

# An approach to Context-Related, Agent-Initiated Off-Activity Talk Suggestion in a Conversational Quiz-Game Setting

OMID MORADIANNASAB, s2764164

University of Groningen

March 20, 2015

## I. INTRODUCTION

Conversation systems embedded in virtual characters such as chat-bots and virtual assistants are created to equip automated agents with computational linguistic abilities through natural language processing techniques. This capability enables the agent to make task-oriented dialogues with the user and provide services to human around several applications in the fields of education, training, entertainment, help desks, or personalized services. Such systems are usually designed to interact with the user around well-defined domains. However, providing a talk by the agent out of the main activity (an off-activity-talk) can lead to the smoothness of the dialogue and thus raises engagement of the user and also their trust in knowledgeability and intelligence of the agent [2]. More trust and engagement of the user can help with achieving a long-term interaction which is a series of several non-repetitive encounters between the agent and a given user.

In the setting defined for this task (conversational quiz-game), the agent asks the user a multiple-choice question from an open domain. After the user selects one of the choices (which can be correct or not), the agent gives a verbal reaction. This reaction have to be related to the content of the previous interaction and should be a piece of information on the main subject matter extracted from the available online resources to confirm or give the correct answer. It can include a provision of some extra information. An example of a dialogue in this setting and some sample OATs to be uttered by the agent right after this dialogue are presented in the appendix. A dialogue manager component in a conversation system, which takes the major responsibility of controlling the conversation flow, is the consumer of the OAT suggestions. It can choose one of the suggested OATs from the ranked list right after the dialogue. Then, the agent follows the conversation with the next question, unless the user tries to take the initiative in asking the next question.

The objective of my thesis is to propose a tool/technique which automatically generates system-initiated, off-activity-talks (OAT) to be used in conversational quiz-game settings. The initial motivation of this thesis originates from a project aiming at sustaining a long-term interaction between a user (a diabetic child) and a robot through encouraging the user to follow the conversation by smoothening the dialogue on health-related topics [6]. However, this thesis does not take account of specific considerations on the range of the users' age and the result is not supposed to be bound to health-related topics. This makes the output of the current work appropriate to be used in a broader range of applications.

A suitable OAT, as defined in this thesis, is a verbal reaction which is bound to be contextually *relevant* to the content of the previous interaction. It should include a verbal confirmation or signification to the correct answer of the previously asked question. It can also include a provision of some added-information but not necessarily. This verbal reaction can consist of one or more sentences and can be extracted from the freely available online resources. The suggested OAT is not intended to be a social dialogue (small talk) and no information exploitation from the human user is expected, but it is meant to be an informative talk including one or more sentences to be uttered in a single turn by an artificial agent (i.e. a virtual character or a robot). This OAT can include some social features but it does not have to.

Grice's maxim of relation [5] implicitly defines *relevance* as a relation between a set of propositions and a discourse-topic but he mentions that formulation of this maxim brings about questions the treatment of which he finds "exceedingly difficult". [1] views *relevance* as "usefulness with regard to the conversational goals". [13] defines *relevant* information as that information which is worth hearer's attention. In order to assess this notion of *relevance*, I will follow the evaluation practice established in QA track at the Text REtrieval Conference(TREC) [3]. In TREC, as well as in my setting, both question and answer are based on facts and the goal is to provide the most *relevant* OAT. There is a major difference between *relevance* and *appropriateness*. In [7] the authors define *appropriateness* of an answer. In their definition "an evasive, misleading, or an honestly wrong answer" from a character can also be counted as an acceptable output. This definition covers a broader set of possible answers. It includes answers that might not be *relevant*. What is targeted here is the *relevance* and not the *appropriateness* of the OAT.

The first issue to tackle with is extracting candidate OATs for a question-answer pair from online resources. As it will be explained in III, iGNSSMM is planned to be employed for this purpose. Having a question-answer pair and the candidate OATs, a succeeding major sub-problem will be to select the *relevant* sentence(s) to be used as an OAT. This sub-problem can be looked upon as a classification task. Another view to this sub-problem can be as an information retrieval problem. We have a query including a question and a set of answers. This query will be compared to each element of the candidate OAT collection and the most *relevant* one will be returned. The problem can also be viewed as a (semi-)question answering system; that is, an information retrieval system with a natural language interface, to which the user provides the query as a question in natural language, and the system retrieves a few sentences from online or offline resources that are believed to be the most *relevant* to the query. The distinct property of the setting in the current work is that a limited number of answer choices including the correct answer is also available to the system which provides a broader view of the context. Given the agent knows the correct answer and some other potential answers, it is supposed to select an OAT which is the most *relevant* to this context.

To sum up, the final goal of this thesis is to create a tool/technique which, from a set of documents (fetched from online resources), seeks for the most relevant sentence(s) to the context and returns a ranked list of suitable candidates in order to suggest to the dialogue manager to help smoothen the dialogue between an artificial agent and a human user. Such off-activity-talks can be handcrafted in the knowledge base of a virtual character. The output of this work can be used to create (or enrich) such a knowledge base. However, this usage might not be feasible in open domain scenarios. That is why we put the ultimate aim of this thesis is to provide an open-domain solution for real-time OAT suggestion.

## II. RESEARCH QUESTIONS

The major research questions in this work are the following:

- What are the features of a suitable OAT?
- Does providing OATs via sentence selection lead to user satisfaction?
- What can be a good source for such sentences? Which online resources are available for this purpose?
- To what extent using online resources can help with non-repetitive dialogue with human?
- Which methods can be applied to achieve a high quality sentence selection? Which are more effective?
- Is it necessary, and if so, how far can we get with including semantic analysis in the architecture of the proposed tool?

## III. METHODS

Following is the provisional tools and methods to be applied:

- **iGNSSMM:** iGNSSMM is a system developed for topic graph extraction from web content with a web search engine as the back-end by Neumann and Schmeier[10]. Basically, iGNSSMM performs topic-driven search/exploration through online resources.
- Candidate sentences, which will be fetched from online resources through iGNSSMM, will be syntactically processed via The Stanford CoreNLP natural language processing toolkit [8]. This will help in pruning unfavorable syntactical structures (i.e. questions, imperative sentences and sentences including unresolved textual mentions).
- In order to represent the question-answer pairs as queries to the system a simple POS-based approach will be defined. Keyword extraction is necessary to feed iGNSSMM, even though Lucene searching tools can also deal with the whole question-answer expression as the query. This simple POS-based approach will be the baseline, however by applying more syntactic and semantic analysis over the inputs, the author hypothesizes the possibility of defining more efficient query representations.
- In order to rank the candidate sentences, statistical techniques for document retrieval implemented in Lucene Indexing and searching tools will be employed and compared [4]. These techniques are based on term frequencies and vector representations of documents.
- In a further step a semantic analysis will be employed to deepen the understanding of the context in order to make a better matching between candidate OATs and the context. This semantic analysis can possibly range from employment of an open domain ontology (i.e. WordNet [9]) to analysis of the semantic roles of the elements [11] of the query expression. In case of semantic role labeling, the process can be possibly applied to the query and documents in order to check the overlaps of semantic roles and their relations between the query and the documents. This can possibly be done in either of the following ways: first, by taking advantage of boosting methods of Lucene to increase the weights of the keywords in a direct correspondence to the amount of information they carry in the query. Second possible

strategy could be defining a concurrent semantic ranker in parallel with the Lucene-based one and using the weighted sum of the two rankers as the final score for each candidate OAT.

#### IV. EVALUATION STRATEGY

Not all off-activity-talks in a conversation are equally satisfactory. The aim of the evaluation process is to measure how satisfactory the suggested OATs of a model are. For this purpose we can intersperse the suggested OATs in the dialogues of the conversational system proposed in [6]. The logged dialogues of that work plus the interspersed OATs will be exposed to expert (or possibly non-expert) raters. They will be asked to judge the *relevance* of each of the OATs. For this purpose, following TREC evaluation practice, a binary relevance scale (relevant, not relevant, don't know) combined with some relevance criteria will be used. The rate of the relevant OATs will be reported as a means to compare the performance to the baseline. The difference between the scores of the baseline and the proposed model will be tested for significance.

The simplest baseline for this work can be defined as iGNSSMM on its own. However, the methods introduced in III will be combined with a stack-like architecture and the addition of each method is hypothesized to result in an improvement of the performance. Thus, an alternative evaluation strategy can be assessing the tool in an incremental way, reporting the effect of each method on the performance of the tool.

In order to measure the reliability of the assessments, the amount of agreement between raters will also be reported (e.g. via kappa statistics). This value will represent the consistency among the raters.

It might also be feasible to take a step further and ask the raters to judge the *relevance* using a scale similar to the one proposed in [12]. In this case, the average *relevance* of each model can be reported as a performance measurement.

#### V. PROVISIONAL TIME SCHEDULE

Task	Due Date
Literature Review	May 1st, 2015
Baseline Implementation	June 1st, 2015
Final version Implementation	July 1st, 2015
Evaluation	August 1st, 2015
Conclusion and Finalizing the thesis	August 15th, 2015

#### REFERENCES

- [1] Jonathan Berg. The relevant relevance. *Journal of Pragmatics*, 16(5):411–425, 1991.
- [2] Justine Cassell and Timothy Bickmore. External manifestations of trustworthiness in the interface. *Communications of the ACM*, 43(12):50–56, 2000.

- [3] Hoa Trang Dang, Diane Kelly, and Jimmy J Lin. Overview of the trec 2007 question answering track. In *TREC*, volume 7, page 63, 2007.
- [4] Brian Goetz. The lucene search engine: Powerful, flexible, and free. *JavaWorld*. Available <http://www.javaworld.com/javaworld/jw-09-2000/jw-0915-lucene.html>, 2000.
- [5] H Paul Grice. Logik und konversation. 1979, pages 243–256, 1975.
- [6] Ivana Kruijff-Korbayová, Heriberto Cuayáhuatl, Bernd Kiefer, Marc Schröder, Piero Cosi, Giulio Paci, Giacomo Somnavilla, Fabio Tesser, Hichem Sahli, Georgios Athanasopoulos, et al. Spoken language processing in a conversational system for child-robot interaction. In *WOCCI*, pages 32–39, 2012.
- [7] Anton Leuski, Ronakkumar Patel, David Traum, and Brandon Kennedy. Building effective question answering characters. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 18–27. Association for Computational Linguistics, 2009.
- [8] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014.
- [9] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [10] Günter Neumann and Sven Schmeier. A mobile touchable application for online topic graph extraction and exploration of web content. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*, pages 20–25. Association for Computational Linguistics, 2011.
- [11] Martha Palmer, Daniel Gildea, and Nianwen Xue. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103, 2010.
- [12] Ronakkumar Patel, Anton Leuski, and David Traum. Dealing with out of domain questions in virtual characters. In *Intelligent Virtual Agents*, pages 121–131. Springer, 2006.
- [13] Teun A Van Dijk. Relevance assignment in discourse comprehension. *Discourse processes*, 2(2):113–126, 1979.

## VI. APPENDIX

Following is an example of a dialogue in the defined setting :

- {Agent} here's the question: Which of these breeds is most commonly used as a police dog?
- {Agent} the available options are: one, Poodle. two, German shepherd. three, Chihuahua. four, Shihtzuone.
- {Agent} your answer please?
- {User} German shepherd.
- {Agent} very good.

Sample OATs to be uttered by the agent right after this dialogue:

1. The German Shepherd Dog is hailed as the world's leading police, guard and military dog, however, this dependable breed is more than its 9-to-5 job. Derived from the old breeds of herding and farm dogs, the first German Shepherd Dog exhibited in America was in 1907.
2. There are two breeds of dogs which are most commonly utilized in police work, namely the German shepherd and the Belgian malinois.
3. Considered one of the most intelligent and obedient breeds of dog, the German shepherd is easily the most commonly used dog in police work.
4. Most people can easily spot a German Shepherd and associate with police dogs.
5. The German Police Breed is a German shepherd as they are used as police dogs in many countries , and they are German.
6. The German Shepherd is one of the most renowned breeds of dog there is.

Any one of the items in the list above can be chosen by the dialogue manager to be uttered after the aforementioned dialogue.