THE CLASSIFICATION OF DRUG ADDICTS' MESSAGES IN SOCIAL NETWORKS WITH NEURAL NETWORKS

Firsanova Viktoriya Igorevna (vifirsanova@gmail.com)

National Research University Higher School of Economics, Saint Petersburg, Russia

The study presents a pilot AI research. Using VKontakte API, we have collected a training dataset for a supervised learning text classification model, trained two neural network models, Bag-of-Words and Word Embeddings, with a convolutional neural network, using two different types of word vector representations (one-hot encoding and word frequency indexing). We have compared their results and defined the most effective modification. Our program defines automatically if a given text was published in a drug addicts' online community or not, with 96% accuracy. The model recognizes texts published in drug addicts' communities with an accuracy of about 92% and texts published in non-addicts' communities with about 99%.

Keywords: Classification; Bag-of-words; Embeddings; CNN (Convolutional Neural Network)

Introduction

In this study, we analyze publications from drug addicts' online communities and focus on training a supervised learning text classification model in Python. The goal of our program is to automatically classify publications from a social network VKontakte into one of two defined categories: (1) *This text is published in a drug addicts' online community*, (2) *This text is published in a non-drug addicts' online community*. During the experiments in Google Colaboratory, we have trained two models for natural language processing, Bag-of-Words and Word Embeddings, with a convolutional neural network, using two different types of word vector representations (one-hot encoding and word frequency indexing). Then we have compared their results and developed the best modification for the drug addicts' messages classification with 96% accuracy.

With our program, we have also analyzed different types of publications, for example, descriptions of mental or psychological conditions, advice or help requests, life stories. It was important for us to identify, whether the program would be able to distinguish those messages that do not contain the most explicit characteristics of addicts' speech, slang and obscene language. After a series of tests, we have found that our program could implement the analysis of much deeper lexical features. For example, it classifies well those messages by addicts, in which they complain about their psychological and physical condition.

The professional significance of the study is that it presents a pilot AI research in the field of psycholinguistics of altered states of consciousness. The results of the study can be implemented in the field of psychology. We also hope that our AI program will be used as a basis for our future projects. The paper highlights the acute social issue, however, its result cannot be used as a diagnostic tool and its further practical implementation should be firstly discussed with experts.

Literature Review

In 1983 Dmitry Spivak introduced a new term «the linguistics of altered states of consciousness». The term refers to a new field of study that combines psycholinguistics and neuroscience. The purpose of this subfield is to reveal and analyze changes in the speech that take place during the alteration of consciousness. Spivak called an altered not only a mental state produced by a psychoactive substance, but also a state of a human living or working in adverse conditions, for example, in the highlands, and any particular extreme emotional state, for example, anxiety (Spivak, 1983).

Our paper presents a study in the field of linguistics of altered states of consciousness, as well. According to the topic of our data and the contents of messages from our collection, most of the authors of the publications could consume drugs systematically, regularly. All the data from our research is anonymous, so we cannot present any statistical evidence, however, we can give some examples from our database (the examples present original punctuation and orthography):

(1) *Sejchas kurju chasten'ko , no ne bolee togo. (Now I'm smoking quite often, but no more than that.)*
(2) *Mne 22 goda iz nih ja shest' let upotrebljaju narkotiki (I'm 22, and for the last six years I'm consuming drugs)*

In general, psychoactive substances, including alcohol, injure certain areas of the brain. This applies also to centers responsible for speech. In some cases, natural resources of the brain are not sufficient to compensate for the work of these damaged centers (Lurija, 1976; Jakobson, 1974). This fact partly explains widespread speech defects among addicted people (Collins, 1980). We suggest that a speaker's speech production problems might be reflected in his or her internet discourse, as relatively new online communication has a lot in common with live interaction.

The results of studies in the field of the linguistics of altered states of consciousness can be used in psychotherapy. For example, the analysis of speech and texts produced by addicts or people affected by psychoactive substances may shed light on patients' perception of the world. This might further help to identify sources of their problems during the therapy.

Consider the following example. Stigmatization manifestations in the form of formulas («*there are no former drug addicts*») are especially popular among addicts. During psychotherapy, it is important to analyze the patient's speech semantics, to make it possible to implement the destigmatization, to debunk widespread myths about the addiction. This procedure is an important step, which allows establishing a trusting relationship with a therapist (Shajdukova, 2013).

Donald Spence describes a similar idea comparing psychotherapy with linguistic research. To find out as many details about a patient as possible, a good specialist should become an expert on personal patient's language. The decryption of each patient's statement helps to get all its possible meanings. For example, in a sentence «*I'm scared to death*» a specialist should first of all pay attention to the word «*death*». This word is like a key to a patient's attitude towards an issue (Spence, 2013).

Methods

The research continues a work launched in 2018. Our paper published in 2019 describes some features of the jargon, which drug addicts use in social networks. As a part of the work, we have compiled a list of VKontakte public pages. We have analyzed publications from these communities and concluded that most of them could be created by drug addicts (Firsanova, 2019). In this study, we have used the list to collect a dataset for the neural classification.

The list contains web-addresses of 31 public pages. We have divided all of the communities mentioned in the list into three groups. The first one includes 11 pages of so-called trip reports libraries. A publication that describes one's drug experience is called a trip report. Such texts are usually relatively large, about 500 words. The second group includes 14 thematic communities, in which subscribers can share the information and ask for advice or help. The last group includes 6 pages of entertaining content related to the peculiarities of addicts' life. Subscribers can leave in such communities short textual commentaries up to 10 words on images or videos.

We have also collected some data from 43 pages, which are not related to the addiction issues but where people use vernacular and obscene vocabulary as much as addicts. This language feature was rather important, as the training data should be more or less uniform. Otherwise, the neural network model will assign all the publications, which contain slang or vernaculars, to class 1. Among such pages are 23 communities of residential areas and 20 groups where people discuss their private life.

We have collected 23 983 publications from the pages of both classes. Our dataset contains 1 636 998 words, the collection of texts found in non-addicts' communities contains 857 195 words (52% of the data) and the collection of texts found in addicts' groups contains 779 803 words (48% of the data) (Image 1).

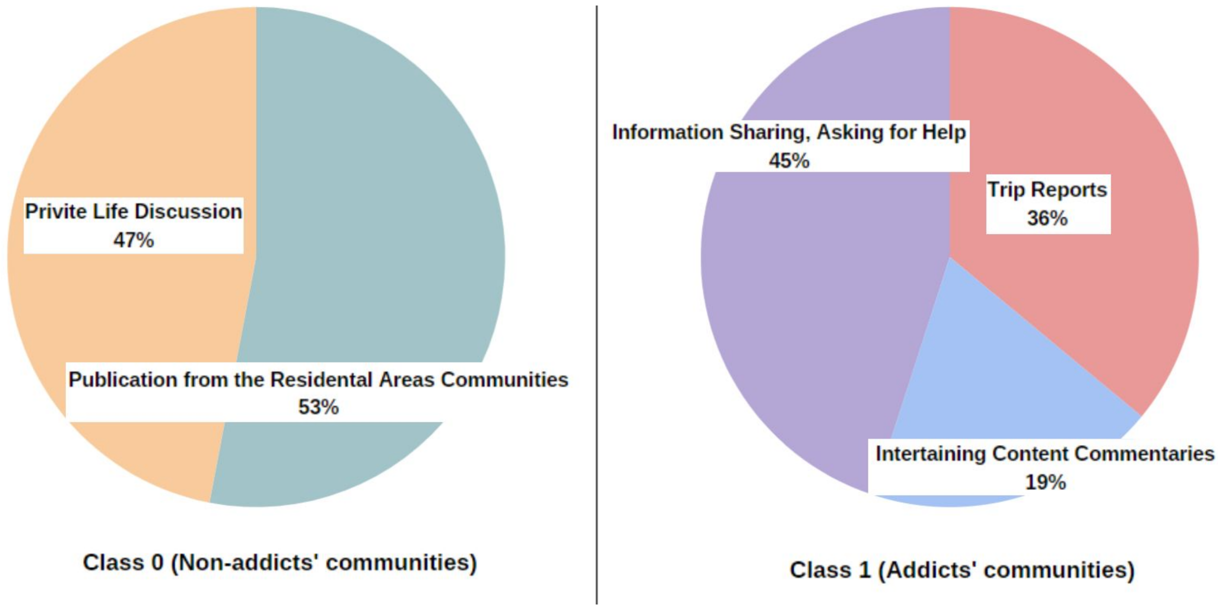

Image 1. Types of publications in the dataset

With the help of the official VKontakte application programming interface (API), we have built a bot on Python to download all the available publications from the groups and get a full dataset. API is a tool that makes it possible to implement computational interaction between two systems (Lauret, 2019).

One of the problems of natural language processing is that a neural network cannot analyze raw letters. One of the solutions is to turn a word or symbol into a digital vector (Webster, Kit, 1992). We have defined a function, which transforms text data into a digital matrix. To implement this, we have created a frequency dictionary of our data and then tokenized all the words from the array into frequency vectors. First of all, we have used a module from the Keras library for the word tokenization. We filtered the symbols (our program analyzed only Cyrillic and Latin letters) and converted all the words to the lowercase so that our data became uniform. Then we have created a frequency dictionary of our training data and defined a frequency index to each token in the dictionary according to its sequence number. Then we have transformed each index into a one-hot vector and transformed them into a matrix.

To avoid the analysis of infrequent words and typos, we have defined a maximum amount of words from the frequency dictionary of our training data, which our neural network would analyze during the training. That was our *maxWordCount (mWC)* parameter, which was equal to 20 000 at first and to 20 500 after the parameter optimization. Our frequency dictionary contains 137 802 wordforms, however, if a given wordform from the input data is not recorded in the dictionary or its frequency index (its sequence number in the dictionary) is greater than mWC value, then it will be ignored by the program.

We split the data into two samples, a training set (80% of the collected data) and a validation set (20% of the data), and we also collected some new publications for our test set to evaluate our models. We split all the data in the training and validation sets into passages of equal length, so that our neural network could analyze short pieces of the text and not the whole array. During the tokenization, we gave a frequency index to each word in a sample. Each passage was transformed into a frequency index sequence. We transformed all the sequences into a one-hot encoding matrix, in which one vector represents one index, to implement the training of the Bag-of-Words model (Table 1). The volume of our training data is 8 998 819 symbols, 1 221 630 words.

Original text	Frequency index sequence (used for Word Embeddings model training)	One-hot encoding (used for Bag-of-Words model training)
TV-3 priglashaet sem'i iz Cheljabinska prinjat' uchastie… (TV-3 invites families from Chelyabinsk to take part in...)	[2 441, 4 228, 747, 23, 335, 1 324, 1 860]	[0. 1. 1. ... 0. 0. 0.]

Table 1. An example of the vectorization process

We have trained two classification models based on a convolutional neural network with a sequential layer (Lewis, Gale, 1994). We have used the Keras library. According to some research, this neural network is quite efficient in text classification (Lee, Dernoncourt, 2016). The first model, Bag-of-Words (BoW-model), compute words usage frequency (Zhang, Jin, Zhou, 2010), but do not take into account sentence structure or word order, however, we assume that this model might analyze valencies filling, as we did not implement the lemmatization. The second one, Word Embeddings (E-model), matches word vectors (Sriram, Fuhry, Demir, Ferhatosmanoglu, Demirbas, 2010). This model can implement deeper lexical analysis and define semantic clusters. The last model considered to be a more effective tool for natural language processing (Turney, Pantel, 2010). That is why we have made a hypothesis that our E-model will classify texts with higher accuracy than our BoW-model. For both models, we have used Adam optimization algorithm from the Keras library, categorical cross-entropy loss function, and accuracy metrics for the model evaluation. We have calculated the accuracy metrics as a percentage of correct answers to the number of the total passages. We have trained both models for 20 epochs with the batch size equal to 200 (Table 2).

Bag-of-Words parameters	Words Embeddings parameters
modelB = Sequential() modelB.add(BatchNormalization()) modelB.add(Dense(200, input_dim=maxWordsCount, activation='relu')) modelB.add(Dropout(0.5)) modelB.add(BatchNormalization()) modelB.add(Dense(2, activation='softmax'))	modelE = Sequential() modelE.add(Embedding(maxWordsCount, 30, input_length=xLen)) modelE.add(SpatialDropout1D(0.25)) modelE.add(Flatten()) modelE.add(BatchNormalization()) modelE.add(Dense(200, activation='relu')) modelE.add(Dropout(0.25))

	modelE.add(BatchNormalization()) modelE.add(Dense(2, activation='softmax'))

Table 2. Model parameters

We have conducted a series of experiments with optimization of the following model parameters: the length of an analyzed passage in a sample (xLen), the offset (step), and the number of words from the frequency list, which our neural network has analyzed during the training (maxWordCount, mWC). According to the accuracy metrics, the Bag-of-Words model showed better results (Table 3). That means that our hypothesis did not confirm. In the most efficient modification of our model, the value of xLen is 50, the step is 50 and the value of mWC is 20 500, which is greater than the initial one, 20 000. We called this modification BoW 50x50 + mWC (Image 2).

Bag-of-Words model accuracy	Words Embeddings model accuracy
Messages from non-addicts' groups: 99% Messages from addicts' groups: 92% Average: 96%	Messages from non-addicts' groups: 97% Messages from addicts' groups: 92% Average: 95%

Table 3. The comparison of class recognition accuracy

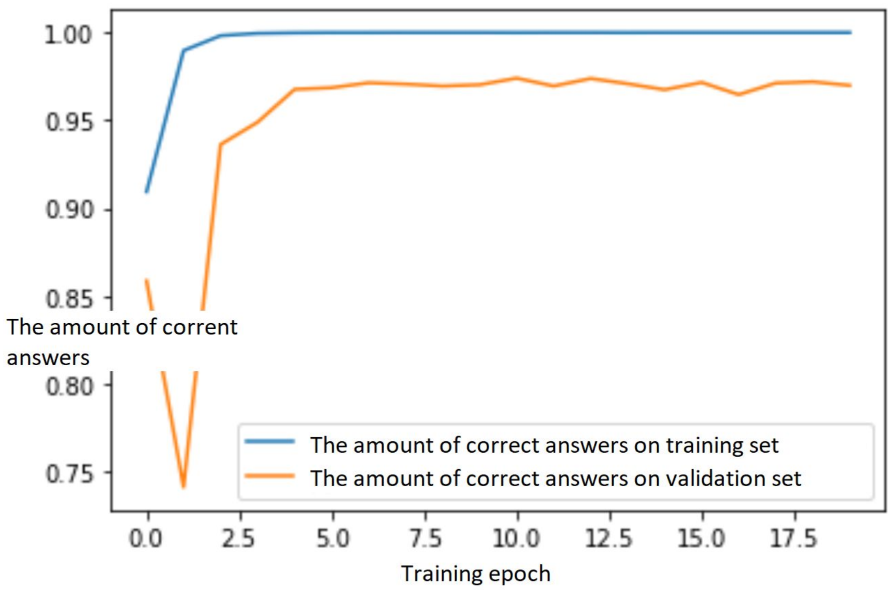

Image 2. The training process of BoW 50x50 + mWC

During the training, the goal of the neural network was to compute weights. The weights are the values of the connection strength between text characteristics and given classes (Ng, Hun, 2014). We have saved this data and the model parameters. This step was necessary for the reuse of the program. We have analyzed some publications from the public pages, which we have used earlier to collect the training data. These texts did not appear in the training set.

Out AI model recognizes drug addicts' publications with high accuracy (96%). Notwithstanding the presence or not of such explicit features as specific slang or vernacular language, the program classifies mostly correctly. We have found that most of the publications from addicts' online communities contain help requests or complaints about psychological or physiological problems. We believe that our program can become a tool for the automatic detection of addicts' communities. For example, this might help experts to contact with addicts and offer them psychological consultation.

Input:	narod! esli tema byla sorjan. kto-to uznaval v zhjek chto s vodoj? pochemu gorjachaja takogo mutnogo cveta? (guys! sorry if the topic was already [mentioned]. did someone find out in the housing office what's wrong with the water? why the hot [water] is so muddy in color?)
Prediction:	[0.9182049 0.08179515]
Class:	0
Output:	Artificial Intelligence Verdict: This post was most likely published in a non-addicts' online community.

Table 4. The example of the program output

Overall, we have developed the most effective modification of our neural model after experiments with parameter optimization. We have confirmed that artificial intelligence can deal with the recognition of language characteristics of a social category. We believe that our program will become a starting point for our further experiments with AI.

Conclusion

All in all, after training a neural network on a large amount of text data and experimenting with model parameters, we have created an efficient classification tool, which can filter the information from social networks. This result opens up several future directions. For example, we can design a program that will classify public pages in social networks and define automatically whether a given public page presents a drug addicts' community or not.

We have trained two AI models, Bag-of-Words and Word Embeddings. During our experiments in Google Colaboratory, we have developed the most effective model modification according to the accuracy metrics. We made a hypothesis that an E-model (Word Embeddings) will be more efficient than a BoW-model (Bag-of-Words). This hypothesis did not confirm.

BoW 50x50 + mWC, Bag-of-Words model, in which xLen is 50, step is 50, and mWC is 20 500, showed the highest accuracy. We have used this model to analyze publications, which did not appear in our training set, and found that artificial intelligence recognizes correctly publications, in which people complain about their health condition and mental state.

Drug addicts need to share their experiences. They also regret their past. Instead of consulting with medical experts and psychologists, they leave messages for unknown people online. Our AI classification allows finding such people and such communities automatically, so we plan to continue our research in the field of linguistics of altered states of consciousness.

References

1. Collins, Patrick J. “A Comparison of the Oral Syntactic Performance of Alcoholic and Non-Alcoholic Adults.” *Language and Speech*, vol. 3, 1980, pp. 281–288.
2. Firsanova, Viktoriya Igorevna. “Funkcionirovanie zhargona narkozavisimyh v social'nyh setjah (The Functioning of the Drug Addicts’ Jargon in Social Networks).” *Politika i kul'tura v sovremennom mire (The Politics and Culture in the Modern World)*, 2019, pp. 185-195.
3. Jakobson, Roman Osipovic. *Main Trends in the Science of Language.* Harper and Row, 1974.
4. Lauret, Arnaud. *The Design of Web APIs*. Manning Publishings, 2019, 400 p.
5. Lee, Ji Y., Dernoncourt, Franck. “Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks.” *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* vol. 16, no. 1, 2016, pp. 515–520.
6. Lewis, David D., Gale, William A. A “Sequential Algorithm for Training Text Classifiers.” *SIGIR ’94*, Springer, London, 1994, pp 3-12.
7. Lurija, Aleksandr Romanovic. *Basic Problems of Neurolinguistics.* Mouton, 1976.
8. Ng, Thomas S., Hung, Us, S. *Safety, Reliability and Applications of Emerging Intelligent Control Technologies*. Elsevier, 2014.
9. Shajdukova, Lejla Kazbekovna. “Sovremennye podhody k reabilitacii narkozavisimyh (Modern Approaches to the Drug Addicts' Rehabilitation).” *Kazanskij medicinskij zhurnal (Kazan Medical Journal)*, vol. 94, no. 3, 2013, pp. 402-405.
10. Spence, Donald P. “Language in psychotherapy”. *Psycholinguistic Research (PLE: Psycholinguistics): Implications and Applications,* 2013, 471 p.
11. Spivak, Dmitry Leonidovich. “Lingvisticheskaja tipologija iskusstvenno vyzyvaemyh sostojanij izmenennogo soznanija (The Linguistic Typology of Artificially Caused Altered States of Consciousness. Message I).” *Fiziologija cheloveka (Human Physiology)*, no. 1, 1983, pp. 141–146.
12. Sriram, Bharath, Fuhry, David, Demir, Engin, Ferhatosmanoglu, Hakan, Demirbas, Murat. “Short text classification in twitter to improve information filtering.” SIGIR '10: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, pp. 841-842.
13. Turney, Peter D., Pantel, Patrick. “From frequency to meaning: Vector space models of semantics.” *Journal of artificial intelligence research*, vol. 37, 2010, pp. 141-188.
14. “VK API Documentation.” *VK Developers*, VK, vk.com/dev/manuals (request date: 30.04.2020).
15. Webster, Jonathan J., Kit, Chunyu. “Tokenization as the initial phase in NLP.” *COLING '92: Proceedings of the 14th conference on Computational linguistics,* vol. 4, 1992, pp. 1106-1110.
16. Zhang, Yin, Jin, Rong, Zhou, Zhi-Hua. “Understanding bag-of-words model: a statistical framework.” *International Journal of Machine Learning and Cybernetics*, vol. 1, 2010, pp. 43-52.