

## Analysis of convergence and divergence in chat conversations

**Liviu-Andrei Niță**

University Politehnica of Bucharest  
312 Splaiul Independenței,  
Bucharest, Romania  
[nitaliviu79@yahoo.com](mailto:nitaliviu79@yahoo.com)

**Ștefan Trăușan-Matu**

University Politehnica of Bucharest  
312 Splaiul Independenței,  
Bucharest, Romania  
and  
Research Institute for Artificial Intelligence  
and  
Academy of Romanian Scientists  
[stefan.trausan@upb.ro](mailto:stefan.trausan@upb.ro)

**Traian Rebedea**

University Politehnica of Bucharest  
312 Splaiul Independenței,  
Bucharest, Romania  
[traian.rebedea@upb.ro](mailto:traian.rebedea@upb.ro)

DOI: 10.37789/rochi.2020.1.1.17

### ABSTRACT

A discussion between several participants is often accompanied by an exchange of information between speakers and, in the case of collaborative learning, a polyphonic interanimation is desired. They have, according to the polyphonic model, different points of view that are convergent or divergent by which one of the participants approves or disapproves of another person that is participating in the discussion. This paper presents an approach for identifying convergent or divergent utterances using neural networks and other machine learning methods in order to help people analyze a dialogue conducted in an online environment. Especially in collaborative chats used in education, this solution allows professors to identify how the participants in the discussion exchange information and how the debate evolved in time.

**KEYWORDS:** dialogue, chat, neural networks, machine learning, convergent, divergent, polyphony.

### ACM Classification Keywords

I 2.7 Natural Language Processing. Text analysis.

### INTRODUCTION

Online conversations are exchanges of ideas in synchronous or asynchronous ways between two or more participants. A pattern observed in a significant number of discussions is the accompaniment of the exchange of information between individuals by different points of view, which are either convergent or divergent, in which a sender approves or disapproves the ideas of another person participating in the discussion. This phenomenon may be viewed as

an interanimation between ideas, similarly to the interanimation of voices in polyphonic music [1,2]:

“a suggestion for how to analyse knowledge building and debates of ideas is to consider the complex weaving of musical polyphony, where several voices/ideas are concurrently developed in time, with the consequence that some dissonances/divergences occur, but eventually a coherent whole, a discourse, is achieved.” [1]

We consider that an utterance in a conversation has a convergent character when put in relation with a message send by another participant in the discussion expresses an agreement of opinions between individuals. The process by which a convergent statement is formed is that of extracting, defining, and emphasizing the information transmitted by the original sender [3]. For example, if a person says (the examples are taken from real chats between students [4], therefore they may contain mistypes and other linguistic errors):

*“I think that blog are there to keep in touch with the sensitive part of the customer”*

and another participant responds with the message:

*“I support you on that”*

between these replies it has been established a point of convergence.

Unlike convergent statements, divergent remarks are defined by a process that generates different points of view to reach a common truth that has the effect to stimulate creativity [5,6]. For example, when a participant in a chat says:

*“I think that blog are there to keep in touch with the sensitive part of the customer...but need a lot of men power which can be a big drawback for small companies”*

and another one re:

*“a small company would need no more than two or three specialists to cope with their customers in a live chat systems”*

then we can say that a divergence exists between the two utterances.

These two types of replies are the basic components of interanimation that characterizes polyphony, a desired phenomenon arising in collaborative chats. By identifying them, different performance indicators for collaborative learning processes may be provided by learning analytics applications [1] allowing teachers to observe, develop new strategies to increase student performance, and ultimately achieving better educational dialogue by developing the ability of students to accept different points of view and increase the level of knowledge that is transmitted in chat conversations and in general [1,8].

#### STATE OF THE ART

The basic structural unit of a dialogue is the utterance (reply). In the format that dialogue is a conversation between at least two people who alternate the roles of speakers and listeners, the utterance becomes the element that transmits a quantity of information. The analysis of a conversation implies the understanding of the effect of the messages transmitted through a reply by studying the (polyphonic) structure of the dialogue, the action sequences, and the adjacent pairs within a dialogue. An adjacent pair consists of a sequence of two sentences produced by different participants in the discussion in which a first action transmitted by one of the speakers requires an answer from another.

With the appearance of the internet, the way that humans communicate completely changed. A key element for this type of communication is represented by its development in real time, the message becoming an important proportion of spontaneity. Although this brings the internet mediated conversation closer to a traditional one, which takes place face to face, there are some key elements that makes it different from a natural one. Extra verbal means of communication are no longer present in their classical form, being replaced instead by stylistic elements borrowed from classical written communication and by symbolic elements through which different emotions are

transmitted. However, online chat conversations allow the existence of more than one thread of discussions in the same time [1] which is a key factor for interanimation.

In the context of the rapid development of computers and increase of computational power, it has become feasible to perform better analysis and processing of texts using computer systems. At the beginning of artificial intelligence, attempts were made to mimic the steps used by a linguist: defining and evaluating the rules of speech and writing by using regular expressions and context-free grammars. With the increase of computational power and of the number of texts in electronic format, it became possible to use powerful statistical techniques in order to study, process, and find meaning in texts. The ambiguity of language found in texts (especially in the colloquial ones) has led to the treatment of natural language processing in a similar manner to random processes, in which the understanding of distributions of probability is essential. Advantages of using this kind of models are the capacity of a system to learn on texts that have not been studied in the past, rising of accuracy by introducing new training data, and the general scalability of the program.

We can use neural networks to build efficient classification models. They are based on the analogy with the structure and function of the human neuron, being systems with the ability to produce results based on statistics without constructing an algorithm or set of rules. Due to the computational power and the huge volume of data, by modeling a structure based on that of the human brain, the purpose of neural networks is to create much better and simpler machine learning algorithms. Neural networks thus offer the opportunity for programmers to replace existing models with some that can perform better, to model through new structures problems of natural language processing and to automatize the process of designing a program through a reduced dependency by a professional linguist. The use of neural networks in the field of natural language processing is currently intensively studied, due to the performance obtained by models that need a large volume of data but do not require linguistic expertise for proper operation [7].

#### IMPLEMENTATION

The conversations that we analyzed are saved in XML format [4], where element *Corpus* represents the whole content of the file, *Dialog* refers to the discussion in which the participants are engaged, *Participants* relates to the people who communicate, *Person* refers to one of the participants, *Topics*

represents the subjects which are discussed, *Turn* cumulates one or more utterances from one individual, and *Utterance* refers to the actual reply in the chat. Besides elements, there are attributes defined, where *genid* is a unique identifier for utterances, *time* refers to the moment when a reply was transmitted, and *ref* describes an explicit link added by a chat participant to a previous reply. This value can be either positive, the value referring to the *genid* of a previous reply, or -1, noting that the author of the current reply did not add an explicit link.

The *convergence* attribute marks the fact that the utterance is in a convergence relation to a certain reply and it can refer to more than one. The *divergence* attribute marks the fact that the sentence is in a divergence relation to a certain reply and, like convergent replies, several previous utterances may be referred to. Below we will illustrate the steps we made for preparing the training and test datasets.

**Step 1**

First, we parse the whole XML corpus, which consists of 63 annotated conversations and we save convergent or divergent utterance pairs. For this, we identify the replies have the attribute of convergence or divergence defined and the reply for which the link is formed.

**Step 2**

Then we pre-process the utterances in which expression like “*I’m*” are transformed into “*I am*”, “*We’re*” into “*We are*”. Then all uppercase letters are converted to lowercase, punctuation is removed, and words that are made up of a single letter, irrelevant to the context but also the names of participants are removed. After applying these steps, the words are lemmatized in order to bring each word to its basic form, thus reducing the size of the vocabulary and grouping together words with the same lemma.

**Step 3**

Because the length of the utterances is variable, from 5-6 words to extremely long ones that have more than 32 words as is showed in Figure 1., their size is limited to a constant value. To further reduce the vocabulary size, words that have a frequency less than a certain value are removed. During testing, this value is 10. It is important to note that very rare words usually do not provide useful information for statistical text classifiers.

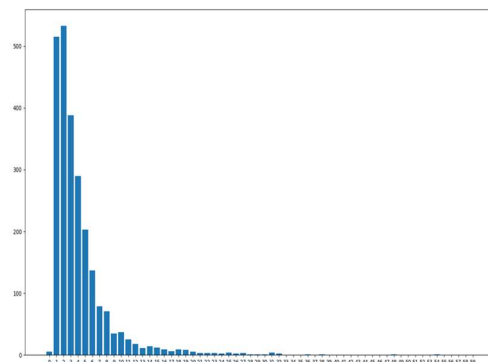


Figure 1. The number of convergent and divergent replies of a certain length

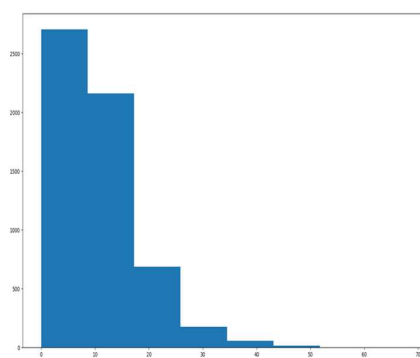


Figure 2. Distances between convergent replies and the utterance for the link is formed

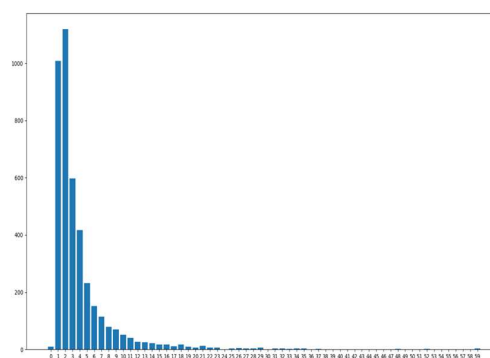


Figure 3. Distances between convergent replies and the utterance for the link is formed

In Figures 1-3 are presented relevant statistics about the utterances in our dataset, including the utterance lengths and distances between replies.

### THE 3-CNN MODEL

The method initially chosen to detect convergence and divergence relations in chat conversations is using a 3-channel neural model consisting of Embedding and Convolutional layers (see Figure 4). This model uses only the utterances that have the attribute convergent or divergent defined. Neural networks that use convolutional layers can be used to classify a statement due to their ability to identify words that are important for defining class membership. This Convolutional layer does not take into account the position of the words in the sequence received as input.

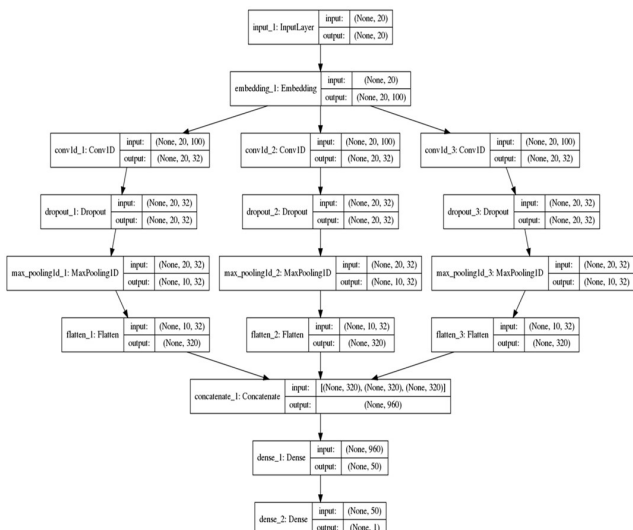


Figure 4. Neural network with 3 CNN channels

Before the convolutional layers, there is an embedding layer with the role of representing a mapping of each word in the vocabulary to a continuous vector space, with similar words being located in nearby regions. Other layers in the network have the role of regularization (Dropout) and sum up the results. We use different sizes for the kernels so we can capture different groups of words, the number of filters is set to 32, and the size of the embedding layer is the same as the length of the utterance. In Figure 4 we provide the complete representation of the architecture.

### THE SIAMESE MODEL

The second considered method consists of using a Siamese neural network model that uses Long Short-

Term Memory (LSTM) cells [9], inspired by the work of Mueller and Thyagarajan [10]. LSTM cells learn contextual dependences between words in sentences

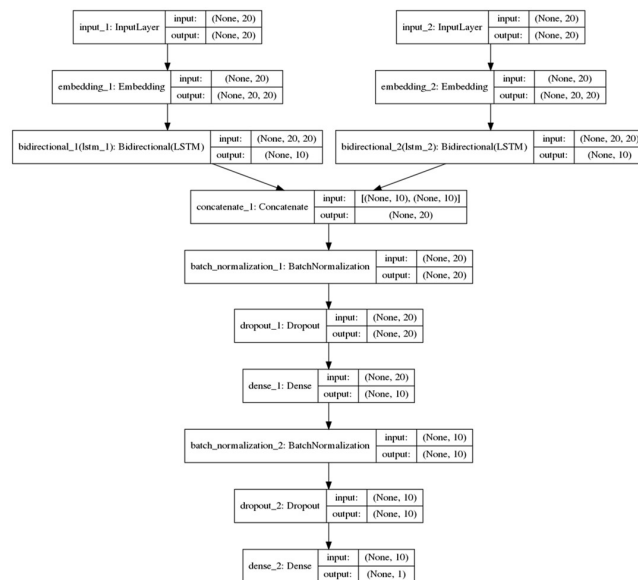


Figure 5. The siamese neural network architecture

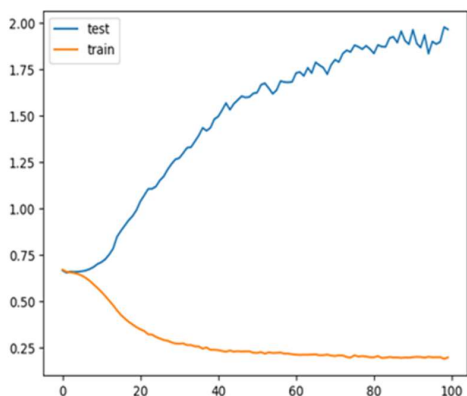
that can pass through the network at a large time difference.

Our siamese architecture employs bidirectional LSTM cells through which both the original and the inverted sequence are passed, thus defining a context for words from both a past and future perspective. Before the bidirectional LSTM layers, the model uses embedding layers that create a mapping for each word in the vocabulary to a continuous vector space. The model uses the pair of utterances that creates a point of divergence or convergence (current utterance and previous utterance mentioned as a divergence or convergence relation). The size of the embedding layers is the same as the number of the words a reply has and the dimension of the LSTM cells is set to 20. In Figure 5 we provide the complete representation of the siamese architecture.

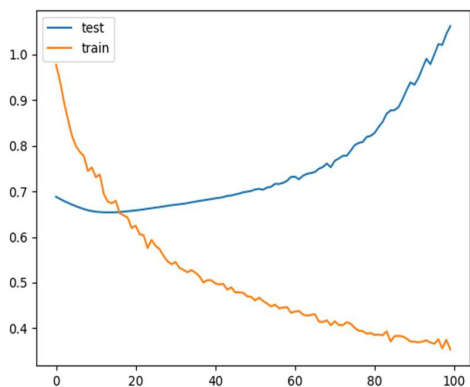
### RESULTS

The corpus of chats used for training and evaluation contains 63 conversations with multiple participants in an educational setting [4]. The dataset is split randomly, using 85% as training data and 15% as test data. The training is done over 100 epochs and the test dataset is used as validation data.

The results using the two neural networks proposed in the previous chapter are shown in Tables 1-4.



**Figure 6.** Evolution of the loss function for replies containing 40 words using 3-CNN network.



**Figure 7.** Evolution of the loss function for replies containing 40 words using Siamese network.

From the charts presented in Figures 6-7 that show the evolution of the loss function for both neural networks we can easily see the presence of overfitting. In the case of the Siamese network, the results are slightly better and we can conclude if we use a larger dataset the results the loss on the test data will be lower. Below are the results of loss and accuracy in both models, as well as precision per class. Maximum size of an utterance will be truncated to a dimension of 20, 30, 40 words in order to reduce the size of the vocabulary.

Test/Size Utterance = 20	3-CNN		Siamese	
Accuracy	0.6265		0.645	
Loss	0.6551			
Precision	Conv. 0.63	Div. 0.35	Conv. 0.64	Div. 0.37

*Table 1 Results for 3-CNN and siamese models using replies that contain a maximum of 20 words*

Test/Size Utterance = 30	3-CNN		Siamese	
Accuracy	0.6395		0.6395	
Loss	0.6558			
Precision	Conv. 0.62	Div. 0.33	Conv. 0.63	Div. 0.34

*Table 2 Results for 3-CNN and siamese models using replies that contain a maximum of 20 words*

Test/Size Utterance = 40	3-CNN		Siamese	
Accuracy	0.6395		0.6344	
Loss	0.6589			
Precision	Conv. 0.63	Div. 0.35	Conv. 0.63	Div. 0.35

*Table 3 Results for 3-CNN and siamese models using replies that contain a maximum of 40 words*

Test/Size Utterance = var.	3-CNN		Siamese	
Accuracy	0.6115		0.6395	
Loss	0.6853			
Precision	Conv. 0.63	Div. 0.35	Conv. 0.64	Div. 0.36

*Table 4 Results for 3-CNN and siamese models using replies that contain a variable number of words*

As a comparison, the results using Naive Bayes are shown in Tables 5-10. The entry Freq of the tables denotes that words that appear less than 10 times in the corpus will be deleted or no words will be removed.

Test/NB Monograms	Size Utterance = 20	
Accuracy/Freq < 10	0.66	
Accuracy/ Freq < 0	0.72	
Precision	Conv. Freq<10 0.75 Freq<0 0.76	Div. Freq<10 0.53 Freq<0 0.60

Table 5 Results of Naive Bayes using monograms for replies that contains a maximum number of 20 words

Test/NB Monograms	Size Utterance = 30	
Accuracy/Freq < 10	0.69	
Accuracy/ Freq < 0	0.73	
Precision	Conv. Freq<10 0.78 Freq<0 0.76	Div. Freq<10 0.56 Freq<0 0.64

Table 6 Results of Naive Bayes using monograms for replies that contains a maximum number of 30 words

Test/NB Monograms	Size Utterance = 40	
Accuracy/Freq < 10	0.68	
Accuracy/Freq < 0	0.73	
Precision	Conv. Freq<10 0.78 Freq<0 0.76	Div. Freq<10 0.55 Freq<0 0.66

Table 7 Results of Naive Bayes using monograms for replies that contains a maximum number of 40 words

Test/NB Monograms	Size Utterance = var	
Accuracy/Freq < 10	0.69	
Accuracy/Freq < 0	0.74	
Precision	Conv. Freq<10 0.78 Freq<0 0.76	Div. Freq<10 0.56 Freq<0 0.66

Table 8 Results of Naive Bayes using monograms for replies that contains a variable number of words

Test/NB Bigrams	Size Utterance = var	
Accuracy/Freq < 0	0.76	
Precision	Conv. 0.75	Div. 0.76

Table 9 Results of Naive Bayes using bigrams for replies that contains a variable number of words

Test/NB Trigrams	Size Utterance = var	
Accuracy/ Freq < 0	0.78	
Precision	Conv. 0.78	Div. 0.78

Table 10 Results of Naive Bayes using trigrams for replies that contains a variable number of words

## CONCLUSION

The paper proposes two models of neural networks that target the problem of detecting convergent and divergent replies in a collaborative chat with multiple participants. The accuracy of the proposed neural models is low because they tend to overfit. The main cause is the small size of the dataset used for training and the noisy character of the data in chat conversations.

For all these reasons, using Naïve Bayes, a simple machine learning technique, we obtain better results for the task at hand. In the future, by adding additional chats annotated with divergence and convergence relations and thus increasing the data used for training, the Siamese model promises good results.

An important conclusion of this paper is that neural methods are still lacking in performance for small datasets that are often used in education or other social sciences where gathering large datasets is difficult due to the complexity of the task at hand.

## REFERENCES

1. Stefan Trausan-Matu (2020) The Polyphonic Model of Collaborative Learning, in Mercer, N., Wegerif, R., & Major, L. (eds.), *The Routledge international handbook of research on dialogic education*, ISBN :978-1-138-33851-7, New York, NY : Routledge, pp. 454-468,
2. Stefan Trausan-Matu, Traian Rebedea (2010), A Polyphonic Model and System for Inter-animation Analysis in Chat Conversations with Multiple Participants, in A. Gelbukh (Ed.), *CICLing 2010*, LNCS 6008, Springer, 2010, pp. 354–363
3. Puccio, G. J., (1998). *Letters from the field*. Roper Review, 21, 85-86.
4. Stefan Trausan-Matu, Mihai Dascalu, Traian Rebedea, Alexandru Gartner, Corpus de conversatii multi-participant si editor pentru adnotarea lui, *Revista Romană de Interactiune Om-Calculator*, Vol.3, No.1, 2010, pp. 53-64, MatrixRom
5. Runco, M. A., (1990). *The divergent thinking of young children: Implications of the research*. Gifted Child Today, 13, 37-39.
6. Runco, M. A., and Albert, B. S., (1989). *Independence and the creative potential of gifted and exceptionally gifted boys*. Journal of Youth and Adolescence, 18, 221-223.
7. Brownlee J., (2017). *Deep Learning with Python: Develop Deep Learning Models on Theano and*

*TensorFlow Using Keras*. Melbourne, Australia: Machine Learning Mastery.

8. Mohammad Hamad Allaymoun (2016). *Analysis of rhetorical, altruistic convergent and divergent dimensions in CSCL chats*, Ph.D. Thesis.
9. Hochreiter S., Schmidhuber J. (1997). *Long short-term memory*. Neural computation. 9(8):1735-80.
10. Mueller J., Thyagarajan A., (2016). Siamese recurrent architectures for learning sentence similarity, in *Thirtieth AAAI Conference on Artificial Intelligence*.