Knowledge-Based Systems 000 (2016) 1–16

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Context-Specific grounding of web natural descriptions to human-centered situations

Rui Liu, Xiaoli Zhang*

Department of Mechanical Engineering, Colorado School of Mines, Golden, CO 80401, USA

ARTICLE INFO

Article history:
Received 4 March 2016
Revised 25 July 2016
Accepted 29 July 2016
Available online xxx

Keywords:
Assistive robot
Decision making
Human-centered situation
Web information retrieval
Environmental context

ABSTRACT

Human-centered situation, which describes the surrounding world of a person, indicates his undergoing activities. Understanding of human-centered situations helps an assistive robot with its decision making. Existing methods, such as learning from human demonstration, are economically expensive, time-consuming, and have limited scalability. To address this problem, we developed a web-to-situation (W2S) method with which web natural descriptions are grounded into human-centered situations in a context-specific manner. By comparing the learned knowledge from the web and the survey, we proved that W2S is effective in extracting reliable knowledge in an efficient and low-cost manner. By implementing the W2S-retrieved knowledge in 60 web-collected situations and 60 real life situations, we proved that W2S is effective in situation understanding. Given that the web contains huge amounts of information, W2S is expected to effectively scale up a robot's knowledge.

1. Introduction

Humans' surrounding world, including surrounding objects and environmental conditions, is defined as "human-centered situation" ("situation" for short), which indicates what a human is doing. For example, when a person is *in the kitchen*, standing near a *coffeepot* and grasping a *cup*, it is likely that he is preparing a drink (defined as "drink" situation) [1,2]. Understanding of human-centered situations is essential for robot assistance in daily life [3,4].

Currently, some research has been finished for human-centered situation understanding. Scene-setting knowledge like "cup is in the kitchen" was predefined to describe human-involved situations for robots [5]. Activity-related situation knowledge like "handshaking" was learned from human demonstrations in either a one-way manner (a human teaches, and a robot observes) [6,7] or a two-way manner (a robot observes and meanwhile interacts with a human) [8]. Social manners in a human-involved situation were learned by using crowdsourced surveys [9,10].

Though these methods effectively collect the robots' knowledge and support their understanding towards human-centered situations, their practical applications for human behavior understanding encounter two challenges. First, the existing knowledge collection manners are expensive, time-consuming, and cannot easily scale up to cover a large number of human activities. Collecting a large amount of knowledge by predefinition, demonstration, or through a crowdsourced survey is with high costs of labor, money, and time [11,12], which constrain the knowledge scale [13]. Second, the learned situations are not truly human-centered because the human factor is inadequately considered, leading to inaccurate situation understanding. In a human-centered situation, human factor is an important component since it determines the human-situation interaction manner and interaction type (the undergoing activity) [14,15]. The consideration of only scene conditions (e.g., object location and object spatial relations) without exploring the human-scene relation (e.g., the object's relative distance to the human) may cause unsatisfied situation understanding and robotic assistance [16].

Given the above problems, in this paper we developed a novel context-specific web-to-situation (W2S) method in which a large amount of web natural descriptions were explored and grounded into human-centered situations. A human-centered situation in W2S means a human's instant surrounding environment during an activity performance. Given that the Internet contains a large number of natural descriptions about human-centered situations, it is possible for a robot to acquire human-centered knowledge for various situations. In addition, web descriptions could be obtained with a very low cost and be processed in a short time, especially when compared with learning from human demonstrations. To consider human factors, our method includes four important factors, including undergoing human activity such as 'drinking', human-correlated environmental information such as "hot day" and "before a person's bedtime", human-correlated objects

* Corresponding author. Fax: +1-303-273-3602.
E-mail addresses: rliu@mines.edu (R. Liu), xlzhang@mines.edu (X. Zhang).

http://dx.doi.org/10.1016/j.knosys.2016.07.037
0950-7051/

such as “cup in one’s hand”, and objects’ temporal/spatial relations such as object flow “cup → coffeepot” and “cup → in office”. The reason for exploring the environmental context is that context normally indicates or even triggers a human-centered situation [15,17,18]. For example, “in office” usually indicates a “work” situation, and ‘high room temperature’ usually triggers a ‘drink’ situation [6]. Exploration of object involvement is used because objects are the connection between the real world and a human’s mental plan. Furthermore, a human’s object operation directly describes the human-centered situations [19,20].

In our method, situation knowledge includes object/context involvement, object-object temporal relation, object-context co-appearance relation, and context-context co-appearance relation. They are mainly extracted from texts on the website wikiHow [21] and the open information databases ConceptNet5 [22] and WordNet [23]. WikiHow contains about 180,000 activity-related scenario descriptions covering nearly every aspect of our daily life, which makes it a good information source for situation-related knowledge learning [24,25]. ConceptNet5 is a semantic network, which provides a large amount of relations among things in the real world [26,27]. WordNet is an English lexical database [28], and is usually used as knowledge support for robots to perform automatic text analysis [27]. In this paper, we aimed to validate the following hypotheses.

(1) This novel W2S method is effective in grounding web natural descriptions into human-centered situations. With the grounded knowledge, human-centered situations could be effectively understood.
(2) The influence of environmental context towards humans’ activities could be measured. The context involvements in situation representation could improve the accuracy of situation understanding.
(3) The proposed semantic analysis method could evaluate the situation-involvement degree of the knowledge entities (i.e., objects and environmental context) and their relations. Based on the involvement degree, the key features including objects, environmental context, and their mutual relations could be extracted to accurately represent a situation.

2. Related work

(1) Web information retrieval for robots’ knowledge learning

Most previous work related to web information retrieval focuses on generation of robot-centered instructions for task execution, however the research in this paper focused on human-centered knowledge for situation understanding. In other words, existing works are devoted to adapting task variety in automatic robotic execution, while this work is devoted to adapting human variety in intuitive human-robot interaction. Typical existing works are as follows. Beetz and Tenorth explored daily life guidance websites like eHow [29] and online shopping websites like GermanDeli [30] to learn object physical properties such as shape, appearance, and object grasping pose, which helped a robot to understand its working space in task execution [31,32,33]. Samadi and Kollar explored the general webpages to learn the probabilistic relations between objects and locations in our daily scenes, assisting a robot for object finding [34,35]. Saxena’s group explored the interior design website Google 3D Warehouse [36] to collect object-operation-related information like object trajectory [37] and object physical status [2], enabling a robot to understand object arrangement in our daily life.

(2) Web information retrieval methods

In many existing web-related information retrieval methods, keyword detection and the Part-of-Speech (PoS) tagging methods were used for entity/relation extraction [38,39,40]. For example, according to the keywords “cup, kitchen” and keywords’ PoS tags “NN, NN”, a spatial relation between cup-kitchen was extracted from the sentence “enjoy a cup of coffee in kitchen” [40]. However, due to the ignorance of semantic analysis of sentence meanings in previous works, the method basing on keyword/PoS is unreliable that the entity/relation is highly likely to be misinterpreted [39,40]. For example, sentences like “To help celebrate and enjoy the Rugby World *Cup* 2015, Ghaf *Kitchen* have created the perfect menu for you to enjoy with friends and family” and “in grams instead, using a *kitchen* scale, rather than measured in *cups*”, …}, the entity “cup, kitchen” and the spatial relation “cup-kitchen” could be incorrectly extracted but actually the entity and spatial relation do not exist. This was because semantic features like “a cup of”, “coffee”, and “in kitchen” were ignored.

(3) Situation modeling methods

A statistical learning technique was adopted in many existing methods [34,35,41]. Based on the occurrence frequency, the object-location probabilistic distribution was calculated to describe a situation. It is reasonable, yet may not be adequate enough, for accurate situation understanding [2,41]. Generally speaking, the frequently-appearing features do not necessarily mean they are helpful in distinguishing a situation from others as they may also frequently appear in other situations. For example, in both situations “dish washing” and “hand washing”, the object “tap” is frequently involved. With a statistical learning method, “tap” is likely to be counted as an important feature for describing any of the two situations “dish washing, hand washing”. However, given that “tap” cannot be distinguished from either of the two situations, “tap” is not informative enough to describe either of the situations “dish washing, hand washing”. Therefore, an accurate situation-modeling method is in urgent need.

(4) Our previous work

In our previous work [2], we have validated the feasibility of collecting human-centered situation knowledge from the web. Compared with the previous feasibility study, the W2S method introduced in this paper focuses on solving the open problems, including lack of semantic analysis in information retrieval, lack of consideration of feature dependencies in situation modeling, and lack of advanced learning method, which are critical to advancing this web-supported learning method toward practical applications. Instead of only focusing on keyword occurrences in about 1470 webpages [2], a semi-supervised semantic method is adopted to analyze the keywords’ situation involvement in about 50,000 pages. In addition, instead of using a Naïve Bayes Network algorithm to simply assume the situation features are mutually independent, a MRF algorithm is adopted to describe the mutual dependencies among situation features. Moreover, instead of simple statistical learning, advanced structural learning is conducted to extract representative features for modeling accurate human-centered situation.

The summarization of the above related work is shown in Fig. 1.

3. Web-to-situation (W2S) framework

The W2S’s framework is shown in Fig. 2. Input of this method is the web natural descriptions which likely depict the targeted human-centered situations. The output is the knowledge of the grounded human-centered situations. In web description collection, natural language descriptions about situations such as “drink” and “dish washing” are collected from application-related websites such as wikiHow. The piecemeal knowledge, such as “drink could be expressed as {drink, drinking, drunk, drinks}” and “drink is related to container cup”, is collected from conceptNet5 and Word-

Please cite this article as: R. Liu, X. Zhang, Context-Specific grounding of web natural descriptions to human-centered situations, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.07.037