Smart Topic Detection for Robot Conversation

Elise Russell, Richard J. Povinelli, and Andrew B. Williams
Opus College of Engineering
Marquette University
Milwaukee, Wisconsin 53233

***Abstract*—In order for humanoid robots to have believable conversations with humans, the robots will need a reliable method for detecting the topics shared in the interaction to formulate a relevant response. This paper presents a novel application of intelligent indexing and ontology analysis for use in conversational topic detection for human-robot interaction. We evaluate a method for training on a corpus of transcribed phone conversations and using a concept association matrix to determine the strongest common-sense linkages to words in the conversation. This model is placed within the conversation and emotion interface of our humanoid robot, MU-L8, and tested with users. Evaluation is performed both computationally, with the corpus, and perceptually with users, and the promising results are presented in this paper.**

I. Introduction

This paper presents a new approach for conversational topic detection in human-robot interaction. The goal of the approach is to provide, for each utterance recorded and transcribed by the robot's interface, an appropriate English word ("reply word") for use in the robot's reply.

Such a reply word must be both semantically related to the content of the user's utterance and conceptually reasonable to use in this context. This pilot approach attempts to address both of these requirements by the use of a two-level process. A Latent Semantic Indexing (LSI) model is used in the first layer, and the second layer involves concept lookups in the association matrix of ConceptNet 5. The reply word returned by the system is then inserted into one of eight predefined reply templates and spoken aloud to the user. In this way, the approach attempts to create a natural flow of conversation as directed by the subject matter of the user's utterances.

II. Approach

Queries in this approach consist of conversational utterances recongized and text-transcribed by the humanoid robot MU-L8's interface [1]. Each utterance is filtered for stopwords and then converted to a bag-of-words vector, which is then processed by the two layers of the model.

In the first layer, the utterance's vector is used to obtain a list of semantically relevant words from an LSI model built from a text corpus. The corpus used is the Fisher English Training Transcripts corpus [2], which consists of 11,699 text-transcribed, natural language phone conversations between strangers about assigned topics, where each conversation lasted ten minutes. The topics were presented in the form of a prompt at the beginning of each call, and the topics are drawn from a set of 40, with titles ranging from "Pets" to "Life Partners" to "Issues in the Middle East."

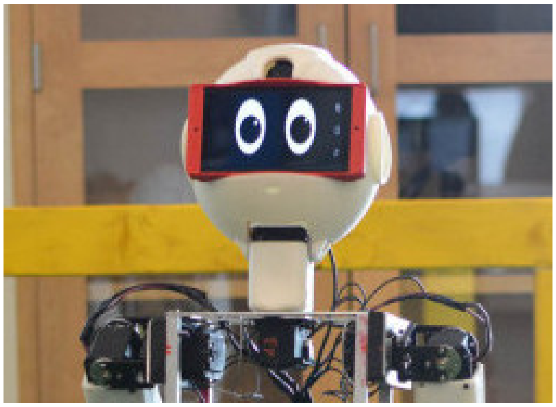

Fig. 1. The MU-L8 robot with conversational and emotional interface.

In the formation of the LSI model, the corpus was cleaned of normal stop words as well as broken words, tokens denoting noise or uncertain transcriptions, and back-channel utterances such as "um" and "uh-huh." All other words were stemmed using NLTK's Snowball Stemmer [3]. The LSI model was then formed by using the log-entropy transformation in the initial weighting step and by setting the final model dimensionality to $k = 100$, to reflect the number of topics that the participants were asked to talk about plus any additional topics that they happened to stray onto during their conversations.

When processing an utterance with this model, the utterance's vector is weighted and incorporated into the model, and the new LSI document vector is then used to find the top 25 words in the corpus most semantically related to the utterance.

The second level of processing requires the Association Matrix, which is a concept-by-concept matrix formed by performing Latent Semantic Analysis on a subset of ConceptNet 5, an online, graph-formed database of common-sense knowledge [4]. In the Association Matrix, each entry reflects the weight of the common-sense association between the two concepts. Such associations arise from relations such as "both cats and dogs are animals, have four legs, and have fur," and thus reflect a different type of association than the semantic, conversation-based similarities represented in the LSI model.

When processing an utterance, the 25 highly semantically related words from the LSI model are used, in combination with the words from the original tokenized utterance, to search

This research supported by NSF grant IIS-1427399.

the Association Matrix for the top ten concepts most associated with each word. Several concepts will appear on multiple top-ten lists, and for these, their association weights are added together. The ten concepts with the highest cumulative association weights are retained as possible reply words; however, concepts that are identical to words from the original utterance are passed over, because in practice this often results in the system parroting the user's words.

One of these top ten concepts is then randomly selected and returned as the final reply word. Random selection is used here because it produces a variation in the generality versus specificity of returned words, since the single concept with the highest cumulative association weight is usually the most abstract word. In interactions with actual users, this random selection was used in order to allow the robot more variability in its replies, which increased user engagement. However, in computational evaluations, randomness was dispensed with for consistency's sake and the single concept with the highest cumulative association weight was used.

The final step of processing an utterance consists of randomly selecting one of eight reply templates, including templates such as "It sounds like you're talking about [reply word]," or "Is [reply word] related to that?" The reply word is inserted into the template and the resulting sentence is spoken aloud to the user via the MU-L8 interface's text-to-speech function.

III. Evaluation

Evaluations were conducted both computationally, with reference to the corpus, and perceptually, in interactions with real users. Computational evaluations compared a test set of utterances with their system-generated LSI wordlists and final topic words for semantic similarity in relation to a set of external documents, and they found a significant topical relationship between them.

Perceptual evaluations were then performed to judge the system in its target scenario: conversations with human users. Ten participants were recruited, and during each evaluation appointment, the participant was seated facing the robot and had three short conversations with it. Each conversation was begun by the robot speaking an adapted topic prompt from the Fisher English corpus to the participant. It then collected the user's subsequent verbal utterances as text strings, until it had at least 15 words to submit to the topic-detection system. It received the reply word from the system and inserted it into a reply template, which was randomly selected from eight templates, and then spoke the resulting string to the user.

After seven replies, the conversation was ended, and two more conversations were conducted in the same way. At the end of the appointment, the participant filled out a questionnaire consisting of nine balanced Likert-style questions in which they rated the ability of the robot to make topic-relevant replies and to be an engaging conversation partner.

Average ratings were 3.95 for topicality of replies and 5.03 for user engagement, each on a scale of 1 (worst) to 7 (best). These results are shown in Figure 2. These average ratings suggest that the robot was perceived to have a middling ability to seem topically relevant and a positive ability to engage the user.

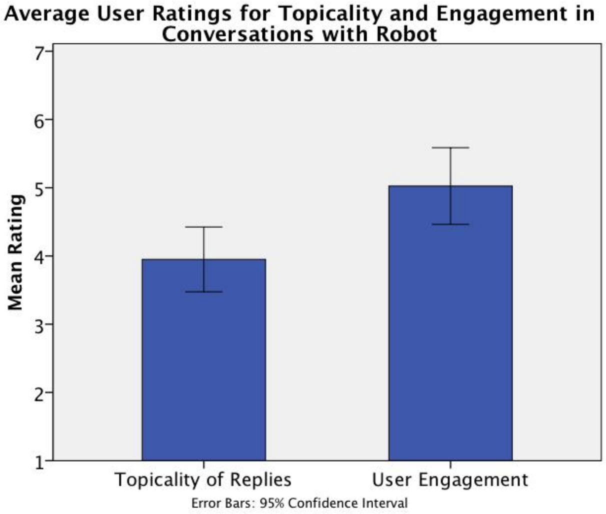

Fig. 2. Average user ratings for topicality of replies and user engagement from the online evaluation of the system.

Comments and suggestions were also collected, and users in general requested greater variety and clarity of replies, as well as better awareness of when a user was still speaking. Some users also attempted to ask the robot questions during the course of the conversation, which suggests that the robot's behavior led users to believe that its intelligence might extend to answering questions. It also suggests that the ability to discover a robot's opinions is valuable to users, and could be another line of study in future work.

IV. Conclusion

This pilot study demonstrated that the use of a Latent Semantic Indexing model in conjunction with ConceptNet 5's Association Matrix is justified in selecting semantically and conceptually appropriate conversational reply words. The system performs well under computational evaluation and decently well under perceptual evaluation. As the user limitations demonstrate, work still needs to be done on the use of these results in actual conversation and user modeling, especially in regard to the generality and grammatical context of the reply words.

References

[1] E. Russell and A. B. Williams, "Effects of SMILE emotional model on humanoid robot user interaction," in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction*, Portland, Oregon, March 2015.

[2] C. Cieri, D. Graff, O. Kimball, D. Miller, and K. Walker, "Fisher english training parts 1 and 2, speech and transcripts," *Linguistic Data Consortium, Philadelphia*, 2005.

[3] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.

[4] R. Speer, C. Havasi, and H. Lieberman, "Analogyspace: Reducing the dimensionality of common sense knowledge," in *AAAI*, vol. 8, July 2008, pp. 548–553.