Knowledge-Based Systems 106 (2016) 116–124

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Grounding the detection of the user's likes and dislikes on the topic structure of human-agent interactions

Caroline Langlet, Chloé Clavel*

LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

ARTICLE INFO

Article history:
Received 18 November 2015
Revised 17 May 2016
Accepted 18 May 2016
Available online 21 May 2016

Keywords:
Human-agent interaction
Sentiment analysis
Embodied conversational agent

ABSTRACT

This paper introduces a knowledge-based system which grounds the detection of the user's likes and dislikes on the topic structure of the conversation. The targeted study is set in a human-agent interaction with the aim to help the creation of dialogue strategies of an agent based on the user's interests. In this paper, we first describe the system based on linguistic resources such as lexicons, dependency grammars and dialogue information provided by the dialogue system. Second, we explain how the system merges its outputs at the end of each topic sequence. Finally, we present an evaluation of both the linguistic rules and the merging process. The system enables a better identification of the target of the user's likes and dislikes and provides a synthetic representation of the user's interests.

1. Introduction

Embodied conversational agents (ECAs) are virtual characters able to engage conversations with human users in various interaction scenarios. They can play different roles – such as companion, tutor or assistant – and can engage either in long-term or short-term relationships with the user. Within these interactions, the management of the affective component of the conversation is crucial: in order to generate relevant affective answer with the agent, it is important to be able to detect the user's affective reactions expressed with non-verbal signals (facial and bodily expressions, acoustic features, etc.) or with verbal content. Regarding the detection issues, most of the solutions that are proposed in the ECA community rely on non-verbal clues [1], even though there is an increasing number of studies dealing with multimodal sentiment analysis using both linguistic, acoustic and visual cues [2,3]. Such a trend is due to the increasing number of web video reviews available under platforms such as Youtube. Beside the existing literature on multimodal sentiment analysis, we have found only two studies, [4] and [5], using a module for the detection of verbal-centred sentiment-related phenomena in human-agent interactions: the studies describe an implementation of "How was your day" Companions integrating a Sentiment Analysis module grounded on the theoretical model provided by [6] and initially designed for the analysis of non-conversational texts. These studies have the merit to show the importance of the sentiment-related detection during a conversation between a virtual agent and a human user. However, as explained in Clavel et al [7], the systems described by [4] and [5] are learned with non-conversational texts and not properly designed to be fit for the conversational context and human-agent interaction.

Indeed, designing a system able to detect verbal expressions of sentiment during face-to-face and speech-to-speech interactions between a human user and a virtual agent requires to deal with various issues. Firstly, the system has to be suited to the agent's communication goals. For example, while in some applications, such as a companion robot, it will be necessary to detect verbal expressions referring to affective states, such as "I'm sad", for other interaction scenarios, such as recommendation in retrieval agent[1], it will be better to focus on the verbal expressions referring to axiological evaluations, such as "This book is a master-piece". Secondly, we need a suitable method for the conversational speech. Each verbal expression of sentiment needs to be considered not individually, but according to the dialogic context. In this way, we have to give attention to dialogue features such as adjacency pairs, topic structure and information progression. In particular, we have to consider that the sentiment-related expressions can be derived, in a pragmatic point of view, from the solicitation or the meaning of the agent's previous utterance. For example, in the following adjacency pair, Agent: "Do you like this painting?" – User: "Yes", the sentiment-related phenomena is entirely grounded on the agent's utterance.

* Corresponding author.
E-mail address: chloe.clavel@telecom-paristech.fr (C. Clavel).

[1] For example, the virtual agent designed in the European Project Aria-Valuspa http://aria-agent.eu/.

http://dx.doi.org/10.1016/j.knosys.2016.05.038
0950-7051/

In [8], we address some of these issues: the selection of the relevant sentiment-related expressions, the likes and the dislikes, and the integration of a first level of dialogic context, the adjacency pairs. We delimit and specify the linguistic phenomenon to detect by focusing on one specific aspect required by ECAs for modelling social relationships: the user's interests that are given by the expressions of the user's likes and dislikes in the verbal content. As we did not find any linguistic model describing likes and dislikes expressions, we rely on the Martin and White model [9]. This model provides a description of sentiment-related phenomena called *Attitudes* which comprises three subcategories: the affects, the appreciations and the judgments. In this model, we consider the likes and dislikes as a subcategory which overlaps the affects, the appreciations and the judgments[2]. The likes and dislikes are considered so far in a binary way and we do not deal with graduation or modality issues.

Then, we proposed a system relying on a rule-based and bottom-up process using symbolic representation of the structure of sentences. The system processes jointly each adjacency pair, which allows us to model the agent's utterances in order to help the detection of the user's sentiment-related phenomena. Even though the system presented in [8] shows interesting results, a larger conversational context needs to be considered for improving the efficiency of the system [10]. In [8], the adjacency pairs were processed without considering neither the progression of the conversation nor the topic structure of the ongoing dialogue. A contextualisation of the likes and dislikes with respect to the topic structure of the conversation – defined by the interaction scenario – can be very helpful for the detection system. Besides, it contributes to build a user's profile that can be used to improve the social relationship of the agent towards the user and to foster the user's engagement.

In this article, we develop a knowledge-based and semantic-based system which grounds the detection of the user's likes and dislikes on the topic structure of the conversation. The system embeds linguistic resources such as lexicons, dependency grammars and dialogue information provided by the dialogue system. The paper is organised as follows: first, we present the theoretical model and the related work (Section 2). Then, we introduce an annotation schema of the topic sequences in human-agent conversations (Section 3). In Section 4, we provide a detailed description of the system. Finally, we present the evaluation of the different levels of the system process and discuss the final outputs of the system on 10 sessions from the Semaine corpus (Section 5).

2. Related work: sentiment analysis (SA) in interactions

A large number of studies in SA uses classical machine learning algorithms for the classification of words, sentences or texts [11–14]. Whether supervised or not, these approaches show interesting results. However, they do not properly manage the detection of the target and the source. In order to prevent these issues, some works provide fine-grained approaches dealing with the sentence structure either in rule-based, machine learning or hybrid algorithms. By taking into account the syntactic and semantic structure of utterances, such approaches can deal with the compositionality principle[3] and improve the calculation of the polarity. While in [15,16] and [6], the authors define semantic rules grounded on the hierarchic relations within the sentences, in [17], they train a recursive neural network on a sentiment treebank. In both cases, they can deal with polarity reversal and propagation. These approaches also appear well-suited for the identification of the source and the target. In [18] and [19], the authors design a hybrid method, jointly using extraction patterns and conditional random fields to maximize the identification of the source and the target. Other hybrid methods such as Sentic computing [20] have also shown convincing results.

As explained in Section 1, designing a method able to detect the user's expressions of sentiment requires to tackle various issues, such as the selection of the relevant expressions of sentiment and the consideration of the dialogue context. Few studies design or use SA methods for human-agent interaction purpose and the methods they provide do not differ from the methods used in written and non-conversation texts. In order to provide a first SA method suited to the interaction issues, we choose to use a rule-based and bottom-up process using symbolic representation of the structure of the sentence. The interaction context can thus be integrated using these rules in order to help the detection system. In the next paragraphs, we examine two notions frequently used in order to model the conversation and the interaction in the research field of ECA: the *liking dimension* and the notion of *topic*, and we discuss how these notions and SA can benefit from each other.

2.1. Modelling social relationship through the liking dimension

In the research field of ECA, several studies – *e.g.* [21] or [22] – aim to design applications involving social relationships between the ECA and the human user. They use different dimensions in order to model relationships. The liking dimension is one of the most widely used. The definition of this concept is frequently grounded on the Heider's *Balance Theory* [23] which is concerned with "the way relations among persons involving some impersonal entity are cognitively experienced by the individual" [24]. It considers the relations between a person P, which is the focus of the analysis, another person O and an entity X, which can be either an event, a process or a physical object. These relations can be modelled by a triangle where the vertices are P, O and X. The edges are the liking relations between them: P's liking toward O, P's liking toward X, and O's liking toward X. For Heider, "a balanced state exists if all three relations are positive in all respects or if two are negative and one positive" [23]. Thus, if P likes X, O likes X and P likes O, then the state between P, O and X is balanced. Similarly, if both P and O dislike X, and P likes O, then the state is balanced. Otherwise, the state is unbalanced and that can be a source of strain for P. As the theory assumes that to reduce the tension, people tend to keep balanced states, some social agent systems [21], [22] define scenarios where the user (P)'s liking toward the agent (O) is determined by the user and the agent's feeling toward an impersonal entity (X) . If the agent and the user share the same feeling toward the impersonal entity (X), then the user's liking toward this agent is positive, otherwise it is negative.

The detection of the user's likings toward different impersonal entities will contribute to help the agent to balance the states and reduce tensions. In such a context, the analysis of the user's verbal content has a key role as a major source of information of the user's likings and for the management of long-term relationships.

2.2. Topics and user's interest and engagement

As explained by Brown et al. [25], different definitions are attributed to the notion of topic. The notion is defined at either the sentence level or the discourse one. At the sentence level, the notion has been widely used for the description of the sentence structure, by establishing a distinction between the *topic* and the *comment*: "the speaker announces a topic and then says something about it" [26]. In the discourse studies, the notion is used for

[2] More details about the Martin and White model and the way we use it can be found in our paper [8].

[3] This principle defines the meaning of a sentence as built by the meaning of its constituents.