

Development of a Model to Predict Intention Using Deep Learning

Nikolay Karpov, Alexander Demidovskij, and Alexey Malafeev

National Research University Higher School of Economics, Russia,
nkarpov@hse.ru, monadv@yandex.ru, amalafeev@yandex.ru
Home page: <https://www.hse.ru/en/staff/nkarpov>

Abstract. This paper presents a method to analyze discussions from social network by using deep learning. We have prepared a new dataset by collecting discussions from a social network and annotating remarks of the discussion. The annotation consists of two types of labels for each message: intention type and direction of intention. Using this dataset and pre-trained word embeddings we have evaluated two neural network structures. On the basis of evaluation, we chose a model to automatically predict intention types and direction of intention of an arbitrary message from any social network.

Keywords: natural language processing, intention analysis, deep learning

1 Introduction

There is a currently growing interest in social network analysis due to the expanded role of the latter. The fundamental trend is that people are mostly communicating and collaborating inside these social networks. As the majority of people now have at least one account at a social network and people communicate there via exchanging text messages, it seems to be extremely important to be able to analyze this type of data and reveal its hidden properties.

One of the most popular formats of communication in a social network is a phenomenon of a post. It is an arbitrary message, expressing thoughts and ideas of a speaking person. Such a post usually appeals to people's emotions and the audience starts to actively discuss it by putting more and more comments. There is a significant peculiarity of such discussions that the topic of the discussion usually changes very fast and at the end is no longer connected with the subject of the source post.

Why is it so important to predict the intention of the given text? The idea of manipulating the discussion and the message it brings is currently an active area of research in the field of political linguistics. Indeed, it is extremely important to make sure that the dialog of the candidate for the place of the President with the audience makes the right message and manipulation. Taking into account the variability of discussion topic described earlier, it is vital to make sure that the discussion still stays on the necessary path for the author of the post theme.

From our point of view, the first step in solving this task is to be capable of predicting the intention of the speaker automatically.

At the same time, an increasing amount of publications in the sphere of Internet texts analysis reveals quite interesting properties of modern texts. In particular, every speech act has its own intention - the aim and will to express any idea. In addition to it, each intention has its own direction, which means that any phrase of the social network user can be directed towards the author, himself etc. More importantly, due to a huge amount of these texts, it is quite easy to collect a large database which is vital for modern means of data analysis.

While the debate over the intent analysis seems to gain popularity, there is a gap in absence of mathematical models and applied instruments for automated prediction of the intention. Although there are existing studies of making text classifiers, the task of predicting the intention of the given arbitrary text stays unaccomplished.

We consider a building machine learning algorithm to predict intentions in social network. Our main contribution is the following:

1. Make specific dataset with each remark of the discussion annotated;
2. Annotation consists of two types of labels for each message: intention and direction of intention;
3. Successfully apply machine learning algorithm to predict intention and direction of intention;

The remaining part of the current study is the following. In Section 2 we will give the detailed overview of the existing approaches and significant theoretical trends. Further, in Section 3 we will describe the experiment methodology we elaborated to perform experiments. The detailed overview of the deep learning architecture will be given in the Section 4. Results of the proposed approach will be shown in the Section 5. Finally, we will make a conclusion and further research directions analysis in the Section 6.

2 Related Works

A foundation of our research is dual. First, it is based on the research in a psychology of communication. Second, we use modern methods for automatic natural language processing.

Let us start from psychology. Early works go back to at least sixties of last century when J. Austin and J. Searle [1, 10] made the theory of speech acts. J. Austin classifies speech acts into three types: locutionary, illocutionary and perlocutionary acts. Our research deals only with illocutionary acts in the field of political discourse. Like some other researchers [14] we use a finite set of intention types. Oleshkov M.Yu. proposed to typify not a speech acts, but to typify a communicative strategy as a general goal of speech act [8]. We use his set of intention types with additions, proposed by Radina N.K. [2]

The task to automatically identify a type of communicative strategy from the finite set of types can be formulated as a classification problem. A classification problem of short messages is a well-known problem in the natural language

processing field. This problem is traditionally solved by using machine learning approaches. For instance, sentences can be classified according to their readability using prebuilt features and SVM, Random Forest, Classification and others [4]. Short messages from social networks can be classified according to its sentiment polarity [5].

A recent success of neural networks application showed that transfer learning allows to improve classical machine learning methods. [12, 3]. Essentially it is important when training dataset is too small. That is why in our study we decide to use neural networks as a classifier.

3 Experiment Methodology

3.1 Dataset Creation

One of the key elements of our research was a creation of the appropriate dataset. It was decided to use the most popular Russian social network - VKontakte¹ as the source of discussions and texts. Raw data was downloaded using VKMiner program². However, we performed rigorous filtering of these discussions. In particular, all the discussions with the phrase quantity less than 40 were rejected as well as meaningless comments, e.g. empty messages or photos instead of texts.

There is no doubt that the expert labeling of each text in accordance with the intention class that it represents is important [8]. Moreover, as we have already mentioned, it is vital to also mark the direction of the intention[2]. That is why during the labeling process, experts used both the letter, which represents the intention type, and the digit, which represents the direction of the given intention. Such a labeling shifts focus from traditional methodology of intention analysis to the hybrid one, which enables the opportunity to classify the text very precisely from an intentional point of view.

The way how the given dataset was used in building the automatic intention classifier will be described in the next section.

3.2 Classification Settings

To predict intentions automatically from a finite set of intentions we should create a classifier of a user message. Oleshkov M.Yu. proposed 25 intention types [8]. Radina N.K. proposed to group these intention types into 5 supertypes [2] according to Habermas as shown in Table 2. Additionally we have 4 directions of the intention shown in Table 1. Therefore, we have the following classes:

- $25 \times 4 = 100$ intention types and directions;
- 25 intention types;
- 5 intention supertypes;
- 4 directions of the intention;

¹ <http://vk.com>

² <https://linis.hse.ru/soft-linis>

As the classifier, we use two types of neural networks with traditional architecture for text classification task. To evaluate results of classifier we use precision, recall, and f-measure score.

Table 1: Intentions classification by direction

No.	Meaning	Example
1	"I"/"We" - the comment author about themselves and "us"	I disagree I think We can together etc.
2	"They" - author(s) of the article/post	They are right They are not right They do wrong etc.
3	"They" - other comment authors	They do not understand They talk nonsense etc.
4	Someone/Something in general, not "I"/"We", not "They"	Russia the USA capitalists etc.

Table 2: Intention types and supertypes

Information reproducing type, To Reproduce the observable in speech	Emotionally consolidating type, Suggestion of the own world picture for cooperative collaboration	Manipulating type, Social domination, hierarchy establishment	Order-directive type, To encourage addressee to an action, make changes in the fragment of reality	Control-reactive type, To express evaluative reaction regarding the situation
A Surprise Question	F Selfpresentation	K Abusement	P Inducement to a positive action/ recomendation	U Acceptance Accolade
B Showing disagreement Hesitation	G Attention attraction (discourse, rhetorical questions)	L Frightening, threats	Q Solicitation to negative	V Sarkasm Malevolence
C Agreement expression Support	H Auditorium aussuagement Reassurance	M Discredit (authority disruption)	R Accusation	W Criticism
D Non-acceptance, rejection from communication	I Forecasts, claims for truth	N Force demonstration (without obvious threat)	S Consequencies caution	X Irony
E Commiseration, sympathy	J Justification (as self-justification, e.g. without accusation)	O Moralisation, homily	T Accusation offset (if is accused)	Y Exposure

4 Deep Learning Architecture

In this section, we briefly describe our choice of architecture, a regularization method and a training algorithm.

4.1 Embeddings

The idea of representing each word in a text or whole texts as vectors have come to occupy the central place in modern methodologies of text analysis. Such vectors, which represent words, are called embeddings, and can be easily trained with word2vec, Glove etc. [7, 9]. In general, these vectors can be trained on the given dataset, however, the dataset needs to be quite big, e.g. Russian Wikipedia, which includes 600 millions of words, can be used, to provide sensible results. Although our final dataset contains 21192 texts, there are also 100 classes, which means that there is not enough data to train our word vectors on. That is why it was decided to use the existing embeddings collection³ trained on the Ruscorpora [6].

However, exploration of the given embeddings resulted in the necessity of preprocessing each word from the source dataset in a way there is a matching vector in the embeddings collection. In particular, each word has to be in a form "infinite form + form name", e.g. "ran" should be translated to "run_Verb". This processing required the use of the automatic tool capable of performing the morphologic analysis of each word in a given text. As a part of this research, the MyStem⁴, a tool developed by Yandex, let us make necessary word forms and reuse already trained embeddings.

4.2 Convolution Layers

Convolutional neural network are state-of-the-art semantic composition models for text classification task [13]. We use three series-connected composition cells with maxpooling.

4.3 Recurrent Layers

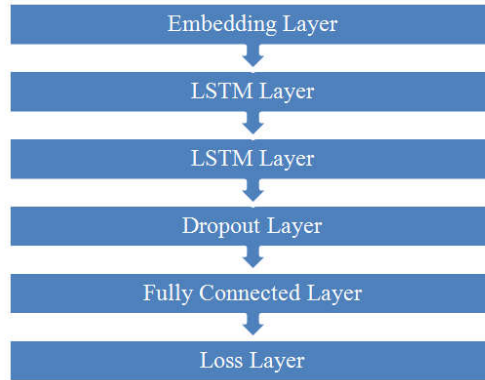
Recurrent layers are proved to be useful in handling variable length sequences [13]. We use two series-connected long short-term memory (LSTM) cells to compute continuous representations of tweets with semantic composition.

4.4 Regularization

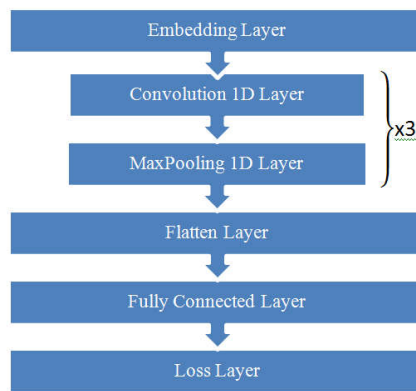
We use dropout as the regularizer to prevent our network from overfitting [11]. Our dropout layer selects a half of the hidden units at random and sets their output to zero and thus prevents co-adaptation of the features.

³ <http://rusvectors.org/en/models>

⁴ <https://tech.yandex.ru/mystem/>



(a) Structure of the LSTM network.



(b) Structure of the CNN network.

Fig. 1: Structure of the networks utilized in experiments (LSTM, CNN).

4.5 Training Algorithm

We initialize our embedding layer with the help of pre-trained vectors on the Russian National Corpus and Russian Wikipedia. Other layers in our neural networks were initialized randomly. Then we trained them on the train subsets using Adam method for stochastic optimization of an objective function.

5 Results

In the part to follow the experimental results will be shown in Table 3 as well as the detailed explanations.

The anticipated results are intended to bring capabilities to solve classification task of using two models (CNN 1b and LSTM 1a) in the foreground. Firstly, both models were used to predict the class of the intention by the given text and the result is quite poor for all models (accuracy less than 0.05). This proves the hypothesis that the dataset used is too small (21192 texts) for such a huge number of classes (100 classes). Considering this fact it was decided to continue experiments with fewer classes. The authors elaborated three strategies to overcome this obstacle.

The first strategy was to try to predict only the intention type. Instead of trying to predict one of 100 classes, we can try to predict only 25 based on the general classification of intention types - Table 2. This experiment brought us to slightly better results, although the top accuracy was still lower than 0.1. In the second strategy, we made an attempt to predict only the direction of the intention. In this case, models should have predicted one of 4 classes of intentions direction - Table 1. The validation of the second strategy hypothesis proved that the main reason of low prediction performance in previous attempts was the huge gap in a concordance of the dataset size and number of classes to predict. The final strategy was to try to predict the intention type by a supertype. Indeed, the intention classification by Habermas (Table 2) provides the way of joining intentions types in 5 superypes. The results of this experiment showed that when trying to predict one of 5 superypes, both models can correctly classify intentions of every third text.

Taken together, the interim results lead us to the conclusion that LSTM model over-performs the CNN model in almost all cases, although it usually takes much more time to train LSTM model even for several epochs and brings us to the conclusion that LSTM is much more suitable architecture for text classification tasks and quite expensive from the computational standpoint.

Finally, the creation of the first-ever dataset in Russian language labeled according to the specific intentions types and the directions is extremely important. The dataset contains 21992 items, which represent 100 classes. Also, the trained model and the source code are available publicly for usage and enhancement⁵ under the business-friendly license (MIT).

6 Conclusion

The aim of this research was to find a model, suitable to predict intentions of users which they express in discussions in social networks. We have prepared a new dataset by collecting discussions from a social network and annotating remarks of the discussion. The annotation consists of two types of labels for

⁵ <https://github.com/demid5111/intentions-analysis>

Table 3: Experimental results on CNN and LSTM in dependence with the epoch quantity

num. of classes	Intention supertypes		Directions of the intention		Intention types		Intention types and directions	
	5		4		25		100	
	LSTM CNN		LSTM CNN		LSTM CNN		LSTM CNN	
num. of epochs	5	20	5	20	5	5	100	5
precision	0.35	0.06	0.63	0.36	0.07	0.01	0.06	0.01
recall	0.35	0.24	0.65	0.60	0.15	0.10	0.05	0.07
f1-score	0.34	0.09	0.63	0.45	0.08	0.02	0.05	0.01

each message: intention and direction of intention. All discussions were dedicated to political topics. Using this dataset and retrained word embeddings we have built two models of a neural network to automatically predict an intention of an arbitrary message from any social network user. Experimental results showed that model based on LSTM allows to obtain better results. The classification by the directions of intention showed the best accuracy. We explain it not only by a low number of classes, but because of the fact that directions are often represented using explicit words.

Acknowledgments

The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2017 (grant N17-05-0007) and by the Russian Academic Excellence Project "5-100".

References

1. John Langshaw Austin. *How to do things with words*. Oxford university press, 1975.
2. Radina Nadezhda K. Intent alysis of online discussions (using examples from the internet portal inosmi.ru). *Mediascope*, (4), 2016.
3. Nikolay Karpov. Nru-hse at semeval-2017 task 4: Tweet quantification using deep learning architecture. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 681–686, Vancouver, Canada, August 2017. Association for Computational Linguistics.
4. Nikolay Karpov, Julia Baranova, and Fedor Vitugin. Single-sentence readability prediction in russian. In *International Conference on Analysis of Images, Social Networks and Texts-x000D.*, pages 91–100. Springer, 2014.
5. Nikolay Karpov, Alexander Porshnev, and Kirill Rudakov. NRU-HSE at SemEval-2016 Task 4: Comparative Analysis of Two Iterative Methods Using Quantification Library. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 171–177, San Diego, California, June 2016. Association for Computational Linguistics. bibtex: karpov-porshnev-rudakov:2016:SemEval.

6. Andrey Kutuzov and Elizaveta Kuzmenko. *WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models*, pages 155–161. Springer International Publishing, Cham, 2017.
7. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
8. M. Yu. Oleshkov. *Simulation of the communication process: monograph*. Nizhny Tagil gos.sots.-ped.akademiya (et al.), 2006.
9. Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.
10. John R Searle. *Speech acts: An essay in the philosophy of language*, volume 626. Cambridge university press, 1969.
11. Nitish Srivastava. *Improving neural networks with dropout*. PhD thesis, University of Toronto, 2013.
12. Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, and Ivica Dimitrovski. Finki at semeval-2016 task 4: Deep learning architecture for twitter sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 149–154, San Diego, California, June 2016. Association for Computational Linguistics.
13. Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *EMNLP*, pages 1422–1432, 2015.
14. Latynov V.V., Cეpcov V.A., and Alexeev K.I. Words in action: Intent-analysis of political discourse. 2000.