

Analysis and prediction of politeness in conversations

Cristiana - Paula Iordache
University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
cristiana.iordache@gmail.com

Ștefan Trăușan - Matu
University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
and
Research Institute for Artificial
Intelligence
and
Academy of Romanian Scientists
stefan.trausan@upb.ro

ABSTRACT

This paper describes an approach for the analysis of the politeness level in a conversation and a prediction model measuring the probability that a conversation will evolve into a conflict between its participants. A sliding window method is used for analyzing the most recent utterances in a conversation, in order to assign a politeness score for the conversation in every moment of its development. The approach presented in the paper enhances the analysis by using a punctuation tuning method for each utterance in the sliding window and by including a polyphonic analysis in the conversation score computation. A graphical interface model is introduced for empowering the user to conduct a detailed politeness analysis of the development and evolution of conversations. Potential uses of the analysis are described, with an emphasis on chat moderator bots capable of alleviating conflictual situations in a conversation, while also suggesting further improvements to the analysis model.

Author Keywords

Conversation analysis; analysis of politeness; prediction of politeness; polyphony analysis;

ACM Classification Keywords

I 2.7 Natural Language Processing. Text analysis.

DOI: 10.37789/rochi.2021.1.1.4

INTRODUCTION

Online chat applications have become some of the most popular interactive computer programs, available to and used by billions of users worldwide. This kind of applications, especially those containing group chat functionalities, often encounter conflicts between their users, tense situations which can lead to personal attacks and might determine some of the users to stop using the applications.

A solution to alleviating these conflictual situations is including a moderator functionality for the conversations, either in the form of a chat bot or a human moderator, capable of intervening in the conversation to alleviate it. In order to early predict a future conflict, before its eruption, and to determine the right way of interceding in the conversation, both human and chat bot moderators would require a flexible application able to offer detailed information about the

politeness level of the conversation and its evolution in time, with easily changeable settings, to ensure relevance to a wide range of conversations.

This paper proposes an analysis and prediction model for the probability that a conversation will evolve into a conflict, wrapped up in a easy to use and adjustable graphical interface. The model attributes a conversation politeness score at each moment in the conversation development, computed using a sliding window method (only the most recent utterances influence the score). For each utterance comprising the sliding window, a politeness score is associated, using the *politenessStrategies* module implemented as part of Cornell Conversational Analysis Toolkit (ConvoKit) [1]. The research presented in this paper further enhances the utterance score through a punctuation tuning method, which identifies punctuation patterns capable of distinguishing underlying stronger emotions conveyed by the conversation participants through exaggerated or repeated punctuation. Another enhancement this paper brings to the conversation score computation resides in a polyphony analysis [8, 9], capable of identifying the transmission of politeness related themes and motifs from one participant to the conversation to others, idea spreading which gives insight into potential escalation of negative emotions or possibly the conversion of previously impolite users to polite levels of speech.

The graphical interface model presented in this paper allows for powerful customization of algorithm parameters, both before and during the running of the program, parameters such as the sliding window length and the impact of punctuation tuning and polyphony analysis on the score computation. One of the important interface functionalities is delineated by the plotting of a graphical representation of the conversation's evolution in time.

STATE OF THE ART

The analysis of the politeness level of a conversation has been discussed in several papers [2, 3, 4]. Danescu-Niculescu-Mizil et al. [4] focus their analysis on correlating social factors to the level of politeness in a conversation. While not sharing the same purpose as ours, this research nevertheless served as a starting point for the analysis and provided a politeness annotated corpus, while also

representing the basis of the ConvoKit *politenessStrategies* module used in the implementation presented herein.

Zhang et al. [2] and Chang and Danescu-Niculescu-Mizil [3] focus, as this paper does, on predicting the probability of a conversation to go awry, targeting the assigning of a label to the conversation (it will turn into a conflict at some point in the future or not). The analysis conducted by these papers starts from pragmatic devices used in the initiation of a conversation, such as politeness strategies and rhetorical prompts.

The approach presented in this paper likewise identifies politeness strategies through the usage of the ConvoKit classifier, but it does not take into consideration the rhetorical prompts, as these are most relevant for the debut of a conversation and this paper's analysis focused on a real time evaluation of a conversation, at any point in its evolution. This paper presents improvements of the analysis achieved by the denominated papers by assigning a politeness score for the entire conversation in every moment of its development and by including other politeness relevant factors in the analysis, most notably scrutinizing each utterance also on a punctuation level and by identifying the spread of the politeness motifs amongst the participants, through the polyphony analysis. Not in the least, this paper augments the previous mentioned work [2, 3] through the creation of a powerful and flexible graphical interface, allowing for clearer interpretation of the results. The customization of the algorithm parameters (such as the sliding window length, the punctuation tuning and polyphony analysis influence) empowers the user to perceive a more complete picture of the conversation's politeness details and permits more accurate score computation for a wider range of conversations, as the setup values most relevant to a conversation depend on its particularities. Furthermore, the previously discussed papers focus mostly on empowering human moderators to analyze conversations from a politeness perspective, while this paper is capable of providing both human and chat bot moderators with a powerful conversation analysis tool.

The polyphony analysis has been previously researched in several papers [8, 9] and it is based on Mikhail Bakhtin's ideas of dialogism, inter-animation, ventriloquism, and polyphony [5, 6, 7]. Bakhtin surfaced similarities between written or oral discourse and the interactions of instruments and voices in orchestral concerts, which mimic melodic lines from each other, turning them into motifs. In the context of this paper, polyphony has been used to identify the spreading of politeness related ideas and syntagms amongst the conversation participants, indicating the direction the conversation is heading to and offering precious insight into the level of politeness the discussion will exhibit in the near future.

IMPLEMENTATION

The developed program empowers the user to obtain detailed information about the evolution of conversations regarding the politeness level, with functionalities such as obtaining the conversation politeness score for the most recent utterances comprising a sliding window, the list with the most impolite speakers and a graphical representation of the conversation's evolution in time.

Obtaining the politeness score for the current development of the conversation is computed as a weighted sum of four major factors. This computation, alongside the inclusion of punctuation tuning and polyphony analysis, and the creation of the customizable graphical interface, are improvements brought by this paper to the state of the art work. The first factor, with the highest weight, is represented by the utterances score, which considers the individual politeness score of each utterance from the sliding window and is likewise computed as a weighted sum, with the most recent utterances having the highest impact, the significance of utterances decreasing as they are older. While still influencing the politeness score, older utterances usually represent situations that have been solved and subjects that have been discussed, which might not be relevant to the current evolution of the conversation and its potential to digress into personal attacks. The politeness score for each utterance is obtained with the help of the ConvoKit classifier and is further enhanced through the use of a punctuation tuning method, which captures the use of punctuation patterns indicating underlying emotions transmitted by the speaker or the emphasis of the sentiments conveyed through the text of the utterance.

The second important factor in the computing of the conversation score is represented by a polyphony analysis score, capturing the spreading of politeness related themes among the participants, giving insight into the influence over the conversation's future development. The weight of this factor is controllable by the user and can be even deactivated, with the weight being redistributed to the utterances score. The polyphony analysis score is computed as a difference between a positive polyphony score (highlighting the spreading of polite ideas amongst the conversation participants) and a negative polyphony score (capturing the adoption and usage of the negative ideas by more and more speakers). Both the positive and negative polyphony motifs and syntagms can be customized in two separate files, with default syntagms being provided by the baseline software program. This allows for fine tuning acknowledging the conversation's particularities and subject, in accordance to the participants' background, level of education and culture etc.

The third and fourth factors are represented by the positive speakers score and negative speakers score, respectively. These scores are computed as the mean of the politeness scores of the participants exhibiting a high level of politeness and, correspondingly, a high level of impoliteness, capable

of influencing the other speakers and, as a consequence, the evolution of the conversation in the near future. The importance of these factors is justified by the frequently identified situations (in direct, in-person group conversations) where, if many speakers impose their level of politeness on the conversation, they will likely cause the other speakers to assimilate and adopt the politeness level the influencing group is using (the vocal and opinioned group imposes the direction the conversation will take, changing the stance of the more reserved and easily influenced participants).

The program allows for powerful and in-depth customization of the algorithm parameters, at any point in the running of the program, through the modification of a setup file. This customization renders the program flexible, suitable for a wide range of conversations and capable of painting a much more detailed image of the conversation's politeness development, offering all the necessary information for the user to take the right decision of whether to intervene or not in the conversation to alleviate it and helping to determine what is the best solution to attenuate the conflicts. The parameters left for the customization of the user include the sliding window length, the level of the punctuation tuning impact on each utterance and the weight of the polyphony analysis on the conversation score, the numbers of points plotted on the graphical representation and the distance in utterances between those points. All customizable parameters contribute strongly to obtaining different results for different setup values. While the best setup values for a certain conversation depend strongly on the particularities of the conversation (the type of the conversation, the number of speakers, the number of redundant utterances, how fast the

conversation changes from a polite tendency to conflict situations etcetera), it is heartily recommended to try different setup values in the same running instance of the program (changing the setup values is very quick and easy), in order to gain a powerful detailed image of the conversation's evolution and to be able to more accurately predict its future politeness path.

An example of the conversation score and impolite speakers list visualization is presented in Figure 1, while a graphical representation of a conversation's evolution in time is visible in Figures 2 and 3. By clicking on any point of a graphical representation such as the one illustrated in Figure 3, the user can visualize the utterances composing the sliding window that generated the point's politeness score, allowing for a validation of the algorithm's prediction (this functionality can be observed in Figure 2, with the green bordered point representing the part of the conversation visible in the pop-up window). The utterances responsible for determining the current politeness score of the conversation can also be visualized in the command line interface (pictured in Figure 1), by entering the command "utterances" in the terminal.

```
(thesis) cristiana@Arcadia:~/Licenta$ python3 politeness.py setup.txt
050/050 utterances processed

Impolite users:
David Grant with the score: 0.351
Anthony Jones with the score: 0.369
Jane Green with the score: 0.345
Will Barnes with the score: 0.361

The conversation score is 0.374
```

Figure 1. Example of program prediction of the conversation score and impolite speakers list visualization

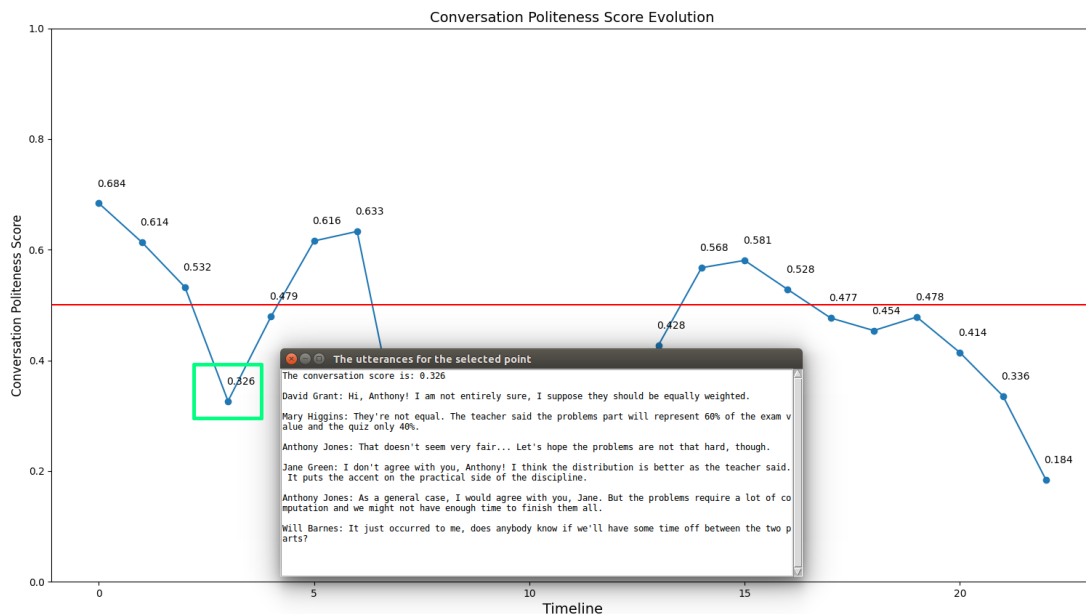


Figure 2. Graphical representation of the evolution in time of a conversation, showcasing the visualize composing utterances functionality

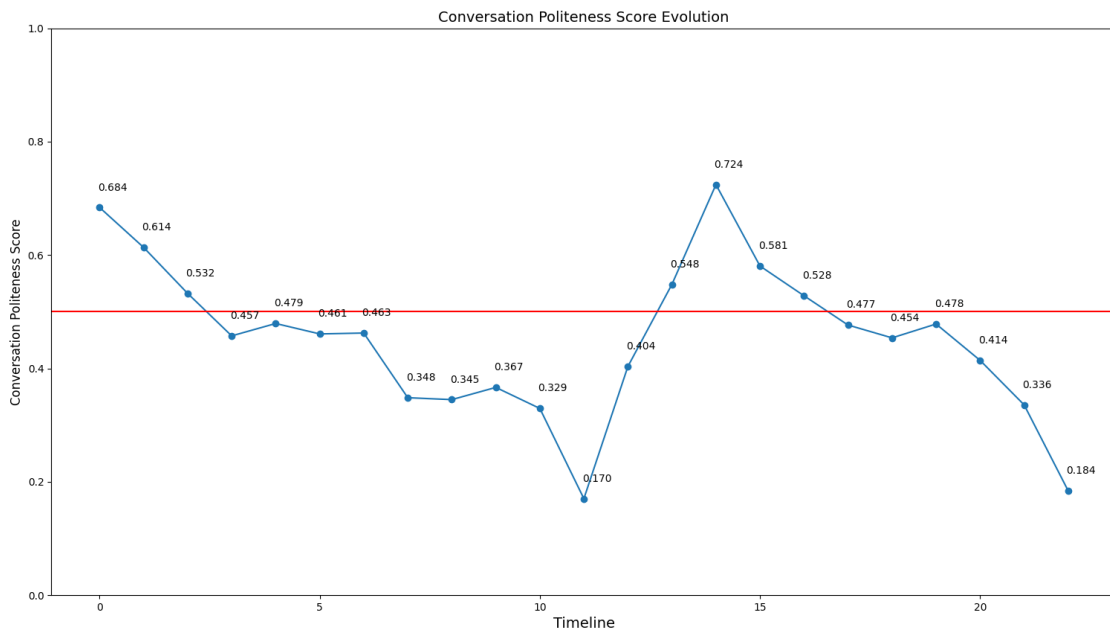


Figure 3. Graphical representation of the evolution in time of the conversation used in the analysis section, with a sliding window of length 6, the distance between points on graph equal to 2 utterances and where both the punctuation tuning and the polyphony analysis are ignored.

ANALYSIS AND RESULTS

The results discussed in this section have been obtained by running the program on a conversation with 50 utterances and 6 speakers. The speakers are students, starting the conversation before the beginning of a two-part exam, exchanging messages during the pause between the two parts and finishing the conversation after the second part of the exam. The conversation opens on a neutral tone, with questions and subsequent answers regarding the examination process, followed shortly by a quickly alleviated tense situation occurring between two of the speakers. Some of the other participants manage to intervene in time by changing the immediate subject of the discussion. The conversation pauses during the first part of the exam and reignites in the break between the parts, manifesting a strong conflict caused by the same two speakers responsible for the previous disagreement. The conflict starts drawing in more participants and becomes increasingly aggressive, exhibiting personal attacks and insults. As the second part of the exam draws closer, the conflict starts to abate, with the end of the exam bringing with it apologies from the impolite users (with the negative attitude justified as an effect of the stressful exam) and a positive view over the conversation's future. While maintaining a civil tone for a while, the conversation starts again to digress, finishing with another conflict.

As it can be observed in Figure 3, the evolution of the conversation, as it is described in the above paragraph, is represented in great detail, closely following the recounted changes in the politeness level and the conflictual situations

occurrences. Still, the resulted graphic does not capture neither the early conflictual situation between the two speakers causing the ensuing strong animosity, nor the subsequent alleviation succeeded through the changing of the immediate subject. These shortcomings are surmounted by including the polyphony analysis in the computation of the conversation score, as it is visible in Figure 4.

Figure 4 captures the inclusion of the punctuation tuning on a level of 5 out of the maximum 10 value and the addition to the conversation score computation of the polyphony analysis score with a weight of 0.225 out of 1 (corresponding to level 8 out of the maximum 10 for the polyphony analysis). The green bordered area indicates the early conflict between the two speakers, followed shortly by the successful moderator attempt, shown with a yellow marking. Those two conversation areas are not correctly assigned through the representation in Figure 3, as the identification of the impolite situation and its subsequent settlement has only been achieved by analyzing the spread of politeness related ideas and syntagms, captured through the polyphony analysis. Furthermore, the conflict alleviation occurring after the strong strife at the middle of the conversation is attenuated in its positive prediction, as visible in the area bordered with black. The justification for this attenuation lies in the continuous existence of impolite motifs spreading through the conversation participants, as the conflict has not been completely overcome. As such, it can be stated that the inclusion of the polyphony analysis painted a more reliable image of the conversation's evolution and has been able to offer an earlier warning of impoliteness to come.

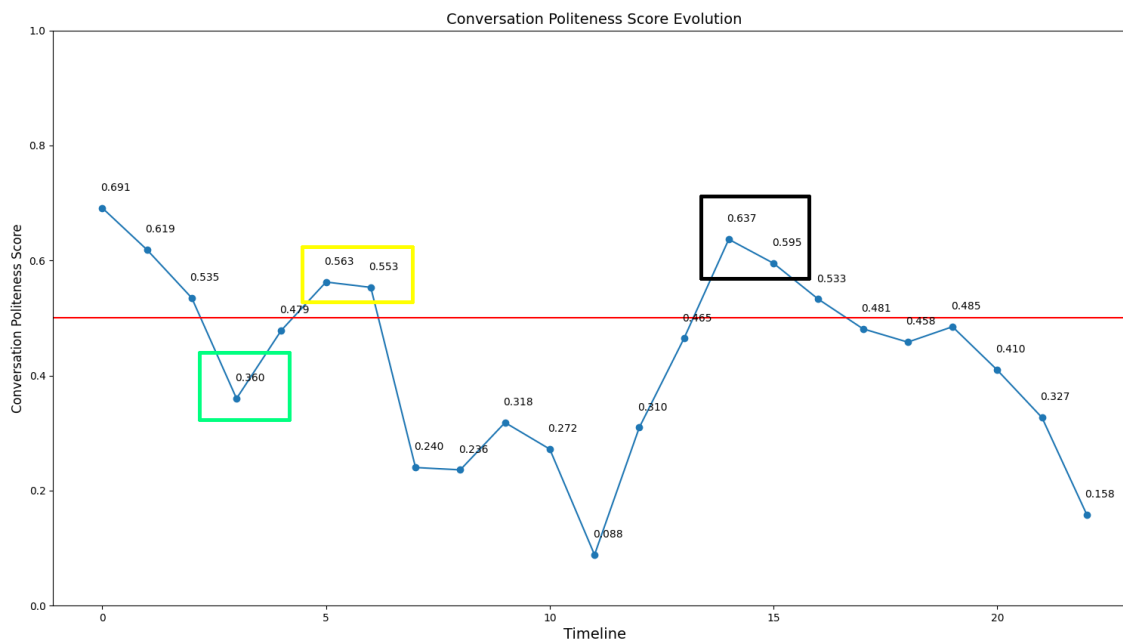


Figure 4. Graphical representation of the evolution in time of the conversation used in the analysis section, with a sliding window of length 6, the distance between points on graph equal to 2 utterances and where the punctuation tuning is assigned a level of 5 out of 10, while the polyphony analysis weight on the conversation score is assigned a value of 0.225 out of 1.

Adding the punctuation tuning in Figure 4 caused the adjustment of points on graph corresponding to areas exhibiting emphasized punctuation. These areas usually represent the politest sections of the conversation and the most conflictual. As such, the punctuation tuning adjustment has usually an effect of stretching the range of the scores on the graphic and creating spikes of low and high points. The punctuation tuning corrects the conversation’s prediction and generally accentuates it in the direction of its classification, but for some points, the adjustment causes a reclassification of the conversation (either passing from a polite prediction to an impolite one, or vice versa).

For the conflictual sections of the conversation, the punctuation tuning has adjusted the score by more than 0.05, sometimes even causing the inclusion of additional speakers to the impolite speakers list. The polyphony analysis has also corrected the impolite sections with as much as 0.04, while including both the punctuation tuning and the polyphony score caused the modification of the conversation by a value of 0.075. These significant alterations improve the accuracy of the model and as it can be observed by comparing Figure 3 and Figure 4, generate a more thorough and clear examination of the conversation’s evolution.

CONCLUSIONS AND FUTURE WORK

This paper presents an improved analysis and prediction model of the politeness level of a conversation, adding further analysis of the conversation’s particularities and enveloping the program in a strong and flexible graphical interface, empowering its users to closely monitor a

conversation’s politeness development and accurately predict its future evolution. The improvements over state-of-the-art work include the adding of a punctuation tuning method to identify the usage of punctuation patterns aimed at emphasizing the conveyed emotions, the inclusion of a polyphony analysis to uncover the spreading of politeness related themes through the conversation participants and the computation of a politeness score at every point in the development of the conversation.

A promising area of future work embodies the inclusion of the algorithm this paper proposes and the graphical interface that accompanies it into chat moderator bots, capable of monitoring the development of conversations and intervening when detecting levels of impoliteness foreboding conflicts.

Further improvements to the model might include the adding of other analysis factors such as the analysis of the politeness profile of each participant, the analysis of the conversation’s subject, analysis of social factors pertaining the predicted politeness level of the conversation right from its start etc.

One limitation of the solution described by this paper is represented by its restriction to only one language (English), as the politeness strategies used to assign a politeness score to each utterance are heavily dependent on the particularities of a language. This limitation can be overcome through creating multiple classifiers, for each language present in the conversation. Each utterance would, therefore, require a language annotation, indicative of the classifier that should be used to assign it a politeness score. However, this

potential solution would still encounter difficulties with mixed language utterances (sentences written in multiple languages).

REFERENCES

1. Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, Cristian Danescu-Niculescu-Mizil. "ConvoKit: A Toolkit for the Analysis of Conversations". Proceedings of SIGDIAL 2020, pp.57-60.
2. Justine Zhang, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, and Nithum Thain, Yiqing Hua, Dario Taraborelli. "Conversations Gone Awry: Detecting Early Signs of Conversational Failure". Proceedings of ACL 2018, pp. 1350-1361.
3. Jonathan P. Chang and Cristian Danescu-Niculescu-Mizil. "Trouble on the Horizon: Forecasting the Derailment of Online Conversations as they Develop". Proceedings of EMNLP 2019, pp. 4743-4754.
4. Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec and Christopher Potts. "A computational approach to politeness with application to social factors". Proceedings of ACL 2013, pp. 250-259.
5. Mikhail M. Bakhtin. "The Dialogic Imagination Four Essays". University of Texas Press, Austin, 1981
6. Mikhail M. Bakhtin. "Speech Genres & Other Late Essays". University of Texas Press, Austin, 1986
7. Mikhail M. Bakhtin. "Problems of Dostoevsky's Poetics". University of Minnesota Press, Minneapolis, 1993
8. Stefan Trausan-Matu, "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, 2020
9. Stefan Trausan-Matu, Gerry Stahl, and Alan Zemel, "Polyphonic Inter-animation in Collaborative Problem Solving Chats", Research Report, Drexel University, Philadelphia, 2005, https://www.academia.edu/4312894/Polyphonic_Inter_animation_in_Collaborative_Problem_Solving_Chats, last accessed at 15.06.2021
10. Cristiana-Paula Iordache, "Analysis and prediction of politeness in conversations", Bachelor thesis, University Politehnica of Bucharest