Analyzing motivating texts for modelling human-like motivation techniques in emotionally intelligent dialogue systems

Patrycja Swieczkowska[1,2], Rafal Rzepka[1,2], and Kenji Araki[1]

[1] Hokkaido University, Sapporo, Japan
http://arakilab.media.eng.hokudai.ac.jp
{swieczkowska,rzepka,araki}@ist.hokudai.ac.jp
[2] RIKEN AIP, Tokyo, Japan

Abstract. In this paper, we present studies on human-like motivational strategies which eventually will allow us to implement motivational support in our general dialogue system. We conducted a study on user comments from a discussion platform Reddit and identified text features that make a comment motivating. We achieved around 0.88 accuracy on classifying comments as motivating or non-motivating using SVM and a shallow neural network. Our research is a first step for identifying computational features of a motivating piece of advice, which will subsequently be useful for implementing the ability to support the user by motivating him, imitating real human-to-human interactions.

Keywords: Human motivation understanding, Artificial agents, General dialogue systems.

1 Introduction

Emotional intelligence is defined in psychology as, among others, the capability of individuals to recognize emotions of others and use emotional information to guide thinking and behavior [1]. In the context of various agents, from cognitive architectures to dialogue systems, this means being able to respond to various emotional states of the user, which already became an important research topic [2, 3]. The authors of [2] dub systems capable of this "relational agents" and define them as "computational artifacts designed to build long-term, social-emotional relationships with their users." They recognize the importance of implementing human-like emotional intelligence in machines, which in case of dialogue systems is especially crucial, as they are specifically designed to interact with humans.

Likewise, we recognize the need to provide an artificial agent with the necessary skills to establish a successful cooperation with a human user. Specifically, we aim to design a dialogue system that would motivate the user to complete tasks on their schedule, regardless of the type of task or reason for being unmotivated, while employing motivating strategies inspired by real human utterances.

The user will tell the system about their lack of motivation using natural language, and the system will produce a response meant to give the user some advice pertaining to the problem at hand. To be able to do this, the system has to know how to create motivational utterances. Therefore, the first step in our research was to examine texts containing motivational advice to find out what particular features they possess. Once we discover what makes an utterance motivating, we will be able to use that knowledge to generate motivational advice for the user.

The concept of motivation has been extensively studied in psychology (mostly involving gamification [4]), but not so much in the field of dialogue systems. While there exist papers suggesting various approaches to influencing motivational states in users [5, 6], they do not contain experiments confirming their hypotheses. Therein lies the novelty of our research; to the best of our knowledge, ours will be the first dialogue system that motivates people to perform all kinds of tasks. Our system should be able to imitate actual human interactions, especially with respect to providing emotional support.

2 Datasets and Features

To be able to implement the ability to motivate users into our system, first we needed to determine what makes an utterance motivating. To achieve this, we analyzed posts and comments from an online discussion platform Reddit. Specifically, we accessed the subreddit r/getdisciplined[3], where the users post their issues with being unmotivated to perform various tasks, such as exercising, studying, going to work and so on. The posters then ask the commenters to provide them with some motivational advice. We chose this subreddit because the posts and comments are very close to the type of input and output of our end-goal dialogue system.

We noticed that best-ranked comments from r/getdisciplined usually had two things in common: they provided very specific, practical advice for the poster and included expressions of being able to relate to the poster's struggles, usually because the commenter had to deal with the same problem in the past. Consequently, we operationalized these characteristics into several features in our experiment. We also expanded our data with comments from non-motivational subreddits to compare them against our original motivational dataset.

2.1 Datasets

To train our classifier for the purpose of distinguishing motivating/advice comments from other texts, we created several datasets. Table 1 below is an overview of the number of comments in all our datasets. Specifically, we had two datasets of motivational/advice-giving comments from subreddits r/getdisciplined and r/relationship_advice[4], and two datasets of other comments from subreddits

[3] https://www.reddit.com/r/getdisciplined/
[4] https://www.reddit.com/r/relationship_advice/

r/pics[5] and r/todayilearned[6]. Both of the latter subreddits require posts to be interesting or amusing pictures and screenshots, which makes the comments unlikely to contain any advice.

Table 1. Overview of the datasets.

Subreddit	Training set	Test set
r/getdisciplined	1,352	573
r/relationship_advice	2,470	1,022
r/todayilearned	3,395	1,435
r/pics	2,255	1,144

2.2 Features

Before analyzing the comments, we pre-processed each one by detecting sentence boundaries, assigning part-of-speech tags, and, for some features, removing stopwords. From now on, we will use names *wordlist_withstops* for a list of all words in the comment, *wordlist_nostops* for the same list with stopwords removed, and *sent_list* for a list of sentences in the comment. The entire feature set included 13 features which were as follows:

Sentics scores of **aptitude, attention, pleasantness and sensitivity** measured with the Sentic library for Python on *wordlist_nostops*. The library is an API to the SenticNet knowledge base[7]. All the values fall on the scale between -1 and 1.

Sentiment score also provided by Sentic and measured on *wordlist_nostops*. The results fell on the five-point scale of strong negative / weak negative / neutral / weak positive / strong positive, which we converted accordingly to integer values between -2 and 2.

Relatability score measured by the percentage of first person pronouns (including possessive pronouns) in *wordlist_withstops*. The score range is 0 to 1.

Imperative score measured by the percentage of imperative/advice expressions in the comment text. Specifically we looked for clauses beginning with non-infinitive verbs but not ending in question marks, the word *please* preceding a verb, the phrase *why don't you* and phrases comprised of *you* or *OP* (*Original Poster*, which is a popular way of referring to the author of the post on Reddit) and a modal verb. Since most of these are bigrams, we counted the percentage on number of all words divided by 2. The score range is -1 to 1; negative values come from deducting points for question marks.

Specificity scores including six features: **Average Semantic Depth (ASD)** and **Average Semantic Height (ASH)** calculated from scores for each word in

[5] https://www.reddit.com/r/pics/
[6] https://www.reddit.com/r/todayilearned/
[7] http://sentic.net/api/

the sentence as retrieved from the WordNet ontology's hypernymy/hyponymy hierarchy, **Total Occurrence Count (TOC)** measured by obtaining occurrence count in WordNet for each word and adding up three lowest scores in a sentence, **Count of Named Entities (CNE)** and **Proper Nouns (CPN)** in the sentence, and **Sentence Length (LEN)**. These calculations were performed for each sentence in the comment using *sent_list* with stopwords removed. The final scores for the entire comment were obtained by adding up all the sentence scores. We then divided ASD, ASH, TOC and LEN by 100 and CNE and CPN by 10 to put the scores in the same numerical range as other features. Specificity score was first proposed by [9] to help extract suggestions and complaints from employee surveys and reviews. The goal was to find sentences containing specific content, which we adapted in our experiment for the purpose of finding specific motivational advice in comments. We calculated the scores as described in [9] with only slight modifications.

3 Experiments and Results

To test our method, we used two classifiers: a Support Vector Machine and a custom-made fully connected shallow neural network with two layers. We chose an SVM because they are robust thanks to their large margin optimization technique and perform well in classification problems. We then implemented the shallow neural network to see whether we could improve on the SVM results.

The neural network had an input layer of one unit per each of our 13 features, a hidden layer of 10 units using the tanh activation function and an output layer with one unit using the sigmoid activation function. For training, we used the parameters learning rate = 0.2 and number of iterations = 20,000. For the SVM computations we used an RBF kernel with the parameter C=20. All the parameters were chosen based on algorithm performance.

We achieved accuracy of 0.86 in the r/getdisciplined vs. r/pics and r/getdisciplined vs. r/todayilearned experiments using SVM. This score rose to 0.88 with the neural network.

To increase the amount of learning data, we then combined the r/getdisciplined and r/relationship_advice datasets into one *all-motivational* class, and r/pics and r/todayilearned into another *non-motivational* class. The performance of both SVM and the neural network was slightly worse, decreasing to 0.84.

Table 2 summarizes our results.

4 Discussion

In the test set for r/getdisciplined vs r/todayilearned we had 265 misclassified comments. Around 78% of them (207 comments) were r/getdisciplined comments misclassified as not motivating. A closer look at the data suggests that most such comments often had very similar values of ASD (*Average Semantic Depth*) and ASH (*Average Semantic Height*). While ASD and ASH are not dependent on

Table 2. Accuracy scores for different datasets.

Dataset	Training examples	Test examples	SVM	Shallow NN
r/getdisciplined vs r/pics	3,607	1,717	0.86	0.87
r/getdisciplined vs r/todayilearned	4,747	2,008	0.86	0.88
all-motivational vs non-motivational	9,472	4,174	0.84	0.84

each other, perhaps similar values of these features make it somehow harder to classify the comment properly.

Moreover, some misclassified comments may have contained unusual punctuation that influenced calculating individual features. For example, some comments had Imperative Score indicating that the algorithm did not detect some non-infinitive imperative verbs if they came after a quotation mark. This indicates that there is a need to improve the part of our algorithm responsible for calculating scores for this feature.

For the CPN (*Count of Proper Nouns*) calculations we used a chunker available in the nltk[8] library for Python. We specifically looked for NNP (proper noun, singular), NNPS (proper noun, plural) and CD (cardinal number) tags. However, in some cases capitalized words were counted as proper nouns even though they do not belong to this category. Therefore, it is important to improve calculations for this feature, perhaps by using a different tagging tool.

Further analysis revealed that error rates on both training and test sets were in the same close range at a relatively high value of 0.13. This suggests a bias problem in our algorithm. A way to fix this would be to use a deeper neural network, or one with a different architecture. Alternately, we can add more input features to our algorithm. For example, since similar ASD and ASH seem to be causing problems, perhaps we could additionally combine them into another feature that would be more informative to the algorithm, as well as adding some new features.

Lower scores for a bigger dataset (*all-motivational* vs *non-motivational)* suggest there might be also more noise in the data. Around half of the comments labeled as motivating (because they came from the r/getdisciplined subreddit) turned out to not contain any motivational advice, which in turn greatly contributed to the high error rate in our results. It is reasonable to expect that this problem gets bigger with bigger datasets. To counter this issue, in future research we can use only a few top-rated comments for each post, thus ensuring their motivational quality.

5 Conclusions

In this paper, we proposed an algorithm for classifying texts as motivational (advice-giving) and non-motivational. This classification is a part of bigger research into motivational features of texts. Using a SVM and a shallow neural

[8] https://www.nltk.org/

network, we achieved 0.88 accuracy on our test sets. As such, we have successfully determined a large subset of text features that make it motivating. Therefore, our research is an important step for giving machines the ability to support and motivate a human to perform various tasks, which can be thought of as computationally implementing a part of emotional intelligence.

To further improve the results, we are planning to add more features to our algorithm, as well as trying out different deep neural network architectures. In the long-term, after determining the features that make a text motivating, we will use this knowledge to construct a language generation module in our dialogue system to provide users with motivational advice suited to their needs.

Advice provided by Reddit commenters proved to be an invaluable source of knowledge how human users can be motivated. Artificial emotional intelligence can greatly benefit from knowledge provided by large online resources, not only by learning how to recognize human emotions, but also by acquiring knowledge about dealing with these emotions efficiently. We believe that in a long run, Wisdom of the Crowd-based knowledge might become a useful source for simulating emotional intelligence in cognitive architectures, improve their understanding of human behavior and enrich human communication with non-biological entities.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number 17K00295.

References

1. Coleman, A.: A dictionary of psychology. 3rd edn. Oxford University Press (2008)
2. Bickmore, T., Picard, R.: Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI) **12**(2), 239–327 (2005)
3. Callejas, Z., Griol, D., Lopez-Cozar, R.: Predicting user mental states in spoken dialogue systems. EURASIP Journal on Advances in Signal Processing **2011**(1), 6 (2011)
4. Mekler, E. D., Bruhlmann, F., Tuch, A. N., Opwis, K.: Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior **71**, 525–534 (2017)
5. Tielman, M., Neerincx, M., Brinkman, W.-P.: Generating situation-based motivational feedback in a PTSD E-health system. In: International Conference on Intelligent Virtual Agents, pp. 437–440 (2017)
6. He, H. A., Greenberg, S., Huang, E. M.: One size does not fit all: applying the transtheoretical model to energy feedback technology design. in: Proceedings of the SIGCHI conference on human factors in computing systems, pp. 927–936 (2010)
7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J: Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 3111–3119 (2013)
8. Quoc, V. L., Mikolov, T.: Distributed Representations of Sentences and Documents. In: International Conference on Machine Learning, pp. 1188–1196 (2014)
9. Deshpande, S., Palshikar, G. K., Athiappan, G.: An unsupervised approach to sentence classification. In: International Conference on Management of Data COMAD 2010, p. 88. Computer Society of India (2010).