Vol.2018-MUS-118 No.2
Vol.2018-SLP-120 No.2
2018/2/20

# Positive Emotion Elicitation in an Example-Based Dialogue System

Nurul Lubis, Sakriani Sakti, Koichiro Yoshino, and Satoshi Nakamura

Nara Institute of Science and Technology

**Abstract:** An emotion-sensitive dialogue system is highly potential in addressing user emotional needs through HCI. For example, by providing a general emotional support through positive emotion elicitation. To date, works on emotion elicitation have only focused on the intention of elicitation itself, but are yet to examine the process that gives rise to the change of emotion in the first place. In this paper, we utilize examples of spontaneous human emotional responses to elicit a positive emotional impact through dialogue system interaction. Text-based human subjective evaluation with crowdsourcing shows that the proposed dialogue system elicits an overall more positive emotional impact.

## 1 Introduction

Recently, there has been an increasing interest in eliciting user's emotional response [1, 2]. However, emotion appraisal that gives rise to the elicited emotion is not yet studied for this task. That is, the relationship between an utterance, which acts as *emotional trigger*, and the resulting *emotional response*. By examining this, it would be possible to reverse the process and determine the appropriate trigger to a desired emotional response.

In this paper, we attempt to elicit a positive emotional change through computer interaction by exploiting examples of emotion appraisal. We collected dialogue sequences containing emotional triggers and responses as examples in a dialogue system. Subsequently, we augment the traditional response selection criterion with emotional parameters: 1) user's emotional state, and 2) expected future emotional impact of the candidate responses.

## 2 Emotion Definition

In this work, we define the emotion scope based on the *circumplex model of affect* [3]. Two dimensions of emotion are defined: *valence* and *arousal*. Valence measures the positivity or negativity of emotion; e.g. the feeling of joy is indicated by positive valence while fear is negative. On the other hand, arousal measures the activity of emotion; e.g. depression is low in arousal (passive), while rage is high (active). Readers are referred to [3] for further details of this emotion model.

Figure 1 illustrates the valence-arousal dimension in respect to a number of common emotion terms.

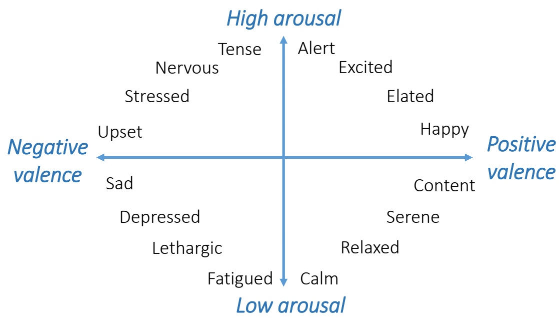


Figure 1: Emotion dimensions and common terms.

## 3 Example-based Dialogue Modeling (EBDM)

EBDM is a data-driven approach of dialogue modeling that uses a semantically indexed corpus of *query-response* pair examples instead of handcrafted rules or probabilistic models [4]. At a given time, the system will return a response of the best example according to a semantic constraint between the query and example queries.

Lasguido et al. have previously examined the utilization of cosine similarity for response retrieval in an example-based dialogue system [5]. In their approach, the similarity is computed between TF-IDF weighted term vectors of the query and the examples.

Given a query, the cosine similarity scores between its term vector and each of the example queries' in the database are calculated, and treated as the example scores. The response of the example with the highest score is then returned to the user as the system's response.

This approach has a number of benefits. First, The TF-IDF weighting allows emphasis of important words. Such quality is desirable in considering emotion in spoken utterances. Second, as this approach does not rely on explicit domain knowledge, it is practically suited for adaptation into an affective dialogue system. Third, the approach is straightforward and highly reproducible. On that account, it serves as the baseline in this study.

# 4 Proposed Dialogue System

We make use of *tri-turn* units in the selection process in place of the *query-response* pairs in the traditional EBDM approach. A tri-turn consists of three consecutive dialogue turns that are in response to each other. In this work, we exploit the tri-turn format to observe emotional triggers and responses in a conversation. Within this work, the first, second, and third turns in a tri-turn are referred to as *query*, *response*, and *future*, respectively. The change of emotion observed from *query* to *future* can be regarded as the impact of *response*.

In designing our method, we consider an analogy in human-human communication: emotional impact of a response is heavily dependent on semantic and emotional context of the conversation. For example, an apology could have contrasting outcomes depending on what happened preceding the apology as well as the emotional state of the listener. This suggests that the emotional impact in our examples are specific to the tri-turn context. In other words, semantic and emotional similarity is pre-requisite for the response to yield consistent impact in real interaction.

In addition to semantic constraint as described in Section 3, we formulate two types of emotional constraints: (1) emotion similarity between the query and the example queries, and (2) expected emotional impact of the candidate responses.

We measure *emotion similarity* by computing the Pearson's correlation coefficient of the emotion vector between the query and the example queries. Secondly, we measure the *expected emotional impact* of the candidate responses by calculating the difference of emotional states between each of their respective query and future.

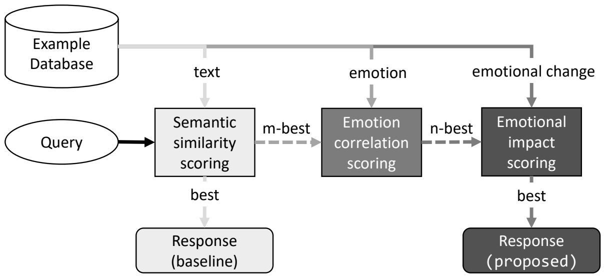


Figure 2: Steps of response selection.

Figure 2 illustrates the steps of response selection of the baseline and proposed systems. We perform the selection in three steps based on the defined constraints. For each step, a new score is calculated and re-ranking is performed only with the new score, i.e. no fusion with the previous score is performed.

# 5 Experimental Set Up

We utilize The SEMAINE database, consisting of dialogues between a user and an agent in a Wizard-of-Oz fashion [6]. There are 5 agent characters with distinct personalities; cheerful Poppy, angry Spike, sad Obadiah, and sensible Prudence. Poppy and Prudence tend to draw the user into the positive-valence region of emotion compared to Spike and Obadiah. To promote positive emotion, we exclusively use sessions of Poppy and Prudence to construct the example database. The training set and test set comprise 29 (15 Poppy, 14 Prudence) and 4 (2 Poppy, 2 Prudence) sessions, respectively. We construct the example database exclusively from the training set, containing 1105 tri-turns.

We utilize the transcription and emotion annotation provided from the corpus as information of the tri-turns to isolate any recognition errors. We sample the emotion annotation of every dialogue turn into 100-length vectors. For the n-best filtering, we chose 10 for the semantic similarity constraint and 3 for the emotion.

# 6 Human Subjective Evaluation

A total of 50 queries are evaluated through crowdsourcing. The queries are presented in form of text, along with the responses from the baseline and proposed systems. The evaluators are asked to select the better response in terms of naturalness, potential emotion connection, and positiveness of elicited emotion. 50 judgements are collected per query. The final judgement of each query for each question is based weighted majority voting of the judgements.

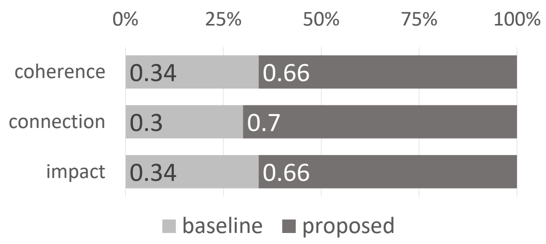


Figure 3: Subjective evaluation result.

Figure 3 visualizes the evaluation result. It is shown that in comparison to the baseline system, the proposed system is perceived as more coherent (66% of the time), having more potential in building emotional connection (70%), and giving a more positive emotional impact (66%). Furthermore, we observe that the queries where the proposed system wins have far stronger agreement than that where the baseline system wins (Fleiss' Kappa of 0.35 vs. 0.16). Higher agreement level suggests a stronger win, where bigger majority of the evaluators are voting for the winner.

# 7 Conclusions

We presented a novel attempt in eliciting postive emotional impact in dialogue response selection by utilizing examples of human appraisal in spoken dialogue. We augment the response selection criteria to take into account emotion similarity between query and the example query, as well as the expected future impact of the candidate response. Human subjective evaluation showed that the proposed system can elicit a more positive emotional impact in the user, as well as achieve higher coherence and potential emotional connection.

# Acknowledgement

This research and development work was supported by the MIC/SCOPE #152307004.

# References

[1] Marcin Skowron, Mathias Theunis, Sebastian Rank, and Arvid Kappas, "Affect and social processes in online communication–experiments with an affective dialog system," *Transactions on Affective Computing*, vol. 4, no. 3, pp. 267–279, 2013.

[2] Takayuki Hasegawa, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda, "Predicting and eliciting addressee's emotion in online dialogue.," in *Proceedings of Association for Computational Linguistics (1)*, 2013, pp. 964–972.

[3] James A Russell, "A circumplex model of affect.," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161, 1980.

[4] Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee, "Example-based dialog modeling for practical multi-domain dialog system," *Speech Communication*, vol. 51, no. 5, pp. 466–484, 2009.

[5] Nio Lasguido, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura, "Utilizing human-to-human conversation examples for a multi domain chat-oriented dialog system," *Transactions on Information and Systems*, vol. 97, no. 6, pp. 1497–1505, 2014.

[6] Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder, "The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *Transactions on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.