Categorizing Events using Spatio-Temporal and User Features from Flickr

[Abstract] *

Steven Van Canneyt[1], Steven Schockaert[2], Bart Dhoedt[1]

[1] Ghent University – iMinds, Belgium
[2] Cardiff University, United Kingdom

steven.vancanneyt@ugent.be, schockaerts1@cardiff.ac.uk
bart.dhoedt@ugent.be

ABSTRACT

We introduce a method for discovering the semantic type of events extracted from Flickr, focusing in particular on how this type is influenced by the spatio-temporal grounding of the event, the profile of its attendees, and the semantic type of the venue and other entities which are associated with the event.

1. INTRODUCTION

Even though the problem of event detection from social media has been well studied in recent years, few authors have looked at deriving structured representations for their detected events. We envision the use of social media for extracting large-scale structured event databases, which could in turn be used for answering complex (historical) queries. In this paper, we study how the semantic type of events can be extracted from social media, as a first step towards automatically extending and creating structured event databases.

2. METHODOLOGY

Evidence about the semantic type of an event can be obtained by analyzing social media documents, such as Flickr photos taken at the event, which we consider in this paper, or tweets that have been sent about the event. In particular, we represent an event as a set of social media documents related to that event, together with its associated characteristics. A set of social media documents related to an event may for instance be automatically extracted from social media or may be extracted from existing event databases such as Upcoming. Most initial work about discovering the semantic types of events only used textual information, which may lead to poor performance when the text is noisy (e.g. in some Twitter posts) or absent (e.g. in some Flickr photos). However, social media documents also contain metadata which provide an indication about the spatio-temporal and attendees features of an event. The hypothesis we consider in this paper is that in many cases the event type can be discovered by looking at characteristics, such as timing, the location of the event or properties of attendees, which can be readily obtained from social media sources. For example, when an event occurs on a Saturday inside a sport complex and it has basketball players as main actors, it is very likely that this event is of type 'basketball game'.

Even though our methods can be applied more generally, we restrict ourselves in this paper to experiments with Flickr photos. In particular, the spatio-temporal grounding of the event, the profile of its attendees, and the semantic type of the venue and other entities which are associated with the event are estimated using its associated Flickr photos, and these characteristics are then used to describe the event. To estimate the type of a given event, we use an ensemble of classifiers, one for each of the considered descriptors.

3. EVALUATION

We consider two use cases. First, the trained classifiers are used to analyze in detail to what extent our methodology is able to discover the semantic type of known events that have no associated semantic type. This is useful, for example, to improve existing event databases such as Upcoming, for which we found that about 10% had no known type. When using our methodology instead of a baseline which only uses the text of the Flickr photos related to an event to estimate its semantic type, the classification accuracy increases significantly. We observe that considering the type of the events visited in the past by the participants of the event lead to the most substantial improvement over the baseline approach. The classification performance is further improved when the types of known events organized nearby the event, the textual content of the photos taken in the vicinity of the event, and the time and date of the event are considered. Second, the model is used to estimate the semantic type of events which have been automatically detected from Flickr and are not mentioned in existing event datasets, which could substantially increase the applicability of existing methods for automated event detection.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] S. Van Canneyt, S. Schockaert, and B. Dhoedt. Categorizing events using spatio-temporal and user features from flickr. *Information Sciences*, 328:76–96, 2016.

---

*A full version [1] of this paper has been accepted and will be published (Jan. 2016) in Information Sciences.