

Crowdsourcing Fraud Detection Algorithm Based on Ebbinghaus Forgetting Curve

Li Peng^{1,2}, Yu Xiao-yang¹, Liu Yang² and Zhang Ting-ting²

¹ Higher Educational Key Laboratory for Measuring and Control Technology, Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, 150080 Harbin, China

² School of Computer Science and Technology, Harbin University of Science and Technology, 150080 Harbin, China
{pli, yuxiaoyang}@hrbust.edu.cn.

Abstract

In the practical application of crowdsourcing, some unreliable workers have emerged due to profit driven. Their results seriously reduce the quality and bring about the initiator's judgment biases. In this paper, we creatively put forward a crowdsourcing fraud detection method based on psychological behavior analysis to find out the spammer according to the psychological difference between deception and reliable behavior by means of Ebbinghaus forgetting curve. Furthermore, we constructed an online crowdsourcing experiment platform to verify the validity of our method. As a result, filtration recall and precision attained 0.93 and 0.79 respectively. Moreover, the overall LAM was decreased by 3.7% and AUC increased by 8% before and after filtration. Experimental results show that our method is effective and feasible.

Keywords: Crowdsourcing, Ebbinghaus Forgetting Curve; Fraud Detection

1. Introduction

Crowdsourcing, a new organization form and cooperation pattern in the process of enterprise production, has grown with the rapid popularity of Internet [1]. Enterprises actively utilize many online user resources and allocate the outsourcing task to the interest groups by means of crowdsourcing technology, solving some limitations of traditional outsourcing services. In recent years, crowdsourcing technology has become a focus of research and researchers around the world have realized it in some practical applications. For example, Rensnik combined monolingual crowdsourcing and targeted paraphrasing to improve the quality of pure machine translation [2]; Hwang introduced mobile-based crowdsourcing into the field of environmental audio recognition, to improve the performance of mobile devices in filtering background noise [3]; Fritz applied crowdsourcing to global land cover, to solve the problem of ignoring potential land in statistical work [4]. Moreover, crowdsourcing technology is also applied to other fields such as software testing, content screening, and labeling training data for machine learning [5]. The rapid development of crowdsourcing technology drew the Text REtrieval Conference's (TREC) attention and crowdsourcing track has attracted many groups around the world to join the competition in TREC 2013.

Unfortunately, some unreliable workers have emerged due to profit driven in the practical application of crowdsourcing. Their results seriously reduce the quality and bring about the initiator's judgment biases [6]. In recent years, many researchers conducted in-depth

exploration and have achieved some progress on improving the quality of crowdsourcing results. For instance, Matthew from the UT Austin provides a method for detecting cheats and it think that worker should pushed confidence score button according to the confidence degree of answers submitted [7]. Bell labs researchers noticed that a large number of user information can be obtained from mobile devices and this information contributes to finding better workers [8]. Jeroen *et al.*, use the average squared ordinal distance between workers' judgments to calculate each worker's random score for detecting random spammers [9]. At present, crowdsourcing technology is still in the early stages and it is important for us to improve the quality of crowdsourcing results by some means.

Detecting spammers is important for enhancing the quality of crowdsourcing results and workers' psychological state will change when they are cheating. The essence of deception is a manifestation of people's psychological activity and psychological methods can make effective judgment on it in a certain extent. Therefore, we think that using the psychological method may be a realistic way to solve this problem and put forward a kind of crowdsourcing fraud detection method based on psychological behavior analysis. The method applying Ebbinghaus forgetting curve to determine the behavior of crowdsourcing worker is normal or not and find out the difference between deception and reliable behavior.

2. Behavior Analysis based on Psychological Behavior Analysis

Ebbinghaus forgetting curve is one of the classical methods in psychology field, it reveals the regularity of forgetting information memorized. In recent years, some researchers have introduce this method into the field of computer science, such as Zeng and Lin designed an interactive vocabulary learning system based on word frequency lists and Ebbinghaus' curve of forgetting [10]; Luo and Yuan utilize Ebbinghaus forgetting curve to mine Internet users' term interest and provide personalized search results [11]. In this article, Ebbinghaus forgetting curve is applied in crowdsourcing fraud detection. The principle is the differences between deceptional worker and trusted worker in forgetting curve when they do crowdsourcing tasks.

2.1. Ebbinghaus Forgetting Curve

Memory can be divided into short-term and long-term according to the length of maintaining information periods. The memory process is shown as Figure 1. Received information will become short-term memory after attention. The memory will be forgotten without timely review. Otherwise, long-term memory will form.

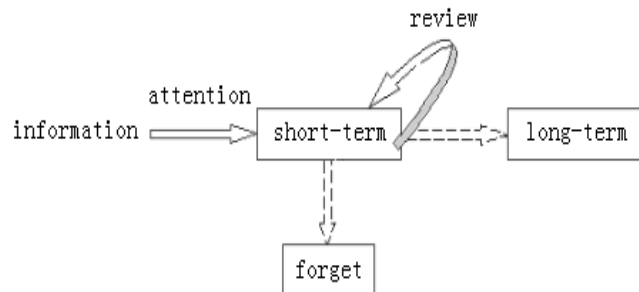


Figure 1. Human Memory Process

The German psychologist H.Ebbinghaus researched the basic rule of human memory and oblivion, and put forth "the function of time and memory" as shown in the formula (1). *OriginalLearning* stands for the number of writing from memory, when he remembered all materials in the first time. After a while, *Relearning* is the number. Thus retention scores are obtained and represented by *SavingScore*.

$$SavingScore = \frac{OriginalLearning - Relearning}{OriginalLearning} \quad (1)$$

According to the formula (1), The Ebbinghaus forgetting curve can be drawn (see Figure 2). In this figure, the longitudinal axis represents the memory retention scores and the horizontal axis represents elapsed time since learning. The curve conducts a quantitative expression for the forgetting rules in learning process resulting oblivion can be calculated. Human oblivion is an unbalanced development, Memory is forgotten very quickly in the initial stage, and then slows down gradually, after a certain time almost no longer forgotten.

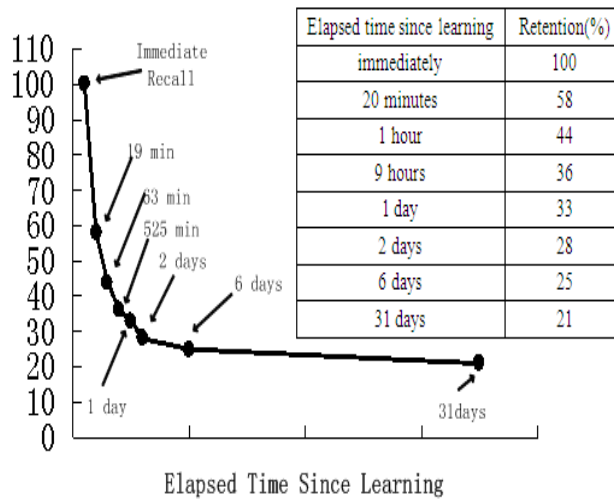


Figure 2. Ebbinghaus Forgetting Curve

2.2. Our Fraud Detection Algorithm

Fraud is usually divided into two basic types. One is random spammer, they submit results randomly and it is difficult to identify them; the other is uniform spammer, they just submit results regularly and this type is easy to be found. Therefore, we only focus on the former.

The principle of our detection method is crowdsourcing participants will produce different memory rules owing to different psychological state in the process of judgment to the crowdsourcing task. Their behavior is a specific embodiment of mental activity, no matter he is a spammer or credible person. The trusted participant will strictly comply with the requirements and think hard when judging the relevance of pairs, so it will produce a deep memory in their mind. This is the general process of human memory in accordance with Ebbinghaus forgetting rule. However, the spammer will only spend little or no effort on crowdsourcing and complete tasks mechanically. They lack the understanding memory about task content and their forgetting states do not comply with Ebbinghaus forgetting rule. According to above psychological differences and quantitative expression of Ebbinghaus forgetting curve, the flow chat of our algorithm is shown in Figure 3. Workers will rejudge a

certain amount of repeated pairs in limited time, and then the work's SavingScore and threshold is compared, low SavingScore workers are considered as the spammer.

There are four sets A, O, W and T in the flow chat, all query-document pairs are waiting for labeling in A. Putting the pairs which are judged for one time into O. The pairs are waiting for labeling again in W. Putting the pairs which are judged for two times into T. In crowdsourcing task, platform randomly selects topic-doc pairs and recommends some pairs that have the less number of judgments for all workers. Firstly, query-document pair would be judged and kept into O, then deleted from A. Secondly, if some pair stay in O about 20 min, it will be put into W and removed from O. At this time, W is not empty and platform may offer worker a pair from W. It means that the pair has been judged. Workers should rejudge the pair within the given limit time so that they have to judge it depending on memory barely rather than reflection. At last, the pair which is judged twice will be placed into T and removed from W.

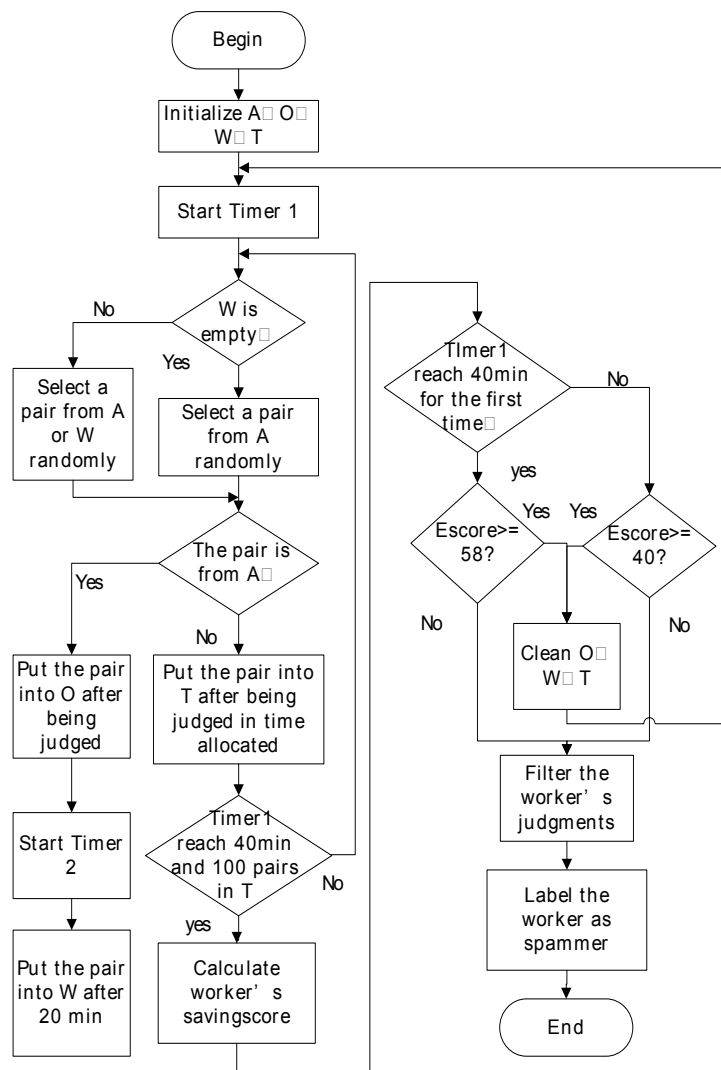


Figure 3. The Flow Chat of our Algorithm

In our algorithm, if workers do repeated judgment for at least 100 query-document pair, the pairs of W reached 100 and the platform will calculate the workers' SavingScore. If the worker operates 40 minutes but haven't reached 80 minutes, and the workers' SavingScore ≥ 58 , we determine the worker has no cheat in this period. If the workers have worked for more than 80 minutes, only his SavingScore ≥ 40 can we think that they didn't have cheat. Platform will adopt judgment results of workers who meet the requirements, then emptying O , W and T and conducting a new round of detecting fraud for the workers.

Finally, we calculate the ratio of the deviations sum of the topic-doc's final label value and its 5 label values, then the product of the maximum deviation of 6 relevance labels and the number of the judgment. At last, 1 minus the ratio gets the relevance probability of final label and judgments.

3. The Results and Analysis of Experiment

3.1. Construction and Evaluation Index of Experiment

We constructed an online crowdsourcing experiment platform to verify the validity of our method. In the Foreign Broadcast Information Service data, a total of 1380 documents were selected from 4 themes as task document collection. All of these documents have the standard relevance labels, and contain 1120 non-relevant documents, 260 relevant documents. The experiment platform randomly assigned these tasks to 76 online students with computer background and let them label pairs.

In order to evaluate the effectiveness of our method filtering spammers, we utilize LAM (Logistic Average Misclassification) and AUC (receiver operating characteristic curve) to evaluate the effectiveness of our method.

$$\text{logit}(p) = \log \frac{p}{1-p} \quad (2)$$

$$\text{logit}^{-1} = \frac{e^x}{1+e^x} \quad (3)$$

$$fpr = \frac{|FP|+0.5}{|FP|+|TN|+1} \quad (4)$$

$$fnr = \frac{|FN|+0.5}{|FN|+|TP|+1} \quad (5)$$

$$LAM = \text{logit}^{-1} \left(\frac{\text{logit}(fpr) + \text{logit}(fnr)}{2} \right) \quad (6)$$

$$AUC = \int_0^1 \frac{TP}{P} d \frac{FP}{N} = \frac{1}{P \cdot N} \int_0^N TP dFP \quad (7)$$

Here, FP (False Positive) represents the number of relevant pairs misjudged by workers; TN (True Negative) represents the number of irrelevant pairs judged by workers; FN (False Negative) represents the number of irrelevant pairs misjudged by workers; TP (True Positive) represents the number of relevant pairs judged by workers. So the number of virtual positive $P = TP + FN$, while the number of negative $N = TN + FP$.

3.2. Experimental Results and Analysis

In the detection of spammers, the system detected 13 deceptional workers. As we can see in Figure 4, six of them were found when they were working for 40 to 80 minutes. During 80 to 120 minutes, only two workers were detected. Our method found 5 workers between 120 and 160 minutes. After then, just one was detected. After investigation on workers, they generally feel tired after working for 120 minute, so the third column is higher than the second one. Moreover, some workers are reluctant to complete the task, so they just submit the task as soon as possible using random strategy. Consequently, their fraud was detected after 40 minutes. Otherwise there are some impatient workers and they often cheating when working for a period of time. The other kind is trusted-worker, they are always careful to complete task. Therefore, only one worker was detected in the later period. In addition, several workers in the experiment don't deliberately label or choose default value on a few documents, so it has little effect on their credibility.

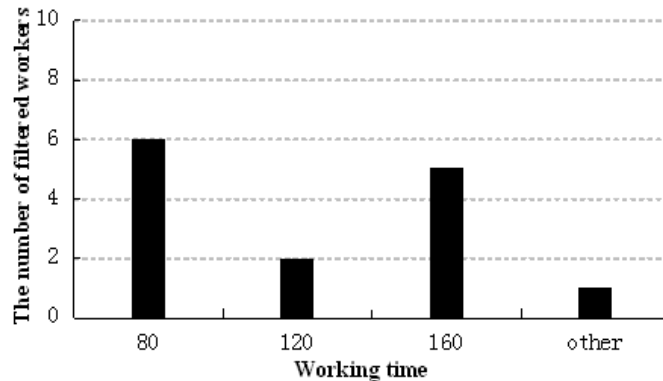


Figure 4. The Detection Results of our Algorithm

In 13 filtered workers, 11 are real spammers and two are trusted workers. Actually, another 3 real spammers are not found by our method. Filtration recall and precision attained 0.93 and 0.79 respectively. After an inquiry on the 3 workers, we know that they labeled most query-document pairs as irrelevance. In fact, the number of irrelevant pairs more than the relevant in our task data set, the ratio is 4.3:1. Therefore, some spammers achieved high accuracy due to the character of data.

Final labels were selected by majority vote before and after filtration, using AUC and LAM assessed 4 topics respectively, and then assessed the whole. The evaluation result is showed in Table 1. As a result, the overall LAM decreased by 3.7% and AUC increased by 8%.

Table 1. Evaluation Results before and after Filtering

Data Set	LAM		AUC	
	Before Filtering	After Filtering	Before Filtering	After Filtering
357	0.19	0.16	0.79	0.86
358	0.25	0.21	0.76	0.82
360	0.14	0.08	0.82	0.91
351	0.17	0.14	0.74	0.87
all	0.19	0.15	0.77	0.85

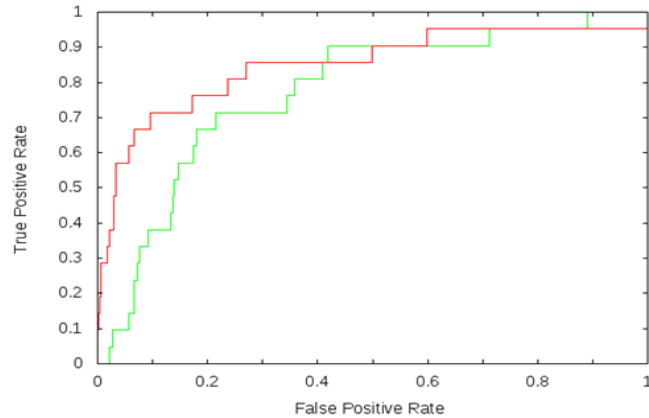


Figure 5. ROC of Crowdsourcing Results

In Figure 5, the green and red ROC is before and after filtering spammers' judgments respectively. Almost the whole green line is under the red one. It means that our method is conducive to improving the quality of crowdsourcing result and can be used to control crowdsourcing quality.

4. Conclusion

This paper proposed an effective solving strategy on crowdsourcing fraud detection by means of psychological behavior analysis method. We creatively apply Ebbinghaus forgetting curve to find out the spammer according to the psychological difference between fraud and reliable behavior. This is an exciting exploration because we successfully applied the psychological method to the field of computer science. We also develop an online crowdsourcing experiment platform to verify the validity of our method, the overall LAM decreased by 3.7% and AUC increased by 8% in test data sets. Experimental results show that our method is contribute to improving the quality of crowdsourcing result and can be used to control crowdsourcing quality.

Acknowledgements

This paper is partially supported by National Natural Science Foundation of China (61103149), China Postdoctoral Science Foundation (2011M500682), Postdoctoral Science Foundation of Heilongjiang Province (LBH-Z11106), Technological Innovation Foundation for Youth Scholars of Harbin (2012RFQXG093), Foundation for University Key Teacher of Heilongjiang Province (1252G023), Postgraduate Innovation Research Project of Heilongjiang Province (YJSCX2012-126HLJ), Research Fund for the Doctoral Program of Higher Education (20102303120005), Province Natural Science Foundation of Heilongjiang (QC2013C060) and Science Funds for the Young Innovative Talents of HUST (NO.201304).

References

- [1] V. Maja, L. Jim, R. Yaoping, H. Milton and R. Sriram, "Assessing Service Development Readiness Using Enterprise Crowdsourcing", Proceeding of the 2013 IFIP/IEEE International Symposium On Intergrated Network Management, Ghent, Belgium, (2013) May 27-31.
- [2] R. Philip, B. Olivia, K. Yakov, H. Chang, A. J. Quinn and B. B. Bederson, "Using Targeted Paraphrasing and Monolingual Crowdsourcing to Improve Translation", Computer Communication Review, vol. 3, no. 4, (2013).

- [3] H. Kyuwoong and L. Soo-Young, "Environment Audio Scene and Activity Recognition through Mobile-based Crowdsourcing", IEEE Transaction on Consumer Electronics, vol. 2, no. 58, (2012).
- [4] F. Steffen, M. Ian, S. Christian, P. Christoph, S. Linda, S. Dmitry, V. D. V. Marijn, K. Florian and O. Micheal, "Geo-wiki.org: The Use of Crowdsourcing to Improve global land cover", Remote Sensing, vol. 3, no. 1, (2009).
- [5] D. Peng, L. Christopher, Mausam and W. Daniel, "POMDP-based control of workflows for crowdsourcing", Artificial Intelligence, vol. 202, no. 52, (2013).
- [6] Z. Zhi-Qiang, P. Ju-Sheng, X. Xiao-Qin and Z. Yong, "Research on Crowdsourcing Quality Control Strategies and Evaluation Algorithm", Jisuanji Xuebao/Chinese Journal of Computer, vol. 36, no. 8, (2013).
- [7] M. Lease and E. Yilmaz, "Crowdsourcing for Information Retrieval", ACM SIGIR Forum, vol. 45, no. 2, (2011).
- [8] G. Dinesh, K. Naidu, N. Animesh, N. Girija and P. Viswanath, "MoneyBee: Towards Enabling a Ubiquitous, Efficient, and Easy-to-use Mobile Crowdsourcing Service in the Emerging Market", Bell Labs Technical Journal, vol. 15, no. 4, (2011).
- [9] V. Jeroen and D. V. Arijen, "Obtaining High-quality Relevance Judgments Using Crowdsourcing", IEEE Internet Computing, vol. 16, no. 5, (2012).
- [10] Z. Liren and L. Ling, "An Interactive Vocabulary Learning System Based on Word Frequency Lists and Ebbinghaus' Curve of Forgetting", International Workshop on Digital Media and Digital Content Management, Hangzhou, China, (2011) May 16-18.
- [11] L. Na and Y. Fuyu, "Detection User's Long-term Interest Based on Ebbinghaus Forgetting Carve", ICIC Express Letters, Part B: Applications, vol. 2, no. 5, (2011).