A Social Network Water Army Detection Model based on Artificial Immunity

Hanhua Zhang¹ and Tao Li² and Yuqiao Wang³

¹1st School of Computer Science and Technology Wuhan University of Science and Technology Wuhan, China
²2nd School of Computer Science and Technology Wuhan University of Science and Technology Wuhan, China
³3rd School of Computer Science and Technology Wuhan University of Science and Technology Wuhan, China

Abstract. The wanton dissemination of water army information in social network seriously affects the authenticity of information received by users. With the continuous renewal of the covert means of the water army, the traditional screening means of the water army are no longer efficient. Combined with the idea of computer artificial immunity, this paper deeply analyzes the deep-level characteristics of the water army, finds out the stable and efficient danger and safety signal, calculates its optimal weight matrix through evolution, fuses DCA algorithm to calculate the antigen maturity, and then identifies the water army. The simulation results show that the recognition model based on the immune risk theory has a certain improvement in accuracy and recall rate, while the time cost is greatly reduced, which verifies its effectiveness and high efficiency.

1. Introduction
Micro-blog (including Sina Weibo, Twitter, etc.) has become the most popular social means in the world. However, while Weibo social networking has such a profound impact on all aspects of the lives of so many people, its features, such as low cost to release, wide range of dissemination and fast dissemination speed, have become a double-edged sword and begun to have negative effects on people. Behind every public opinion event, there are a large number of economic or political interest groups. It has become an urgent task to find an efficient detection method for Water Army. Water army, a group of Internet ghostwriters paid to post online comments with particular content, paid posters, netizens hired to leave fake comments and delete others. This paper proposes a social network water army recognition model based on heterogeneous linear computing method DCA. Four kinds of recognition signals are defined and their optimal weight matrix is evolved to detect water army. At the same time, the recognition accuracy is guaranteed, and the recognition time is greatly reduced.

2. Related research
In recent years, the identification of social network water army has become a focus of scholars. Generally, researchers study the attributes and behavioral characteristics of water army accounts, or the communication characteristics of water army information. Through in-depth analysis of the characteristics of various types of water army to distinguish the characteristics of the water army and
the use of the characteristics of the appropriate classifier to distinguish the water army and the information of it.

Zhang et al. [4] set 6 attributes as the characteristics of the water army classifier, and integrated the bayesian model and genetic optimization algorithm to improve the accuracy of water army identification without sacrificing the recognition rate of non-water army. Cheng et al. [2] proposed a new category of relational graph features on the premise of combining traditional user attributes and behavioral characteristics, and proved that the effectiveness of judging water army was significantly improved with the participation of new features. Han et al. [1] took the probability that the user is a water army as an implicit variable of the user's attribute characteristics and behavior characteristics, and constructed a probability graph model to calculate the probability that the user is a water army. Yuan et al. [5] analyzed a series of obvious characteristics of Weibo water army, used entropy value method to determine the weight of each characteristic index, and established an automatic identification model of Weibo water army combining multi-index comprehensive index method. Yang et al. [8] took a new approach, using sina Weibo's official rumor refuting platform, and selected a large number of micro-blogs that were officially identified as false information for research. They found and used real data to prove that these two new features, "place of release" and "type of blog user end", can provide effective help in distinguishing water army accounts. Similarly, Thomas et al. [11] obtained a large number of common features of water army accounts through detailed studies on more than 1.8 billion micro-blogs and 1 million accounts that were blocked by Twitter. It has laid a solid foundation for the subsequent classification and judgment of naval forces. Castillo et al. [13] analyzed a large number of micro-blogs from the perspective of micro-blog credibility, based on some topics that are easily affected by the water army. Finally, the characteristics of user behaviors such as Posting and forwarding are extracted, and the data processing algorithm of decision tree is used to automatically identify the credibility of microblog content.

On the other hand, Chen et al. [6] designed a semi-local centrality measurement method to balance the measurement method of low correlation centrality and high time consumption. By simulating SIR model and starting experiments in four kinds of real complex social networks to verify the effectiveness of its method for identifying "opinion leaders". Shah et al. [7] proposed a new topological quantity "rumor centrality" and used SIR's variant propagation model to simulate an evaluation model of rumor source. Zhang et al. [9] took the URL in Twitter content as the breakthrough point, analyzed accounts with similar urls, found suspected water army activities and information by studying the characteristics of micro-blogs with URL and their publishers, analyzed their intentions, and verified the validity of this model in judging true and false information by means of machine learning. Chen et al. [3] defined feature vectors and quantified propagation behaviors according to the interaction behaviors between information transmission subjects, and used the method of decision tree to detect the information transmitted by water army.

In addition, Irani et al. [10] studied a large number of social network accounts and established a large case base of static user profile content analysis. Through the comparison of several machine learning algorithms, the most suitable decision tree algorithm is found to distinguish water army users. Morris et al. [12] were keen to find that the relevant characteristics of Weibo account credit evaluation thought by users were not completely consistent with those published by the actual platform. Experiments were conducted by artificially controlling certain micro-blog features to evaluate the impact of each feature on credit rating. Through experiments, it is found that users do not judge the credibility of the account completely based on the authenticity of the content, but are influenced by heuristic such as user name. Thus, a series of characteristic factors that influence users' judgment of information authenticity are found.

3. The micro-blog water army identification model based on the danger theory

3.1. Identification framework of naval forces based on danger theory
Uwe have done the thorough research in Dendritic cells, they introduced the various status of Dendritic cells, in the literature [14] when it comes to Dendritic cells are able to control the immune response of antigen presenting cells. Immature DCs evolves into semi-mature DCs or mature DCs according to different signals collected, and then carries antigens into lymph nodes to provide cooperative stimulation signals and tolerance factors or stimulation factors. Finally, the proportion of mature antigens in total antigens determines whether there is a risk. Compared with traditional artificial immune algorithm, DCA algorithm does not depend on the self collection, for those not to hurt the body of autoantibody will not be processed, and its calculation process is linear, make more lightweight intrusion detection system, and does not require knowledge of the training process and normal and abnormal is in static analysis, so compared with other supervision and learning algorithm, its less time cost, greatly reduced the matching and the scale of the response, to a certain extent, improve the efficiency of the algorithm.

DCA algorithm is more concerned with finding more differentiated "signals", therefore, compared with other supervised learning algorithms, it takes less time and greatly reduces the size of matching and answering. Similarly, in the detection of abnormal accounts in social networks, more attention should be paid to the abnormal behaviors of accounts. By defining the signals of the detection model, the weight matrix of each recognition signal was obtained through evolutionary calculation. Finally, the signals were fused by DCA algorithm, and a probabilistic model of social network water army detection based on immune risk theory was obtained. Based on this, this paper defines four identification signals, two of which are safety signals ($D_{\text{event}}$, $T_{\text{guide}}$) and two of which are danger signals ($R''$, $R_{\text{tightness}}$).

3.2. Definition of danger signal and safety signal

1) event participation signal $D_{\text{event}}$

Due to the nature of social network mercenaries themselves, their existence depends on their intention to create false public opinion direction, mislead public opinion and ultimately achieve the political or economic purposes behind a specific event. Therefore, we believe that all social network water army behaviors are based on specific events. Water army users usually register for a short time, but their participation in a public opinion event is very high. We believe that when users participate in more public opinion events on Weibo and focus more on a few topics or topics within a unit of time, the suspected behavior of this user will increase correspondingly.

$T_{\text{reg}}$ is the registration time of the user, $E_{\text{event}}$ is the total number of events the user participated in, and $E_{\text{event}}$ is the number of events the user participated in. So the definition:

$$D_{\text{event}} = \frac{E_{\text{event}}}{T_{\text{reg}}} \quad D_{\text{event}} \in [0, +\infty) ;$$

(1)

2) second-order correlation signal $R''$

In the microblog network, due to its social attribute, the circle of friends of normal users is often close, and the users they pay attention to, such as those in the circle of relatives, friends, classmates and colleagues, tend to pay attention to each other. Therefore, a normal user will have more second-order correlations, that is, there will be more interrelations between the relevant users. The undirected graph $G=(V,E)$ is used to represent all the second-order concerns of the current user node $n$. Assuming that the degree of $n$ is $k$, then the adjacency matrix of $G$ is $A = (a_{ij})_{k \times k}$, thus, defined as:

$$R'' = -\frac{\sum_{i=1}^{k} a_{ij}}{k} \quad R'' \in \left[-\frac{k(k-1)}{2k}, 0 \right] ;$$

(2)

3) relationship tightness signal $R_{\text{tightness}}$

In order to avoid simple garbage user filtering and influence public opinion, water army users often focus on normal users and buy zombie powder in large quantities to create the illusion of a normal account. However, it is not difficult to find that the normal users who are concerned about the water army users seldom pay attention to the water army users. This leads to an obvious characteristic of
water army users, that is, the degree of close relationship is low, that is, the proportion of mutual fans of water army users is significantly lower than that of normal users.

As shown in the figure above, according to the number of mutual attention between water users and normal users, it can be clearly shown that the relationship between water users is closer than that between normal users. \( Following \) represents the number of followers of the current user, and \( Follower_{\text{mutual}} \) represents the number of users who are fans of the current user. So the definition:

\[
R_{\text{tightness}} = -\frac{Follower_{\text{mutual}}}{Following}, \quad R_{\text{tightness}} \in [-1,0]
\]  

(3)

4) guide tool utilization signal \( T_{\text{guide}} \)

Sina Weibo provides two kinds of "guiding tools", namely the topic symbol "\#" and the external link URL, to achieve the purpose of categorizing users' microblog content and strengthening third-party contact. In general, normal users do not use these two bootstrap tools on a large scale due to their own social diversity. In order to achieve the purpose of topic gathering and heat raising to influence public opinion, the water army often USES these two guidance tools on a large scale.
It can also be seen from the figure above that ordinary users are relatively concentrated in areas with low utilization rate of guiding tools, while water troops are more likely to appear in areas with more use of guiding tools. Hash\textsubscript{sum} represents the total number of microblogs that use "#", URL\textsubscript{sum} represents the total number of microblogs that use url, and Weibo\textsubscript{sum} represents the total number of microblogs of users. So we define:

\[
T_{\text{guide}} = \frac{\text{Hash}_{\text{sum}} + \text{URL}_{\text{sum}}}{2 \cdot \text{Weibo}_{\text{sum}}}, \quad T_{\text{guide}} \in [0, 1];
\]  

(4)

3.3. The optimal weight matrix of recognition signal is obtained by evolutionary computation

In the evolutionary computing model, a 4-row and 3-column matrix Weight is established to represent a weight matrix of 4 recognition signals. Because this model USES two kinds of signals, safety and danger, and is distinguished by positive and negative values, the first column in the three columns is the minimum value of weight 0, and the third column can guarantee the maximum value within the accuracy range of 1%, that is, the difference between the maximum value and the minimum value of the absolute value of all signals in the data set is 100 times. The second column is the optimal weight.

Initialize 10 rows and 4 columns of random population, each row represents an individual, and each column represents the weight of each recognition signal of the individual. Get the weight of each row by cross calculation get the current weight matrix var.

\[
X_i' = a X_i + (1 - a) X_j \quad \text{(5)}
\]

\[
X_j' = a X_j + (1 - a) X_i \quad \text{(6)}
\]

var was substituted into the DCA microblog water army recognition algorithm for calculation and judgment, and each individual was sorted according to the F1 value, and the corresponding length was assigned on the interval (0,1). Then, the individuals who can enter the next generation are selected by random matrix, and the individuals with high F1 value are correspondingly given a greater opportunity to pass on the genes to the next generation. The process of mutation is simulated as the exchange of arbitrary lengths of two random rows in population matrix Population, so as to obtain a new population matrix and conduct the next iteration. Keep updating the optimized recognition signal weight matrix and the highest F1 value. After several iterations, the weight matrix Weight of the recognition signal and its corresponding F1 value stabilize in a small interval, each weight value corresponding to Weight is the optimal weight.

3.4. Micro-blog water army recognition based on DCA algorithm

The greatest advantage of DCA algorithm lies in its linear calculation process. Based on the fusion calculation of danger signal and safety signal with high information gain, the scale of matching and response can be greatly reduced. Through continuous iteration of evolutionary calculation, the optimal recognition signal weight matrix is obtained. Finally, the antigen maturity is calculated, the identification results of Weibo water army were obtained.

\[
MCAV = \sum_{i=1}^{n} \text{Weight}_{i,2} \cdot \text{Signal}_i + b
\]

(7)

4. Experiment and analysis

The experimental operating environment was Windows 10 operating system, 3.2Ghz four-core processor and 8GB memory. The experimental software is Python 3.6 and MySQL 5.5.

4.1. data preprocessing

Through the crawler program, more than 2 million pieces of microblog information and user information were crawled for more than a dozen hot public opinion events. By pruning and data
processing, such as the user will receive the fan number, the number of attention, the attention to each other, each other on proportion, the registration time, friends forwarded when proportion of fans, each content repetition rate, time occurrences, # usage, url utilization rate, forwarding microblogging proportion, forward than not empty, and data is stored in relational database. According to the above definition of danger signal and safety signal, four signal values of each user are calculated and stored in the relational database. In all the existing data sets, 5,000 microblog accounts were selected for repeated manual tagging by multiple people, and 4,677 agreed ones were retained as the training set of evolutionary calculation.

4.2. Evolutionary calculation of recognition signal weight matrix
By analyzing the existing data, the maximum and minimum values of each recognition signal in the data set are obtained. Thus, the weight matrix of the recognition signal is obtained. Since the security signal and danger signal are set in this paper, 0 is taken as the threshold value to judge the microblog water army. The weight matrix selected, crossed and varied in the iterative process of evolutionary calculation is continuously substituted into the DCA microblog water army recognition algorithm, and finally the optimal weight matrix of the convergent identification signal is obtained.

4.3. Effectiveness of identification signals
In order to ensure that the recognition signals required by DCA algorithm are as efficient as possible on the premise of ensuring effectiveness and reduce the amount of evolutionary computation, we also conducted a set of comparative tests to verify the results. One group only uses the four recognition signals \(D_{event}, T_{guide}, R^\prime, R_{tightness}\) proposed in this paper, while the other group, in addition to the four recognition signals, adds the general features commonly used by the water army recognition model such as literature (number of fans, number of followers, number of tweets, proportion of non-empty forwarding, and whether or not the signature is included). This paper also uses the DCA and evolutionary computing fusion of the military recognition model for recognition. The following results are obtained:

![fig 3. Effect Contrast3](image)

The results show that in the control group with conventional features, the recognition effect is less than 1% higher than that of the group using only the four recognition signals proposed in this paper, while the time performance is significantly lower than that of the control group without conventional features. This proves that the four recognition signals proposed in this paper have higher recognition efficiency.

5. Result evaluation
In view of the current new characteristics of naval forces' behaviors, the results of the determination of naval forces' behaviors using this model are compared with the results of the previous traditional model method (FO+FR+NFR+UR model). Precision, Recall and F1 value were selected as the reference indexes for model detection level.
Definition: YY is the water army classified correctly by their respective models, NY is the normal user wrongly judged as the water army, and YN is the water army wrongly judged as the normal user. So:

\[ Prec. = \frac{YY}{YY+NY} \]  
\[ Recall = \frac{YY}{YY+YN} \]  
\[ F1 = \frac{2 \cdot Prec \cdot Recall}{Prec + Recall} \]

The traditional model selects the "fan value (FO), friend value (FR), non-marketing activity participation (NFR), URL usage (UR)" model, and finally obtains its accuracy, recall rate, F1 value and the evaluation index of this model for comparison as shown in the figure.

Through the 10-fold verification of the feature vectors selected by the traditional model, the average results are obtained. Then, the social network water army detection model defined in this paper based on the immune risk theory is used. Through the method discussed in this paper, the optimal weight matrix is obtained by evolutionary calculation, and the microblog water army is judged by DCA algorithm.

The results show that the accuracy recall rate and F1 value of this model are higher than the traditional model to some extent, both exceeding 90%, and the performance of data processing time is significantly higher than that of the traditional model.

Receiver Operating Characteristic (ROC) is also one of the commonly used model evaluation criteria. Compared with the accuracy, recall rate and F value and other indicators, Receiver Operating Characteristic curve is more inclined to care about the score between positive and negative samples rather than the specific score value, which also verifies the performance of classifier from another aspect. Through the test and comparison of Area under ROC curve, the AUC value of 0.941 in this model is also higher than that of 0.8783 in the traditional model.
6. Conclusion
Social networks such as Weibo are already having a profound impact on every aspect of people's work and life, and water army is taking this opportunity to make use of social networks to influence public opinion judgment. In recent years, the identification of the water army has attracted the attention of scholars. However, the complexity of social networks and the constant updating of covert means of water army bring some difficulties to the identification and judgment of water army. Through the combination of computer recognition and artificial immunity, the in-depth analysis of the characteristics of the water army, the use of evolutionary computing to get the weight of each characteristic signal, and the classical algorithm of artificial immunity DCA combined, a social network water army identification model with high reliability and better performance is obtained.

References
[1] Zhongming Han, Minfeng Xu, Dagao Duan. A model of water army identification based on micro-blog probability map [J]. Computer research and development, 2013, 50(s2):180-186.
[2] Xiaotao Cheng, Caixia Liu, Shuxin Liu. Methods of microblogo discovery method based on the characteristics of the relationship graph [J]. Journal of automation, 2015, 41(9):1533-1541.
[3] Kan Chen, Liang Chen, Peidong Zhu, et al. An online social network water army detection method based on interactive behavior [J]. Journal of communications, 2015, 36(7):120-128.
[4] Yanmei Zhang, Yingying Huang, Shijie Gan, et al. Research on the identification algorithm of micro-blog network water army based on bayesian model [J]. Journal of communications, 2017, 38(1):44-53.
[5] Xuping Yuan, Renwu Wang, Boyin Zhai. Automatic identification of micro-blog water army based on comprehensive index and entropy method [J]. Journal of intelligence, 2014(7):176-179.
[6] Chen D, Lü L, Shang M S, et al. Identifying influential nodes in complex networks [J]. Physica A Statistical Mechanics & Its Applications, 2012, 391(4):1777-1787.
[7] Shah D, Zaman T. Rumors in a Network: Who's the Culprit? [J]. IEEE Transactions on Information Theory, 2011, 57(8):5163-5181.
[8] Yang F, Liu Y, Yu X, et al. Automatic detection of rumor on Sina Weibo [C]// ACM, 2012:1-7.
[9] Zhang X, Zhu S, Liang W. Detecting Spam and Promoting Campaigns in the Twitter Social Network [C]// IEEE, International Conference on Data Mining. IEEE, 2013:1194-1199.
[10] Irani D, Webb S, Pu C. Study of Static Classification of Social Spam Profiles in MySpace [C]// International Conference on Weblogs and Social Media, Icwsm 2010, Washington, Dc, Usa, May. DBLP, 2013:591-597.
[11] Thomas K, Grier C, Song D, et al. Suspended accounts in retrospect: an analysis of twitter spam [C]// ACM, 2011:243-258.
[12] Morris, Ringel M, Scott, et al. Tweeting is believing?: understanding microblog credibility perceptions[C]// ACM 2012 Conference on Computer Supported Cooperative Work. ACM, 2012:441-450.

[13] Castillo C, Mendoza M, Poblete B. Information credibility on twitter[C]// International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April. DBLP, 2011:675-684.

[14] Greensmith J, Aickelin U, Tedesco G. Information fusion for anomaly detection with the dendritic cell algorithm[J]. Information Fusion, 2010, 11(1):21–34.