Proposing Chatbot Model for Managing Comments in Vietnam

Phat Nguyen Huu[1(✉)], Cam Do Manh[1], and Hieu Nguyen Trong[2]

[1] School of Electronics and Telecommunications, Hanoi University of Science and Technology (HUST), Hanoi, Vietnam
phat.nguyenhuu@hust.edu.vn, cam.dm165801@sis.hust.edu.vn
[2] National Institute of Patent and Technology Exploitation (NIPTECH), Vietnamese Ministry of Science and Technology, Hanoi, Vietnam
nthieu@most.gov.vn

Abstract. Today, the behavioral culture on social networks is a painful issue. State agencies have been trying to clean up the network environment of country. Many policies are proposed to process videos and clips with offensive content. However, it is a small part of cleaning up the network environment. We often see hateful comments on social media sites. It exists anywhere from social media to online games that are difficult to control and punish because of their big data. There are not too many social networking sites and online games until now. Therefore, it is not too difficult for communities to limit inappropriate words. Therefore, we offer a chatbot model to manage the comments that helps to clean the network environment in the paper. The results show that the proposal model achieves up to 75% accuracy with 100,000 comments.

Keywords: Chatbot · Impolite comment · Natural language processing · AI · Machine learning

1 Introduction

With the explosion of internet, the number of users is increasing. There are about 2.6 billion Facebook users and 1.7 billion people use daily [15,18]. In Vietnam, there are 64 million Facebook accounts per 90 million people. However, there is no shortage of ingredients that always leave offensive comments and go against public opinion that makes readers uncomfortably. In order to avoid harmful effects on society, we need to remove it. Therefore, we propose a chatbot model to help solve this problem.

ChatBot is a computer program that conducts a conversation through instant messaging [12]. It can automatically answer questions or handle situations. Scope and complexity of chatbot are determined by the algorithm of their creators. It is used for many areas such as e-commerce, customer service, healthcare, banking and finance, and entertainment services.

N.-S. Vo et al. (Eds.): INISCOM 2021, LNICST 379, pp. 287–297, 2021.
https://doi.org/10.1007/978-3-030-77424-0_24

Chatbot is divided into two categories:

1. The system targeting on an application domain (Task-Oriented) is called open domain (OP).
   Auto-responder model on OP allows users to participate any topic. Social media networks (Facebook or Twitter) are usually OP and they have many topics. Consequently, the requiring knowledge is created to answer OP dialogues that becomes more difficult. However, the collection and extraction of data from this domain is quite large and simple.
2. The system without a target orientation is called close domain (CD).
   Auto-responder model often focuses on answering questions relating to a specific domain such as health, education, travel, and shopping. In the model, the space for input and output patterns is limited since these systems try to achieve a very specific goal. Technical customer support or shopping assistants are closed to domain applications. These systems are unable to communicate and they only perform specific tasks in the most efficient way. Users are able to ask and answer anything. However, the system is not required to handle them.

Each approach to the problem has a different solution. Inappropriate sentences appear more and more with the growing popularity of social media and comments. However, the difficult problem is that Vietnamese have the ability to magically combine together to create extremely diverse sentences. Depending on the context, it can be understood as an offensive sentence if listing all those words into forbidden and controlling words is completely possible. However, it requires a very large database. Besides, people often try to circumvent the law. They can explain other ways such as antonyms, synonyms, spelling, abbreviations, adding or subtracting words, etc. with the same idea.

In the paper, we propose a method to solve the diversity of objectionable online based on chatbot model that can identify and classify inappropriate statements on the Internet. The rest of the paper includes ve parts and is organized as follows. Section 2 presents the proposal algorithm. Section 3 will evaluate the proposal model and analyze the results. In final section, we give conclusions and future research directions.

2 Proposal Algorithm

2.1 Theoretical Basis

Firstly, we need to set out the requirements for our algorithm. In the paper, our request will be:

1. Automatically detecting sentences that do not match with high accuracy.
2. The program can be integrated into many different languages that is able to be used widely.
3. The maximum amount of time to process per comment is less than 30 s.

Fig. 1. System structure diagram.

From the above requirements, we propose the structure diagram of system as shown in Fig. 1.

In Fig. 1, we have:

Pre-processing Block: Converting the input sentence into an array containing meaningful words. It includes the steps, namely separating Vietnamese words, data cleaning, handling nonsense words, and defining the meaning of each word.

Determining Block: Based on a defining array and a set standard, the system determines the level of whole sentence.

Responding Block: From the level of sentence and components, chatbot will proceed to give the most appropriate answer.

In the scope of study, we have not found any documents for processing sensitive words in Vietnamese. More detail of system structure will be presented below.

Data Collection

The difficulty for testing effectiveness of chatbot is the dataset of comments on social networking sites. We do not have the comments data since we create them from Facebook. Our self-built dataset has 100,000 comments.

Pre-processing Data

Separating Vietnamese Words

Natural language processing includes a lot of problems such as machine translation, text summarization, information retrieval, information extraction, etc. To solve the problems, the word segmentation is very important. It will determine the success of system.

To solve the problem, we need to analyze properties of words of Vietnamese as follows:

- Infinitive word: form and meaning of words are syntactically independent.
- Words include single words, complex words, and compound words.
- Words are structured from language. The recognition of words of Vietnamese is called clustering as shown in Fig. 2.

In Fig. 2, there is more than one way to understand this sentence where the second way has no meaning.

As we all know, Vietnamese text often puts spaces between syllables. A word has more syllables since there are many ways to divide it. This causes ambiguity. This ambiguous resolution is called the word separation problem.

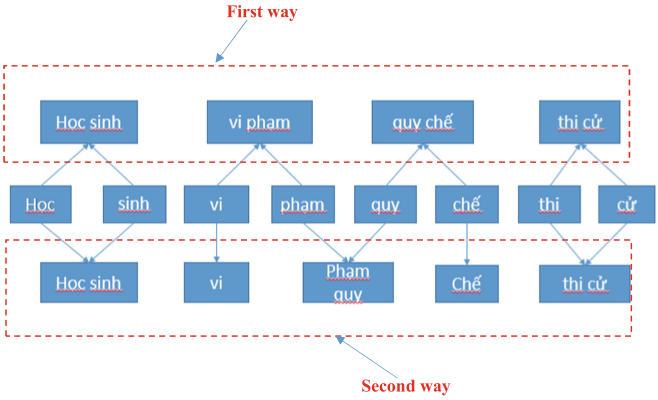

Fig. 2. The problem of clustering in Vietnamese [16].

The most important criterion in word separation is accuracy. We have achieved 97% accuracy on words. However, if calculating according to the sentence, the accuracy is only about 50%. The difference is caused by the complexity of Vietnamese.

Currently, there are several approaches to word separation problem as follows [13]:

- *Maximum match* [11,17]*:* We put words to cover all the sentences and satisfy certain heuristic. The advantage of method is very fast. However, there are many limitations such as low accuracy or inability to handle words that are not in the dictionary.
- *Rules* develops a manual or automatic rule set to distinguish allowing and authorizing combinations.
- *Graphization* builds a graph to represent the sentence and solve the problem of finding the shortest path on the graph.
- *Machine learning* considers the problem of string labeling. The way is used in JVNSegmenter [14].
- *Language model* gives several ways for separating whole sentences. This is the approach of vnTokenizer [2,3].

In this article, we use the *Maximum matching* method based on [14].

Data Cleaning

After separating words, text appears many special characters and punctuation. These ingredients reduce the efficiency of treatment process. In this section, we convert all capital words to low case and remove punctuation marks.

Solving Meaningless Words

This is the key point of the paper. In online comments, acronyms are often used.

If we only split and categorize words, we will miss many offensive sentences that still exist on social networks. There are several avoiding ways to use as:

– Using alternative words to describe the sentence.
– Using marks to center sentences.

These writing styles can fully express the meaning of offensive word. It will not create an offensive word when separating them. The common point of two ways is that the separating words are nonsensical or single words. Therefore, we need to process the separating words.

As shown in Fig. 2, we will have two steps to handle the problem.

– *Matching words:* Applying to words with one to two letters. We put them together into a new word. If that word makes sense, we will re-assemble it. Otherwise, we perform step 2.
– *Swap*: Vietnamese letters are divided into vowels and consonants. There are many Vietnamese words that have no meaning. However, their vowels and consonants are similar to the offensive sounds. Therefore, it will be used as a substitute for offensive words and the reader can still understand their meaning. Based on this point, we separate the vowels and consonants. If the system can match words with offensive meanings, we will update them into dictionary.

As a result, we improve the accuracy of separating from daily comments.

Classify Meaning of Words
As mentioned above, one of the most common ways to be offensive is to use synonyms and antonyms. To solve the problem, we propose to group commonly using synonyms and antonyms. We are dividing them into the following groups: offensive words, bad words, proverbial pronouns, animals, names of vocation, general, body, sensitive words, comparison words, negative words, activity, other words.

Determining Level of Word
Based on the division of antonyms, there are many groups of words that are not offensive. We can see that a few words is able to create offensive statements. There are offensive statements to this person that are extremely vulgar. However, it feels normal. To solve this problem, we proceed to create a norm to determine the level of objectionable sentences. Regulations are shown in Table 1.

The offensive score will be the sum of all vulgarities. Based on that result, we propose to divide it into six levels as follows:

– Level 0 (0–3 points): The sentence is not offensive.
– Level 1 (4–7 points): The sentences do not use disparaging words. A lot of repetition can go up to level 2.
– Level 2 (8–11 points): The sentences are intended to offend others and need to conduct warnings and sanctions.
– Level 3 (12–15 points): The sentences contain vulgar words and need punishment.
– Level 4 (16–19 points): The sentences contain highly offensive words and need a strong punishment

Table 1. The normative table determines the inappropriate sentence.

Word meaning	Example	Lowest mark	Highest mark	Condition
Offensive word	***	14		
Cursing and criticizing words	Stupid	7	14	
Pronouns	Father, mother	2	2	
Animal	Dog, cat	2	2	
Scold	Shut up, go away	3	9	
Body word	Eyes, nose, mouth	1	2	Cursing word or animal
Comparing word	Like, as	2	2	Mark of words is more than 3
Negative	Having negative meaning	2	2	When going with meaningful compliments
Canoe words are not suitable	Damn, fuck	4	8	
Curse	Die, go away	7	7	Going with the pronouns

– Level 5 ($\geq$20 points): The sentence is full of unacceptable offensive words and need deterrent to be an example.

Through the step, we have identified the objectionable as well as the objectionable level of the separate comments. We then can give appropriate handling measures as well as warnings.

2.2 Database Design

Social networks are written by many languages. Therefore, it needs to be able to use for all languages and libraries. Besides, people are intelligent and know how to circumvent the law by different ways of speaking to express objection without violating. Therefore, it is necessary to constantly update and expand in order to ensure the effectiveness of chatbot.

There will be two factors required to ensure:

– The amount of Vietnamese words must be large in order not to lead to confusion,

– The program is able to facilitate frequent updates without cumbersome manipulations.

The first element is a very difficult task. With twelve vowels and seventeen consonants, the number of words is an extremely large number that is hard to enumerate. We can only constantly update and improve over time with the increasing number of comments. Therefore, we focus on the second one. It is be updated regularly without maintaining each time.

We perform the step based on the existing database as shown in Fig. 3 where:

– *Diem_minanddiem_max* are minimum and maximum points for each group,
– *Diem_hien_tai* is score of each meaningful group,
– *Dieu_kien* is condition of meaning group,
– *Tu_chui_tuc* is the word that reacts to group.

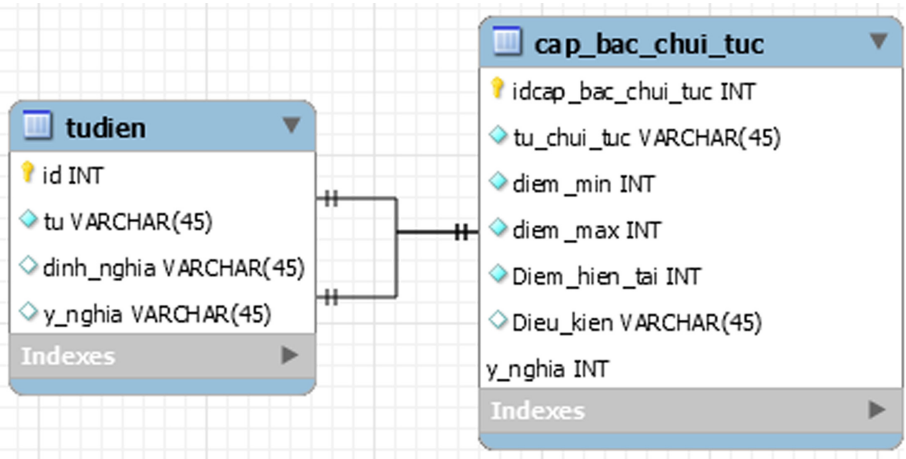

Fig. 3. Database for determining comment level.

Based on the dictionary table, sentences can be divided into meaningful words and phrases. As a result, we update the current score into table corresponding to meaning groups. Therefore, we are able to determine the offensive level of sentence. Answer and level can be used for different penalties. Therefore, the database presenting for penalty function will consist of three tables with two $1-n$ relations as shown in Fig. 4 where:

– *thoi_gian_phat* is time to punish each punishment level per minute,
– *bot_dap* is answer of bot,
– *isbot*, *istuc*, *ischui*, *iscoquan* etc. are existence of factors to check whether the sentence is offensive.

As such, we have dataized the ranking and how the chatbot responds. Depending on user, it is possible to adjust according to them.

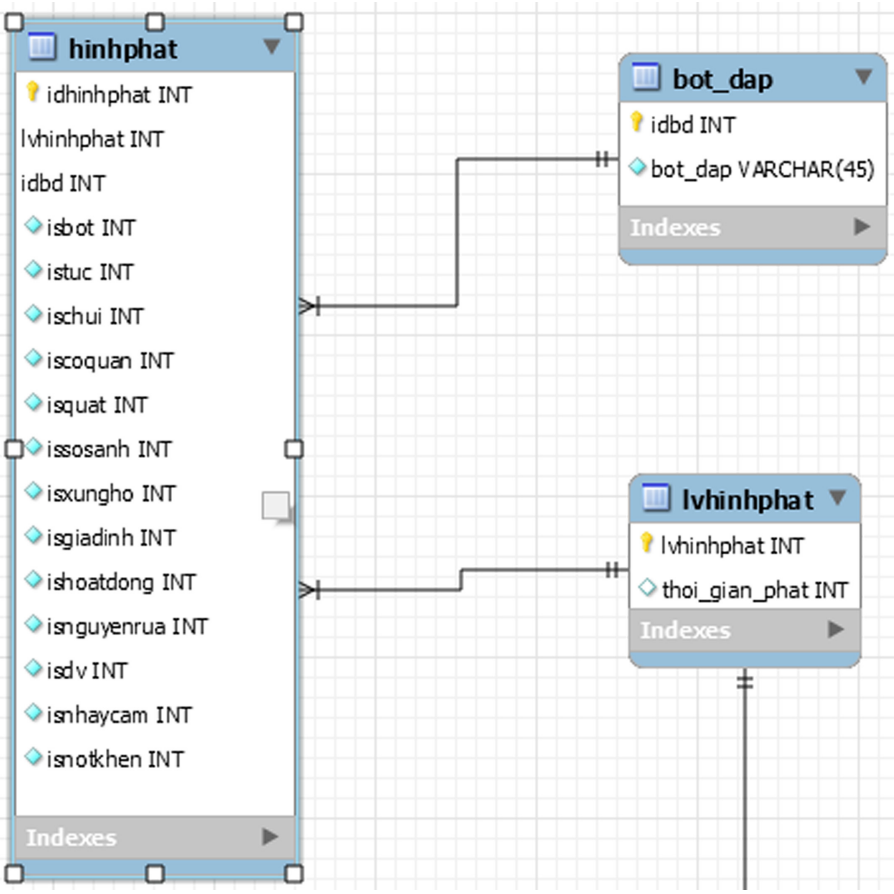

Fig. 4. Database for penalty function.

3 Simulation and Results

We perform for 100,000 comments based on [5]. The training process is as follows:

- Data input includes 100,000 comments taking from articles on Facebook and stored as .xlsx files. We divide into five types (4254 comments are level 1; 2627 comments are level 2; 1574 comments are level 3; 1021 comments are level 4; and 735 comments are level 5).
- Programming language uses php.
- Database system uses my Sql.
- Training tool uses php programming language that will read comments from input data. It then runs through the chatbot program to proceed to get the output data and save it.
- We use Intel Core i7 with 8192 Ram, Inspiron 3543.

– Data output is 5 files corresponding to offensive ranks. The program will divide the comments into offensive ranks and save them to each respective file.

Since we are not able to screen all 100,000 questions, we will conduct an assessment based on the results that are obtained for each level. The results are shown in Table 2.

Table 2. Training statistics and evaluation results.

Number of comments	Expecting processing time (Max)	Number of offensive comments	Exact number of ranks among offensive comments	Correct detection rate	Actual processing time
100,000	80 h	10,211	7696	75.36%	115 h

Detailed results with each level of comments are shown in Table 3. The correct detection rate is estimated by

$$correct\,detection\,rate = \frac{the\,correct\,number\,of\,comments}{total\,number\,of\,comments}. \quad (1)$$

Table 3. Statistics of the results obtained for each level.

Level	Number of comments	Correct detection rate (%)
1	4254	63.3
2	2627	77.54
3	1574	80.74
4	1021	94
5	735	100

In Table 3, we can see that the results are not high with 75% accuracy. The reasons are as follows:

– There are many sentences that do not have offensive meanings but still have offensive words.
– Many words in abbreviations are ignored and cannot be detected.
– The processing speed is still low. The maximum processing time is about 83 h (about 3.5 days) with 100,000 comments.

To solve the problems, it is necessary to improve as follows:

- Reseting comment level splits to be even stricter that helps to cover bad cases.
- Optimizing code, reduced processing time to an appropriate level. Maximum processing for comment is 30 s.
- Keeping to update dictionary in order to get the best accuracy from non-meaning words.
- Applying machine learning and AI to chatbot based on [6,7,10].

Therefore, we can develop the program that can be applied for practice.

4 Conclusion

Today, popular games have several ways to mask inappropriate comments. However, they have not put any effort into the issue. Most of those programs are based on specific words to identify and their effect is not great. Therefore, we propose the chatbot program to manage this comment based on:

- using antonyms.
- separating words with spaces or punctuation marks.
- using alternative words.

Besides, we set a standard to be able to determine the level of comment since optimal treatment can be taken. In the future, we will integrate new algorithms to improve the accuracy based on artificial intelligence (AI) [1,4,8,9].

References

1. Albayrak, N., Özdemir, A., Zeydan, E.: An overview of artificial intelligence based chatbots and an example chatbot application. In: 2018 26th Signal Processing and Communications Applications Conference (SIU), pp. 1–4 (2018)
2. Bakar, J.A., Omar, K., Nasrudin, M.F., Murah, M.Z.: Tokenizer for the Malay language using pattern matching. In: 2014 14th International Conference on Intelligent Systems Design and Applications, pp. 140–144 (2014)
3. Barcala, F.M., Vilares, J., Alonso, M.A., Grana, J., Vilares, M.: Tokenization and proper noun recognition for information retrieval. In: Proceedings. 13th International Workshop on Database and Expert Systems Applications, pp. 246–250 (2002)
4. Bozic, J., Tazl, O.A., Wotawa, F.: Chatbot testing using AI planning. In: 2019 IEEE International Conference On Artificial Intelligence Testing (AITest), pp. 37–44 (2019)
5. Burtsev, M., et al.: DeepPavlov: Open-source library for dialogue systems, July 2018
6. Chen, Y.N., Asli, C., Hakkani-Tur, D.: Deep learning for dialogue systems, pp. 8–14, January 2017
7. van Deemter, K., Krahmer, E., Theune, M.: Plan-based vs. template-based NLG: a false opposition?, August 1999
8. du Preez, S.J., Lall, M., Sinha, S.: An intelligent web-based voice chat bot. In: IEEE EUROCON 2009, pp. 386–391 (2009)

9. Khin, N.N., Soe, K.M.: University chatbot using artificial intelligence markup language. In: 2020 IEEE Conference on Computer Applications (ICCA), pp. 1–5 (2020)
10. Klüwer, T.: From Chatbots to Dialogue Systems, pp. 1–22, July 2011
11. Liu, B., Zhang, T., Han, F.X., Niu, D., Lai, K., Xu, Y.: Matching natural language sentences with hierarchical sentence factorization. In: Proceedings of the 2018 World Wide Web Conference, pp. 1237–1246, WWW 2018, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018)
12. Mauldin, M.: Chatbot (2020). https://en.wikipedia.org/wiki/Chatbot. Accessed 11 Dec 2020
13. Nguyen, C.T., Nguyen, T.K., Phan, X.H., Nguyen, L.M., Ha, Q.T.: Vietnamese word segmentation with CRFs and SVMs: an investigation. In: Proceedings of the 20th Pacific Asia Conference on Language, Information and Computation, pp. 215–222. Tsinghua University Press, Huazhong Normal University, Wuhan, November 2006
14. Nguyen, T., Le, A.: A hybrid approach to Vietnamese word segmentation. In: 2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pp. 114–119 (2016)
15. Phillips, S.: A brief history of Facebook. The Guardian, January 2007
16. Hông Phuong, L., Thi Minh Huyên, N., Roussanaly, A., Vinh, H.T.: A hybrid approach to word segmentation of Vietnamese texts. In: Martín-Vide, C., Otto, F., Fernau, H. (eds.) LATA 2008. LNCS, vol. 5196, pp. 240–249. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88282-4_23
17. Zhong, M., Liu, L., Lu, R.: Shallow parsing based on maximum matching method and scoring model. In: 2008 3rd International Conference on Innovative Computing Information and Control, pp. 408–408 (2008)
18. Zuckerberg, M.: Facebook (2020). https://www.facebook.com/. Accessed 11 Dec 2020