Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 3, No. 3, 2023

Quantum-based Detection of Higly Semantically Similar Social bot

Yulin Liu*

Cyberspace Security College of Chengdu University of Information Technology, Chengdu, Sichuan 610225, China
* **Corresponding author Email:** 1045734823@qq.com

Abstract: Social bot often publish content that has a high degree of similarity in text or semantics. Combining this feature, this paper designs a method to detect social bot. First, the user-published text is used as the input to use sentiment analysis and data cleaning to predict the content. Processing and classification, and then adding the quantum similarity algorithm in the emerging quantum discipline in recent years to the structural system of social bot detection to perform similarity clustering, which greatly improves the operating efficiency of the system, and then obtains the corresponding user attribute characteristics for artificial intelligence. Intelligent algorithm classification, on the collected real dataset, marked two social bot groups for machine learning classification and detection, and listed two traditional similarity algorithms for comparison, the results show that the quantum similarity results Compared with the two traditional similarity algorithms, it has improved, and the average accuracy improvement was about 2%.

Keywords: Artificial intelligence; Social bot; Detection; Machine learning.

1. Introduction

In recent years, with the development of Internet technology and the substantial improvement of living standards, people's tools for obtaining information have gradually changed from the previous TV, newspapers, etc. to the current network application platform. As far as the current social platforms are concerned, such as Twitter, Facebook and Sina weibo makes it easier for users to express their own opinions and understand other people's opinions through the Internet, and provides convenience for users to contact family and friends, so that social platforms have gained a lot of traffic and attention, and this trend will continue to accelerate in the future. In addition to the increasing number of users, it has brought many hidden dangers, among which the public opinion influence of social bot is a very serious problem.

A class of users controlled by artificial software programs is called social bot. At first, social bot was used to provide services to users, such as automatically replying to user questions. However, the rise of malicious social bot has a negative impact on other users and public opinion security and even countries. Security poses a great threat. Manipulators mainly manipulate malicious social bot to do three things out of profit: (1) Focus on employer users. In order to make the account look popular and well-known, employer users will hire operators to follow them and make their own fans. A large amount of expansion, which may cause other normal users to pay attention to employer users. (2) A large number of false comments, false advertisements and rumors are released to mislead other normal users. For example, during the new crown pneumonia, malicious social bot posted more negative content on the Internet than humans, thus amplifying people's emotions and causing panic, and even leading the blame to China [1]. (3) The third category is a very dangerous category. The operators are malicious social bot participating in national politics. The most famous one is the 2016 US election. About 19% of related tweets are from social bot. These Social bots not only affect electoral politics, but also the economy and society [2]. For social bot, especially malicious ones, the research work of detecting them can not only lead the correct public opinion orientation, but also further promote the stability of social order, so the research on its detection work is very necessary.

This paper mainly focuses on the detection of highly semantically similar types of social bot, which must satisfy a specific topic, and the topic meets two characteristics (1) has value, which means that it has research value, and its topic can change the direction of public opinion and the safety of public opinion to a certain extent. cause impact. Such as the new crown virus, Huawei 5G. (2) It is contagious. The literal meaning here is that the topic can spread widely. Because of the high semantic similarity, the experiment will use the similarity algorithm for text clustering to filter normal users. The traditional similarity algorithm is effective in text, but lacks in semantics and cannot be taken into account.

This paper proposes to use the quantum similarity algorithm [3] that takes both semantic and textual similarity into account in the emerging quantum disciplines for experiments, and selects two traditional similarity algorithms for comparative experiments. The results show that the quantum similarity results Compared with the two traditional similarity algorithms, it has improved.

2. Literature References

At present, the mainstream social bot detection is based on machine learning and graph detection methods.

2.1. Machine learning based detection

Jorge [4] et al. used a single-class classifier to detect social bot, and the accuracy rate exceeded 0.89, and proved by experiments that in the face of the diverse types of social bot and the status quo of continuous anti-detection technology changes, single-class detection is better than binary classification. better adaptability. Due to the dataset imbalance problem, BinWu[5] et al. proposed an improved conditional generative adversarial network to expand the dataset, that is, combining Wasserstein distance with gradient penalty and clustering algorithms as a new data augmentation

method, In order to improve the accuracy of social bot detection, experiments show that the F1 score is 97.56%. Feng et al. [6] proposed a method called BotFlowMon to detect social bot from the perspective of network data traffic. The experimental results show that BotFlowMon has an accuracy rate of 96.1% and can see social bot in an average of 0.71 seconds. traffic.

Compared with traditional machine learning detection methods, deep learning detection methods are more suitable for processing big data. Wu et al. [7] designed a social bot detection structure based on deep neural network and active learning. 30 features were extracted from four types of tweet content, tweet comments, likes and other interactive information, and time information to detect social bot, including 9 self-created features, and an active learning method was designed to effectively expand data labels, the accuracy rate reached 98.87%. Sneha Kudugunta et al. [8] proposed a deep neural network based on a contextual long-term short-term memory (LSTM) architecture, which takes both the tweet content and the user's personal information as input, and uses fewer features to detect social bot, but the accuracy of the results is good, exceeding 96%. Greeshma Lingam [9] et al. designed a deep Q-network model, which takes the content attributes of tweets, user personal information content attributes and social relationship attributes as input and deep Q-learning model. Combined, based on the Q-value function to detect social bot, the overall accuracy rate has reached 93%.

2.2. Graph-based detection methods

In graph-based detection, social bot are generally referred to as sybils and are often used to forge or create multiple identities to launch malicious campaigns. Gao et al. [10] designed a framework, SybilFrame, which is mainly based on graphs and combined with machine learning as an auxiliary method to detect Sybil. SybilFrame is a multi-stage classification method. The first stage extracts useful information on the dataset, the second stage is the post-inference layer. Experimental results on large-scale datasets show that SybilFrame performs much better than previous methods, but the disadvantage is that the accuracy depends on the external classifier.

The classic graph-based detection Sybil method believes that Sybil is only connected to a small number of normal users, and this view does not hold in real social networks. In order to deal with this problem, Zhang [11] et al. developed SybilSAN, a two-layer model, It is intended to make full use of the interaction information and connections between normal users and Sybil, and propose a more practical Sybil attack model. Based on the attack model, three random walk-based algorithms are coupled to detect Sybil, and SybilSAN also has Stronger detection robustness, but performance degrades when the Sybil population is not concentrated and dispersed.

El-Mawass [12] et al. designed a hybrid method to detect social bot, combining graphs and traditional supervised learning classification, where graphs are not constructed using user interaction connections on social networks, but the application of user usage The program is similar to the graph, and then the Markov random field model is constructed on this basis. The experiments show that the detection of the supervised learning combined with the Markov random field model in the paper has improved the accuracy and recall rate of the detection only by supervised learning.

3. Experiment

3.1. Work process

Figure 1 shows the specific working steps of the system.

1. The input is the original text. It is a user-published content field in the data set crawled by the crawler, which can be mainly divided into two types, one is to reply to other people's tweets, and the other is to non-reply to the tweets.

2. The second step is preprocessing, the first of which is data cleaning. The purpose is to obtain as much as possible the complete content posted by the user from the original text to prepare for subsequent operations. The main cleaning contents are:

   (1) Remove @user, @ is a function provided by Twitter for users. For example, if user A @user B in a tweet, user B can see reminders and tweet jumps in "message", which contributes to the entire sentence the value is small and needs to be filtered.

   (2) Remove the URL, the original text will have a URL at the end.

   (3) Remove non-English language tweets, because quantum similarity is only for English, so remove non-English sentences.

   (4) Remove irrelevant characters, such as "*", "[", "&" and other strange characters.

   (5) According to the manual observation of this topic tweet, users often regard # keywords, such as #HuaWei, #5G, #HUAWEI, etc. as part of the sentence, so there is no need to remove the #tag, just remove the "#" character. Can.

   (6) Remove stop words, and filter stop words with the help of the functions provided by the word segmentation library. Usually, stop words have less value to sentences, and filtering can improve the efficiency of the program.

3. Sentiment analysis. Take the sentence output in 2 as input, call the VADER [13] library for sentiment analysis, and output the sentiment of each sentence, including positive, negative and neutral sentiments.

4. Participle clauses. The sentence output in 2 is used as input, and the public NLTK [14] library is used to segment the sentence according to the part of speech set in the library.

5. Quantum similarity clustering. The purpose of similarity clustering is to screen social bot into the identification link as much as possible, and eliminate normal users, which can greatly enhance the running speed of the system and improve the recognition accuracy. The outputs of 3 and 4 are used as input, and then clustering is performed to obtain several group categories, calculate the texts one by one for similar grouping. There are the following two clustering results.

   (1) The number of sentences in the group is less than one-tenth of the number of users. This value is not large, and it only represents the opinions of a small number of people, so it is eliminated.

   (2) The number of sentences in the group is one tenth higher than the number of users, reserved.

6. Feature extraction. Find the user corresponding to the sentence from the output in 5, and extract the selected features as the output.

7. Machine learning identification, take the output in 6 as input, and use the logistic regression algorithm in the sklearn library to identify social bot.

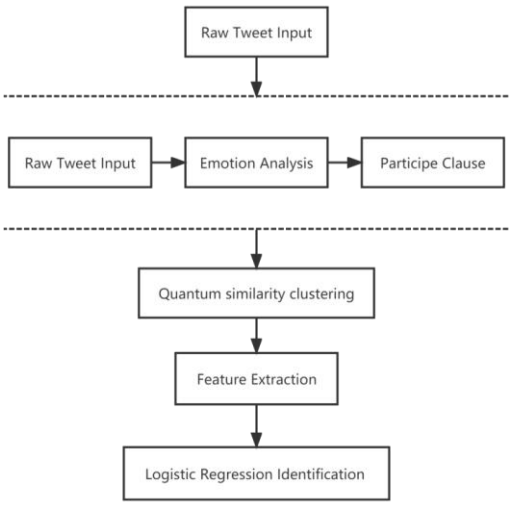

Figure 1. Detection process

3.2. Quantum Similarity [3]

In recent years, quantum disciplines have gradually developed. Among them, quantum theory has a good mathematical modeling framework, which is combined in the fields of economic theory, psychology and information science, and there are some attempts in natural language in artificial intelligence. The quantum similarity algorithm is also known as the sentence similarity calculation method based on quantum theory. It uses the properties of quantum theory in terms of superposition, interference, entanglement and incompatibility to establish a sentence quantum state modeling model, and then introduces the concept of fidelity. To judge whether two quantum states are close, that is, to judge whether two sentences are similar. The main process is to decompose the two sentences into state functions, and then use the density matrix to represent the meaning of a single complete sentence, and then further obtain the fidelity calculation expansion between them, and then represent the word as a semantic word vector (Word Embedding) Do the calculation to get the similarity value of the two sentences.

A natural language sentence is composed of words with different parts of speech, such as nouns, verbs, prepositions and adjectives, etc. This paper decomposes sentences by parts of speech and forms a set $W=\{\omega_1,\omega_2...\omega_i\}$, where ωi is a set of words with different parts of speech, $\omega i=\{\nu_1,\nu_2...\nu_j\}$, where ν_j represents different words of the same part of speech, for ω_i, where the word vectors are superimposed $\sum_j|\nu_j\rangle$, a Hilbert space χ whose basis vectors are orthogonal to each other is obtained, as shown in Table 1.

Table 1. Part of speech classification of natural language sentences

Part of Speech	n	v	adj	conj	adj	n
Sentence 1	I	like	sunny	and	honest	boys
Sentence 2	I	like	handsome	and	cheerful	boys

The Hilbert space of the adjective in sentence 1 is Consists of $|\text{sunny}\rangle$ and $|honest\rangle$ superimposed. The whole sentence can be expressed as "I like sunny boys" and "I like honest boys".

Here, the part-of-speech set W is represented as a state function $|\Psi\rangle=\chi_1+\chi_2+...+\chi_i$, and a weight is assigned to each Hilbert space (here represented by a complex value $p_i=e^{i\theta_i}$, the value is optional), the weight can be understood It is assigned for each part of speech to contribute differently to the sentence as a whole, then the mathematical expression of the function is as follows

$$|\Psi\rangle=\frac{1}{\sqrt{m}}\sum_i p_i|\omega_i\rangle=\frac{1}{\sqrt{m}}\sum_i e^{i\theta_i}|\omega_i\rangle \quad (1)$$

In the formula, $1/\sqrt{m}$ is the normalization coefficient, which ensures that the normalization $\langle\Psi|\Psi\rangle=1$,Where $1/\sqrt{n}$ in $\omega=1/\sqrt{n}\sum_j|\nu_j\rangle$ is also a normalization coefficient, It can be expressed as

$$|\Psi\rangle^{(1)}=\frac{(\frac{e^{i\theta_1}}{\sqrt{2}}(|I\rangle_1+|boys\rangle_1)+e^{i\theta_2}|like\rangle_2+\frac{e^{i\theta_3}}{\sqrt{2}}(|sunny\rangle_3+|honest\rangle_3)+e^{i\theta_4}|and\rangle_4)}{2}$$

$$|\Psi\rangle^{(2)}=\frac{(\frac{e^{i\theta_1}}{\sqrt{2}}(|I\rangle_1+|boys\rangle_1)+e^{i\theta_2}|like\rangle_2+\frac{e^{i\theta_3}}{\sqrt{2}}(|\text{handsome}\rangle_3+|\text{cheerful}\rangle_3)+e^{i\theta_4}|and\rangle_4)}{2}$$

The density matrix is obtained by the outer product of the state vectors, expressed as

$$\rho_s=|\Psi\rangle\langle\Psi|=\frac{1}{\sqrt{m}}\sum_i p_i|\omega_i\rangle\frac{1}{\sqrt{m}}\sum_j p^*{}_j\langle\omega_j|=\frac{1}{m}\sum_{i,j}p_i p^*{}_j|\omega_i\rangle\langle\omega_j|. \quad (2)$$

And then get their fidelity calculation formula

$$\rho^{(1)}=|\Psi^{(1)}\rangle\langle\Psi^{(1)}|=p_i^2(1)|\omega_i^{(1)}\rangle\langle\omega_i^{(1)}|,$$

$$\rho^{(2)}=|\Psi^{(2)}\rangle\langle\Psi^{(2)}|=p_j^2(2)|\omega_j^{(2)}\rangle\langle\omega_j^{(2)}|, \quad (3)$$

Then their fidelity is calculated as

$$F=Tr(\rho^{(1)}\rho^{(2)})==p_i^{2(1)}p_j^{2(2)}|\langle\omega_i^{(1)}|\omega_j^{(2)}\rangle|^2 \quad (4)$$

The similarity between them can be obtained by the inner product $\langle\omega_i^{(1)}|\omega_j^{(2)}\rangle$ of the two right losses. But this is a complex value, but if it is multiplied by its own complex conjugate, A real value can be obtained, which is the definition of fidelity. Represent each word in it as a semantic word vector, then calculate the cosine similarity of the two words, accumulate and average through the loop iteration, you can get the similarity value of the two sentences, and then perform nonlinear transformation to determine whether the two sentences are similar, gather sentences judged to be similar together to form a group.

3.3. Figure

(1) Status number: The total number of tweets, replies and retweets.

(2) Favorite number: The total number of favorite tweets.

(3) Verifier: Verification is mainly a function set by Twitter in order to allow some users to be better recognized by other users. Only after verification can there be a verification label. Social bot cannot pass the verification.

(4) Ratio of followers to following: Followers and following represent the number of followers and

followers, respectively. This can also be understood as the relationship between users. Normal users will not pay attention to social bot, and social bot will pay attention to normal users for profit or disguise. Therefore, social bot usually has a higher fpf value than normal users.

(5) Retweet number: The number of tweets retweeted by other users in practice. Usually, tweets published by normal users are rarely retweeted by other users, so the ret value of social bot will be higher than that of normal users.

(6) The original amount: The original proportion of user tweets. Normal users use Twitter to express their views on life dynamics or events, while social bot will retweet other people's tweets for this purpose, so usually The Originalp value for normal users will be higher.

(7) Quantity: In the selected event or topic, for the purpose of interest, usually social bot will tweet and retweet more than normal users on the event topic, so usually the quantity value of social bot will be higher than that of normal users. higher.

(8) Registration Date: Calculation formula: the total number of days from the registration time to the present time/the number of days in a year (the experimental value is 366). Usually, the registration time of social bot will be related to the topic of the event, such as registration after the event occurs. So usually the regtime value of social bot will be lower than that of normal users.

4. Experimental results

4.1. DataSet

The social platform selected in this paper is Twitter. The data set is based on the API interface provided by the platform and crawled by crawlers. The selected topic is "Huawei + 5G", with a total of 36,677 pieces of data. The selected fields are the original post link, content and key. Words, post time, author, author link, likes, platform, replies, retweets, and author location. The user information data fields are the number of pictures and texts, the number of likes, the number of fans, the number of followers, the number of original tweets, the total number of tweets, whether to authenticate, and the registration time. We select two types of social bot groups for experimental comparison in our experiments: (a) the exact type of tweet replication (b) the type of tweets with the same semantics and roughly similar types of tweet texts. Both social bot groups were manually labeled in the dataset.

4.2. Comparison of experimental results

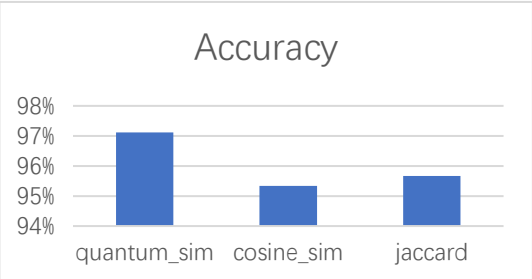

Figure 2. Test results of type a social bot team

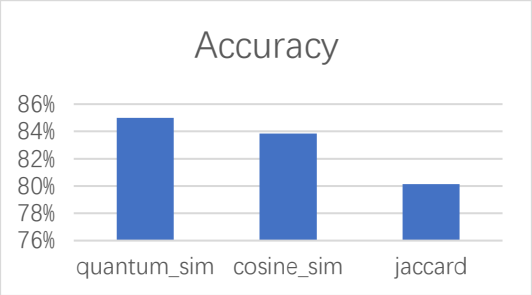

Figure 3. Test results of type b social bot team

The experiments in this paper found that similarity clustering filtering is very important in identifying social bot or eliminating the amount of data, which can reduce a lot of data. Figure 2 is the detection result of the a-type social bot team. The accuracy histogram of the three algorithms shows that the accuracy of quantum similarity is about 97%, which is 2% higher than the other two traditional similarity algorithms. Figure 3 b-type the detection results of the social bot team show that the accuracy rate of quantum similarity is about 85%, which is 3% and 5% higher than the other two algorithms, respectively. Experiments show that the method we designed can accurately identify social bot, and the cross-disciplinary combination the quantum similarity algorithm has better performance than the traditional similarity algorithm in judging the similarity of social bot texts.

5. Conclusion

We briefly introduce the background and meaning of social bot recognition, and then gives an overview of the current related work, and designs a detection method for social bot. First, the user's published text is used as input to eliminate data and then use sentiment analysis. Classify it and then use the quantum similarity algorithm to cluster and filter out the clusters, and finally get the user characteristics corresponding to the text and use the artificial intelligence algorithm for detection. This method combines the quantum similarity algorithm of quantum science and the detection of social bot in computer science. An interdisciplinary combination was carried out, two types of social bot groups were manually marked from the data set, and two classical similarity algorithms were selected for comparative experiments.

References

[1] Shi, W., et al. (2020). "Social bot' Sentiment Engagement in Health Emergencies: A Topic-Based Analysis of the COVID-19 Pandemic Discussions on Twitter." Int J Environ Res Public Health 17(22).

[2] Hagen, L., et al. (2020). "Rise of the Machines? Examining the Influence of Social bot on a Political Discussion Network." Social Science Computer Review 40(2): 264-287.

[3] Bingqing, M. and M. Lei (2021). "Study on sentence similarity based on quantum theory." Journal of East China Normal University (Natural Science) 2021(1): 60.

[4] Rodríguez-Ruiz, J., et al. (2020). "A one-class classification approach for bot detection on Twitter." Computers & Security 91.

[5] Wu, B., et al. (2020). "Using Improved Conditional Generative Adversarial Networks to Detect Social bot on Twitter." IEEE Access 8: 36664-36680.

[6] Feng, Y., et al. (2020). "Towards Learning-Based, Content-Agnostic Detection of Social Bot Traffic." IEEE Transactions on Dependable and Secure Computing: 1-1.

[7] Wu, Y., et al. (2021). "A novel framework for detecting social bot with deep neural networks and active learning." Knowledge-Based Systems 211.

[8] Kudugunta, S. and E. Ferrara (2018). "Deep neural networks for bot detection." Information Sciences 467: 312-322.

[9] Lingam, G., et al. (2019). "Adaptive deep Q-learning model for detecting social bot and influential users in online social networks." Applied Intelligence 49(11): 3947-3964.

[10] Gao, P., et al. (2015). "Sybilframe: A defense-in-depth framework for structure-based sybil detection." arXiv preprint arXiv:1503.02985.

[11] Zhang, X., et al. (2018). Sybil Detection in Social-Activity Networks: Modeling, Algorithms and Evaluations. 2018 IEEE 26th International Conference on Network Protocols (ICNP): 44-54.

[12] El-Mawass, N., et al. (2018). Supervised Classification of Social Spammers using a Similarity-based Markov Random Field Approach. Proceedings of the 5th Multidisciplinary International Social Networks Conference on - MISNC '18: 1-8.

[13] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

[14] Bird, Steven, Edward Loper and Ewan Klein (2009).Natural Language Processing with Python. O'Reilly Media Inc.