

ARGUMENTATION SYSTEM FOR INTELLIGENT ASSISTANTS USING FUZZY-BASED REASONING

T. KOIVUAHO, M. IBRAHIM, F. UMMUL, M. OUSSALAH

*Centre for Ubiquitous Computing, Faculty of Information Technology and
Electrical Engineering, University of Oulu.
OULU, 90014- FINLAND*

This paper addresses the issue of building intelligent assistant that is able to maintain sustainable conversation with the user while taking into account his emotional, personality aspects, and, at the same time, maintaining high level focus. The approach makes use four meta-features; namely, topic, emotion, personality and dialogue-act. A sequence-to-sequence recurrent neural network approach was used to learn answer prototypes from Reddit.com sport corpus, while an ANFIS based approach was developed to extrapolate from the limited configurations used by the neural network to various dialogue utterances.

1. Introduction

The growth of e-commerce as well as the multiplicity of the smart city projects has led to the development of online systems that allow corpora, authorities and SMEs to interact with their users and potential customers, efficiently with minimal operator intervention [9]. Watson et al. [15] investigated mobile marketing platforms that enhance consumers' relationship with a brand through text messaging, mobile advertisements, m-commerce and permission based marketing. Such marketing is found to be useful for companies to easily reach consumers at relatively low cost [10].

The emergence of (intelligent) conversational agents, also known as chatbots utilizing chat, messaging, or other natural language interfaces (i.e. voice) to interact with people, brands, or services in a bidirectional asynchronous context [8], is at the heart of the success story in e-commerce and e-democracy like applications. In this respect, as pointed out by Shopify [14], consumers can chat with company representatives, get customer support, ask questions, get personalized recommendations, read reviews, and click to purchase all from within messaging apps.

Despite the potential of chatbots, their developments are challenged by inherent barriers, which are often rooted back to the limitation of the software agent to comprehend human language. The inherent barrier to understand the motives and the feelings of the user / customer sometimes creates psychological to ensure full user satisfaction and continuation of services. Long conversations are hard to

automate and if in an open domain, the conversation can go in any direction [1]. In most of current systems, human intervention is still needed when things get complex [4]. Besides, since pioneer work of Turing in early fifties, the interaction between computers and humans via natural language is a topic that is extensively researched in information processing community, artificial intelligence and humanities and is a complex task. Besides, issues of cross-domain portability, evaluation and validation have been reported as immature in the field (e.g. Kuligowska [7]). Yet, as chatbots gradually start to expand to the messenger interface, a different research approach is required to quantify its acceptability. On the other hand, the identification of user's reasoning and rationality using standard natural language processing is at its infancy, and more research is still needed in order to open new horizons for chatbots.

This paper aims to contribute to the ongoing research in this direction where a fuzzy based approach is developed in order to leverage the capacity of chatbot to identify user's argumentation while accounting for his emotional state and feelings.

Strictly speaking, investigating recent works on the issue from both academia and commercial like applications reveals that the quasi-majority of such promising prototypes advocate the use of machine-learning like approach where a dedicated configuration of neural network learns the best sentences from a large dataset of question-answers corpus. For instance, Sutskever et al. [13] and Zhou et al. [17] promoted a sequence-to-sequence neural network, where a multi-layer model has been successfully used in translation task. Nevertheless, despite their limited success, such models often ignore the input and produce highly generic responses such as "I dont know what you are talking about" answer to be shortsighted and ignore their influence on future outcomes, which, trivially, negatively impact the ability of the underlying chatbot to calm users down, and built trust relationship as pointed out in [11]. This calls further research on the issue.

The key reasoning advocated in our methodology of handling is to distinguish five key patterns that are extracted from the user's textual statements, namely, personality, emotion, topical discussion, sentiment and dialogue act, using a combination of natural language processing, machine learning and commonsense-based reasoning. These patterns are next integrated into a fuzzy rule based system in order to yield the prototype of potential chatbot answer. The exact answer is randomly selected among those sentences belonging to the prototype class. Without loss of generality and for evaluation purpose, we deliberately restricted to the sport topic. Three different approaches using full rule-based system, deep learning and hybrid (rule-based and deep learning) are compared.

2. Methodology

2.1 General approach and motivation grounds

As pointed out in the introduction section, the general approach relies first on the identification of key discussion patterns, which are used as control variables to derive the answer prototype through a combination of fuzzy rule based and learning like approaches, and then randomly select a candidate answer from the class. The rationale behind the proposal is the following.

First, the fact that the system outputs a prototype answers instead of a single unique “statement” answer agrees with the socio-linguistic findings and commonsense reasoning that any statement can be equivalently conveyed through a set of fully equivalent and distinct other statements. For instance, the statement of agreement can be translated using statements like “OK”, “Sure”, “I agree”, “Fully agree”, etc. Besides, any attempt to force a unique single answer may create a state of boringness that would impact negatively the subsequent dialogue.

Second, the multiplicity of control variables employed in the sequel is motivated by the complexity of comprehending human language as well as the complexity of human emotion, which can change drastically with respect to tiny change of words. For instance, any miss-understanding of user language can trigger strong emotion. More specifically indicators on discussion topics (gathered using Dirichlet latent modelling), dialogue act (using corpus and support vector machine), five-trait personality score using the cumulative discussion of the user up to the current time, emotion as quantified using Linguistic Inquiry Word Count features, were employed as control variables.

Third, the use of fuzzy based reasoning [5] in the process of integration of these control variables is motivated by the inherent subjectivity and absence of accurate uncertainty model pervading these control variables. Therefore, a fuzzy inference based strategy was employed. On the other hand, interestingly, unlike commonly employed fuzzy inference system, the universe of discourses ascribed to the aforementioned mentioned variables are not all defined on interval scale, but some rather use only a nominal scale. For instance, both topical and dialogue act variables use only a set of words to define their universe of discourse. This brings our reasoning more close to Zadeh’s concept of “computing with words” [16]. More explicitly, the detail description of the features and control variables are given below.

2.2 Dialogue Act classifier

For the sake of simplicity and intuitive interpretation, we imitate the 42-dialogue act of the switchboard corpus [12]. Examples of dialogue act include Statement, Backchannel / Acknowledge, Opinion, Appreciation, Yes-No question, Yes answer, No answer). Non-verbal, uninterpretable and third-party talk labels were left out, as they were not seen to be relevant for our case, which leaves 39-dialogue act to be effectively considered. The classifier was trained using the Switchboard Telephone Speech Corpus by Linguistic Data Consortium [3] and then implemented a logistic regression for the purpose of classifying the 39 various

dialogue acts. The training and testing sets contained roughly 129 000 and 43 000 utterances, respectively. Following good accuracy rate obtained elsewhere using the same corpus, the features, extracted for each utterance, and employed in the above classification are the following: i) First 10 tokens, lemmatized, including punctuation, and padded with blank if fewer; ii) First 10 POS-tags, including punctuation, padded with blank if fewer; iii) Presence of a question mark; iv) Previous speaker; v) Previous dialogue act; vi) Predicate verb and respective POS-tag; vii) Subject and respective POS-tag; iix) Object and respective POS-tag. Lemmatization and POS-tagging were done using SpaCy such that the predicate, the subject and the object were determined using SpaCy's dependency parser. The macro accuracy of our classifier using the specified testing material is 0.77.

2.3 Emotion

We used WordNet-Affect, an extension of WordNet domains that concerns a subset of synsets suitable to represent affective concepts correlated with affective words. Especially, the affective concepts representing emotional state are individuated by synsets marked with the a-label *emotion*. Distinct labels were also attributed to those concepts representing moods, situations eliciting emotions, or emotional responses. It is organized as an xml file providing a tree in the set of words. This tree first describes the root of human behavior which can be broken into physical state, behavior, trait, sensation, situation and mental state, while emotions are considered part of the mental states. Divided into positive and negative emotions, the tree holds most of the words that can have an emotional meaning. This approach matches the words in the input text with these words in the tree, when matched with a high similarity, the module returns the emotion behind these words and then returns the root feeling whether it is positive or negative one. This occurs for every word in the input text, as the module tries to match every word out of the input to the tree in the domains files. The result will be for a set of different feelings, then by taking average, we can get a relative estimation about the actual feeling from the text.

The open source WNAffect (implementation of WordNet-Affect) tries to match the adjectives and nouns entered by the user to the wordnet-domains. JJ, an adjective and NN, nouns are called.

2.4 Personality

Taking advantage of the accumulated input text from same user, our purpose was to identify the user personality in terms of the five-personality model. The latter encompasses the following five personality traits:

- Extroversion, which indicates how the person can engage with the world. E.g., he may enjoy being with people and tend to be more enthusiastic, action-

oriented and like to assert himself. A high level of emotional stability means that the person is more confident and sure of himself with less self-doubt.

- Agreeableness, means how much the person is liked, respected and sensitive to others needs. It shows how the person is generous, helpful and friendly. It correlates in a weak way with extroversion.
- Conscientiousness, where people with high values in conscientiousness are likely to value order, duty, achievement, and self-discipline, and consciously practice deliberation and work towards increased competence.
- Openness to experience, which is the tendency of people to try new experiences and knowledges, usually linked to creativity and imaginative thinking.
- Neuroticism, which represents the tendency to experience unpleasant emotions such as anger, anxiety, depression and vulnerability. It is sometimes referred as the degree of emotion stability and impulse control.

We used an application developed by University of Sheffield in cooperation with University of Arizona that uses MRC Psycholinguistic Database, where the Linguistic Inquiry and Word Count (LIWC) is used to get emotions, thinking styles, social concerns, and even parts of speech. At each new utterance of the user, the application outputs the percentage of each of the five personality trait up to the current state of the human-agent interaction.

2.5 Topic

Latent Dirichlet Allocation (LDA) [1] is employed to detect the topics that are present at individual utterances or corpus. The LDA generative model assumes that documents contain a combination of topics, and that topics are a distribution of words; since the words in a document are known, the latent variable of topics can be estimated through Gibbs sampling. We used an implementation of the LDA algorithm provided by the Mallet package [6] adjusting one parameter ($\alpha \sim 0.30$) to favor fewer topics per document, since individual utterance updates tend to contain fewer topics than the typical documents (newspaper or encyclopedia articles) to which LDA is applied. Besides, in order to avoid subsequent unnecessary complexity burden while ensuring high interpretability of the results, we also reduced the size of the vocabulary of the words used by the LDA to the most frequent 500 words, excluding the stopword elements in the Reddit corpus. The algorithm is fully unsupervised, which means no human input is necessary only need a corpus of plain text documents. The purpose of using this model in our application is to identify underlying “topics” that user talked about with our chat agent. Here, we have used separate documents for each user, So that we can find about which topic they are discussing with our chat agent.

2.6 Prototype answers

Without loss of generality and avoiding the complexity of open debate, we restricted our analysis to sport field in Reddit corpus (www.reddit.com/r/sport). Due to its proven performance in other studies, we used neural machine translation model, seq2seq recurrent neural networks [17] which works with encoder decoder principle in which encoder generates “thought vector” which decoder then decodes into vector representation of words. Our neural network also uses attention mechanism in the decoder phase as well as beam search. Our neural machine translation model had 800 000 of sentence-translation pairs as its training material with the vocabulary size of 20 000. All the training material was extracted from reddit.com.

On the other hand, the five-meta features (word topic, personality vector, emotion and dialogue act) are extracted from the original (Reddit) sentence-inputs. This will enable our neural network system to learn the answer prototype according to various input configuration of the meta-features.

Beam search is used for getting best 10 outputs of the neural net instead of one, provided the outputs are sufficiently reliable, up to a threshold level on average classification accuracy obtained on training dataset, otherwise, the prototype will include a small number of instances (less than 10 answers).

Therefore, for each combination of the meta-feature vectors, a prototype answer, constituted of a class of up to 10 instances, is generated. A tensor implementation of the seq2seg recurrent neural network has been used.

3. Fuzzy based reasoning

Given that the meta-features vectors obtained during the prototype answer generation in previous step are far to be complete as many other vector configuration can be generated when using our utterance corpus as quite distinct from the Reddit corpus, therefore the question of which prototype answer to use is well open. For this purpose, and given the already proven results in terms of universal approximation results. This motivates our approach of using a fuzzy inference system as a universal approximator in order to extrapolate new unseen meta-feature configurations from known configurations. More specifically, an ANFIS based methodology [4] was employed. This allows us to learn and optimize the various parameters of the inference system.

In order to ascribe a fuzzy like quantification to each attribute, we considered the inherent classification score generated by the corresponding algorithm for each of

the meta-features. For instance in case of topic meta-feature, the LDA algorithm generates a score value for each word representing the topic, e.g.,

Topic A \rightarrow {Word 1 (0.6), Word 2 (0.4), Word 3 (0.5)}

Therefore, attributes of Topic A will be described, for instance, as

Topic A \rightarrow {Word 1 is High AND Word 2 is Medium AND Word 3 is High}.

Similarly, a configuration of meta-feature –personality such that:

Personality A \rightarrow {Openness (0.7), Conscientiousness (0.4), Extraversion (0.44), Agreeableness (0.2), Neuroticism (0.6)}

is translated into

Personality A \rightarrow {Openness is High AND Conscientiousness is Medium AND Extraversion is Medium AND Agreeableness is Low AND Neuroticism is High}

Similar reasoning applies to dialogue act as well, although we only restrict to the two classes (first and second ranked dialogue act class).

In the case of emotion meta-feature, the WNAffect only outputs the set of emotion words as traversed through WordNet hierarchical taxonomy. Therefore, all the generated words are considered equally high.

The exact boundaries as well as the number of partition is left open as part of the optimization-process of ANFIS system.

Therefore, the accumulated (meta) fuzzy rule looks like

If Topic is T AND Emotion is E AND Personality is P AND DialogueAct is D THEN Answer is A.

where the T, E, P and D are rather defined in a multidimensional space. While, the consequent part A is chosen to be crisp (non-fuzzy), which bring the reasoning close to Zero-degree Takagi-Sugeno fuzzy inference system. The detailed description of the system will be reported at other publications forums.

4. Implementation

ChatInstance-class is used for integrating the functional components of the system, and for maintaining the necessary data structures of a single conversation. ChatInstance implements getter-methods for the different features of the argument vector as well as connection to the NMT-model. ChatInstance is an access point for the user to use lower level functions. A top method called converse calls other methods responsible for getting features, which are: get_emotion, get_personality, get_topic, get_dialogueAct. A user interface is provided in Fig.1.

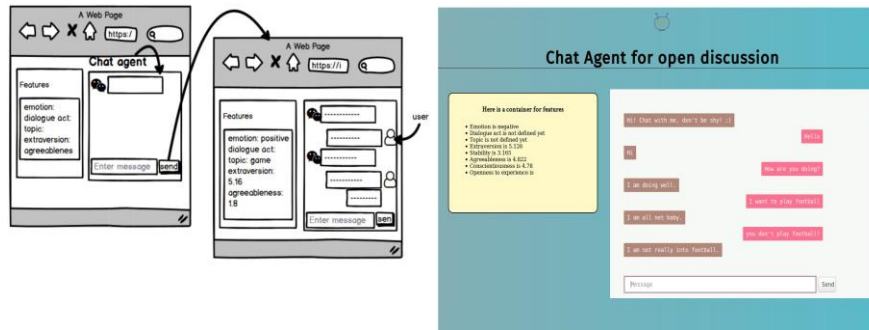


Fig. 1. User Interface

5 Conclusion

This paper describes a fuzzy based approach for intelligent assistant or chatbot design. Especially, four types of meta-features were employed and integrated as part of the methodology. This consists of topic (using LDA reasoning), emotion, personality and dialogue-act. A sequence-to-sequence recurrent neural network approach was used to learn answer prototypes from Reddit.com sport corpus, while an ANFIS based approach was developed to extrapolate from the limited configurations used by the neural network to various dialogue utterances.

Acknowledgment

This work is (partially) funded by the Marie Skłodowska-Curie Actions (645706-GRAGE)

References

1. Blei DM, Ng AY, Jordan MI, J Mach Learn Res 3: 993 (2003)
2. Britz, D. Retrieved at <http://www.wildml.com/2016/04/deep-learning-forchatbots-part-1-introduction/> (2016)
3. Calhoun, S., Carletta, J., Brenier, J. M., Mayo, N., Jurafsky, D., Steedman, M., & Beaver, D. Lang. Res. Eval. 44(4), 387 (2010).
4. Desaulniers, S, Retrieved from <http://www.cnb.com/2016/04/08/chatbots-rise-and-the-future-may-be-re-written.html> (2016)
5. Godfrey, J., & Holliman, E., Switchboard-1 release 2: Linguistic Data Consortium. SWITCHBOARD: A User's Manual (1997).
4. Jang JSR, IEEE Trans Sys Man Cybern 23: 665 (1993)
5. Klir G., Fuzzy Sets and Fuzzy Logic. Theory and Applications, Pren. Hall Pub. (1995)

6. McCallum AK, Available at: <http://mallet.cs.umass.edu>.(2002)
7. Kuligowska, K., *PCBR*, 2(02), 1 (2015).
8. Messina, C., Retrieved from <https://medium.com/chris-messina/2016-will-be-the-yearof-conversational-commerce-1586e85e3991#bsdskkyj> (2016)
9. Pavlou, P. A., Lie, T., & Dimoka, A., SSRN, <http://dx.doi.org/10.2139/ssrn.2380712> (2007)
10. Persaud, A., & Azhar, I., *Mark. Int. & Plan.* 30(4), 418 (2012).
11. Sordoni A., M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J. Nie, J. Gao, and B. Dolan, *Proc. North Amer. Chap. Ass. Comp. Ling.* 196 (2015).
12. Stolcke A. et al., *Comp. Ling.* 26(3), 339 (2000).
13. Sutskever I., O. Vinyals, and Q. V. Le, *Proc. Adv. Neur. Infor. Proc. Syst.* 27 (2014)
14. Shopify., Retrieved from <https://www.shopify.com/encyclopedia/conversationalcommerce> (2016)
15. Watson, C., McCarthy, J., & Rowley, J., *Int. J. Inf. Man.* 33(5), 840 (2013).
16. Zadeh L. *IEEE Tran. Circ. Syst. I: Fund. Theor. Appl.* 46(1), 105 (1999).
17. Zhou H., M. Huang, T. Zhang, X. Zhu, and B. Liu., arXiv preprint arXiv:1704.01074, (2017).