, indicating no significant variance difference, so we refer to the first row of t-test results where Sig.=0.000, confirming significant mean differences. Other features show similarly significant differences.

Methodology: DBN and Fuzzy Sets

Deep Belief Networks (DBN) are widely studied and applied deep learning architectures composed of Restricted Boltzmann Machine (RBM) units. Fuzzy sets mathematically describe uncertainty, and introducing fuzzy set concepts into deep learning can effectively improve prediction accuracy.

Let x_i denote a user from user set X, where $X = [x_1, \dots, x_{R+T}]$, with each $x = [x_1, \dots, x_D]'$. Here, R represents the number of training users, T the number of test users, and D the number of user features. Let Y denote labels for L labeled training samples: $Y = [y_1, \dots, y_L]$, where each $y = [y_1, \dots, y_c]'$ and c represents the number of classes.

DBN training follows a semi-supervised greedy learning algorithm with two main phases:

1. **Layer-wise Unsupervised RBM Training**: Train each RBM layer separately using unlabeled data. The first RBM network comprises input layer h_0 and first hidden layer h_1, receiving raw feature vectors and learning parameters w_1 between layers. After training the first RBM,

its hidden layer h_1 becomes the visible layer for the second RBM, which trains parameters w_2 with h_2 as the new hidden layer. This process continues iteratively: the output of layer $N-1$ becomes the input for layer N, training parameters w_N, thereby initializing the DBN parameter space $W = [w_1, \ldots, w_N]$.

2. **Supervised Fine-tuning**: After layer-wise pre-training, use a BP network for supervised fine-tuning. Error between input features and reconstructed features from the top-layer representation is backpropagated through all RBM layers to adjust inter-layer parameters, yielding optimal DBN parameters [12,23].

For cheater identification, fuzzy sets describe membership degrees of "is a cheater" or "is not a cheater." Let fuzzy sets A (positive) and B (negative) be defined as follows [24]:

Let X be the user set with elements x representing individual users. Positive fuzzy set A is represented by membership function $\mu_A(x)$, where $\mu_A(x) \in [0,1]$ indicates the degree to which x belongs to cheaters. Negative fuzzy set B uses membership function $\mu_B(x) \in [0,1]$ indicating the degree to which x does not belong to cheaters. Both membership functions are computed from the deep architecture' s layer N output $h_N(x)$.

For final cheater identification with two classes ("is cheater" vs. "is not cheater"), layer N dimension should be 2, with class boundary at $h_N = 0$. The distance between $h_N(x_i)$ and this boundary is $d(x_i) = (h_N(x_i)_1 - h_N(x_i)_2)/2$. If $d(x_i) > 0$, x_i is a cheater; otherwise, it is not.

The relationship between membership functions $\mu_A(x)$, $\mu_B(x)$ and distance $d(x)$ is given by [25]:

$$S(d(x); \beta, \gamma) = \begin{cases} 1 & \text{if } d(x) > \gamma \\ \frac{d(x)-\beta}{\gamma-\beta} & \text{if } \beta < d(x) \le \gamma \\ 0 & \text{if } d(x) \le \beta \end{cases}$$

During identification, parameters β and γ must be estimated. In [Figure 3: see original paper] with $\beta = 2$ and $\gamma = 1$, γ represents the boundary where $\mu_A(x) < 1$ and $\mu_A(x) = 1$ (and $-\gamma$ for $\mu_B(x)$). β represents the distance $d(x)$ over which $\mu_A(x)$ transitions from 0 to 1 (and similarly for $\mu_B(x)$). We estimate β and γ by statistically analyzing all users' distance values $d(x)$.

The DBN-based cheater identification is described by:

$$\beta = \xi \times \max |d(x_i)|, \quad i = 1, \ldots, R+T$$

where $\xi \ge 1$ is a constant representing the boundary degree between "is cheater" and "is not cheater," adjustable based on specific datasets.

[Figure 3: see original paper] illustrates the membership functions $\mu_A(x)$ and $\mu_B(x)$. Using equations (3) and (4) to estimate fuzzy parameters, we construct the deep architecture. The L labeled samples and membership functions $\mu_A(x)$ and $\mu_B(x)$ then optimize parameter space W to improve discrimination accuracy. [Figure 4: see original paper] depicts layer $N-1$ incorporating fuzzy set concepts, using membership functions as input functions.

Experiments and Results

To validate our method' s performance, we collected Taobao user purchase data including gender, registration days, credit score, real-name authentication status, purchased product names, review content, ratings, and review timestamps. To obtain verified cheater data for training, we monitored major brushing platforms (Shuangying Network, Baili Network, ShuaKe Network) that profit by publishing merchant orders and task requirements. We identified brushing users from product reviews in participating stores based on task requirements, then collected their basic information and historical reviews using the Taobao query website TaoDake. For normal user data, we selected reputable, high-influence Tmall stores (e.g., Nike flagship store, Xiaomi flagship store) that don' t require brushing. From these stores' hot product reviews, we selected non-anonymous reviewers and verified their historical review patterns via TaoDake. Users with objective review content, few duplicates, and no short-term massive reviewing behavior were classified as normal.

In fuzzy sets, parameter ξ represents the boundary degree between "is cheater" and "is not cheater," significantly affecting accuracy. We tested different ξ values to find the optimum, with results shown in [Figure 5: see original paper].

[Figure 5: see original paper] demonstrates that identification accuracy peaks at 89% when $\xi = 3$, which we adopted for this study.

[Figure 6: see original paper] shows classification results where 1 represents cheaters and -1 represents normal users. Using 100 users as the test set, overlapping points indicate correct identifications while non-overlapping points indicate errors. The first row of non-overlapping solid points shows normal users misidentified as cheaters; the second row shows cheaters misidentified as normal users. The results show 11 misidentified users, achieving 89% accuracy.

We evaluate performance using standard classifier metrics: Accuracy, Precision, Recall, and F-score, combining Precision and Recall via F-score [26]:

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |FP| + |TN| + |FN|}$$

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|}$$

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|}$$

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $|TP|$ is true positives (cheaters correctly identified), $|FP|$ is false positives (normal users misidentified as cheaters), $|TN|$ is true negatives (normal users correctly identified), and $|FN|$ is false negatives (cheaters misidentified as normal users).

compares our method with KNN and SVM. While Precision is slightly lower, F-score is significantly higher, indicating clear performance improvement.

** Performance Comparison of DBN with SVM and KNN**

Method	Accuracy	Precision	Recall	F-score
DBN (Proposed)	89%	84.21%	96%	89.72%
SVM	-	85.42%	75.56%	83.68%
KNN	-	-	-	-

Our method combines deep architecture feature extraction with fuzzy classification, using exponential loss functions during training to maximize class separability. Employing the same deep architecture for both fuzzy parameter estimation and cheater classification enhances process consistency and improves performance, demonstrating that the DBN-fuzzy set combination strengthens discriminative capability.

Conclusion

This paper proposes a fraud transaction identification method combining Deep Belief Networks and fuzzy sets. By analyzing behavioral features of fraudulent transaction participants (cheaters), we identify them from massive user populations and flag their transactions as fraudulent. We extracted 12 quantifiable features from users' historical review and transaction records, constructed a DBN-fuzzy set hybrid architecture, and defined fuzzy sets for cheater identification to enhance recognition capability through fuzzy information training. Experiments using Taobao data demonstrate that our method achieves 89% accuracy, 84.21% precision, 96% recall, and 89.72% F-score, significantly outperforming existing classification methods and effectively achieving fraud transaction detection.

References

[1] China E-Business Research Center. The 2014 Report of China E-Business Market Data Monitoring [R/OL]. [2015-04-08]. http://www.100ec.cn/zt/upload_{data}/20150408.pdf.

[2] "2014 Most Successful E-Commerce Website" Nomination: Taobao.com [EB/OL]. [2014-12-05]. http://miit.ccidnet.com/art/32559/20141205/5693755_1.html.

[3] Jindal N, Liu B. Opinion Spam and Analysis [C]. In: Proceedings of the 2008 International Conference on Web Search and Web Data Mining (WSDM). 2008.

[4] Jindal N, Liu B, Lim E P. Finding Unusual Review Patterns Using Unexpected Rules [C]. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM). 2010: 1549-1552.

[5] Ott M, Choi Y, Cardie C, et al. Finding Deceptive Opinion Spam by Any Stretch of the Imagination [C]. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011: 309-319.

[6] Ren Yafeng, Yin Lan, Ji Donghong. Deceptive Reviews Detection Based on Language Structure and Sentiment Polarity [J]. Journal of Frontiers of Computer Science & Technology, 2014, 8(3): 313-320.

[7] Feng S, Banerjee R, Choi Y. Syntactic Stylometry for Deception Detection [C]. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers. 2012: 171-175.

[8] Fei G, Mukherjee A, Liu B, et al. Exploiting Burstiness in Reviews for Review Spammer Detection [C]. In: Proceedings of the 7th International AAAI Conference on Weblogs and Social Media. 2013, 13: 175-184.

[9] Lim E P, Nguyen V A, Jindal N, et al. Detecting Product Review Spammers Using Rating Behaviors [C]. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. ACM, 2010.

[10] Jiang B, Cao R H, Chen B. Detecting Product Review Spammers Using Activity Model [C]. In: Proceedings of the 2013 International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013). Atlantis Press, 2013: 650-653.

[11] Yu Kai, Jia Lei, Chen Yuqiang, et al. Deep Learning: Yesterday, Today, and Tomorrow [J]. Journal of Computer Research and Development, 2015, 50(9): 1799-1804.

[12] Sun Zhijun, Xue Lei, Xu Yangming, et al. Overview of Deep Learning [J]. Application Research of Computers, 2012, 29(8): 2806-2810.

[13] Dahl G E, Yu D, Deng L, et al. Context-Dependent Pre-trained Deep Neural Networks for Large-Vocabulary Speech Recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30-42.

[14] Collobert R, Weston J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning [C]. In: Proceedings of the 25th International Conference on Machine Learning. ACM, 2008: 160-167.

[15] Krizhevsky A, Sutskever I, Hinton G E. Imagenet Classification with Deep Convolutional Neural Networks [C]. In: Proceedings of the 26th Annual Conference on Neural Information Processing Systems. 2012: 1097-1105.

[16] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets [J]. Neural Computation, 2006, 18(7): 1527-1554.

[17] Li Baoqing. Building Pattern Classifiers with Convolutional Neural Networks [J]. Journal of Dalian University, 2003, 24(2): 19-23.

[18] Zeki S. Splendors and Miseries of the Brain: Love, Creativity, and the Quest for Human Happiness [M]. The 2nd Edition. John Wiley & Sons, 2011.

[19] Mendel J M. On a Novel Way of Processing Data that Uses Fuzzy Sets for Later Use in Rule-based Regression and Pattern Classification [J]. International Journal of Fuzzy Logic and Intelligent Systems, 2014, 14(1): 1-7.

[20] Simpson P K. Fuzzy Min-Max Neural Networks—Classification [J]. IEEE Transactions on Neural Networks, 1992, 3(5): 776-786.

[21] Fu G, Wang X. Chinese Sentence-level Sentiment Classification Based on Fuzzy Sets [C]. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010: 312-319.

[22] Whitrow C, Hand D J, Juszczak P, et al. Transaction Aggregation as a Strategy for Credit Card Fraud Detection [J]. Data Mining and Knowledge Discovery, 2009, 18(1): 30-55.

[23] Leng B, Zhang X, Yao M, et al. A 3D Model Recognition Mechanism Based on Deep Boltzmann Machines [J]. Neurocomputing, 2015, 151: 593-602.

[24] Rutkowska P D. Neuro-fuzzy Architectures and Hybrid Learning [M]. Physica-Verlag HD, 2012.

[25] Zimmermann H J. Fuzzy Set Theory—And Its Applications [M]. Springer Netherlands, 2001.

[26] Wang X Y, Yang H Y, Li D M. A New Content-based Image Retrieval Technique Using Color and Texture Information [J]. Computers & Electrical Engineering, 2013, 39(3): 746-761.

Author Contributions

Zhang Liyi: Conceived research ideas and designed the study. **Liu Chang:** Conducted experiments, collected and analyzed data, drafted the manuscript. **Zhang Liyi and Liu Chang:** Revised the final manuscript.

Received: June 26, 2015
Revised: October 14, 2015

Note: Figure translations are in progress. See original paper for figures.