Extracting Criminal-Related Events from Arabic Tweets: A Spatio-Temporal Approach

Feriel Abdelkoui, Constantine 2 University, Constantine, Algeria
Mohamed-Khireddine Kholladi, Echahid Hamma Lakhdar University, El Oued, Algeria

ABSTRACT

Recently, Twitter as one of social networks has been considered as a rich source of spatio-temporal information and significant revenue for mining data. Event detection from tweets can help to predict more serious real-world events. Such as: criminal events, natural hazards, and the spread of epidemics. Etc. This paper deals with event-based extraction for criminal incidents from Arabic tweets. It presents a framework that supports automated extraction of spatial and temporal information from tweets. The proposed approach is based on combining various indicators, including the names of places and temporal expressions that appear in the tweet message, related tweeting time, and additional locations from the user’s profile. The effectiveness of the system was evaluated in term of recall, precision and f-measure.

KEYWORDS
Arabic Language Processing, Gazetteers, Information Extraction, Processing Tweets, Spatiotemporal Information

1. INTRODUCTION

In recent years, microblogging, as a form of social media, has rapidly increased the attention of the general public as a mechanism for news broadcasting, expressing opinions and promoting contacts between people.

Diffusion of information is a key axis in the prevention of criminal events with terrorists’ acts. The aim is to make information largely shared by scientists, best used by professionals and clearly understood by the public.

Today, twitter has become one of the major prevalent social networking and micro-blogging services, it allows 140 maximum characters for each tweet and enables more than 250 million users to share real-time events happening around the world every day (Ozdikis, Halit, & Pinar, 2013). One of the most significant benefits of Twitter is the rapid transfer of information via the Internet (Lau, 2014). The research results indicate that spread of news is often posted on Twitter first before being disseminated by public media. Other important advantages of Twitter are that it is accessible real-time and provides Real-time detection. Tweets can be used to extract not only temporal information, but also for geolocate real time incidents. Approximately 1% of all tweets has GPS coordinates

DOI: 10.4018/JITR.2017070103
and is expressly geotagged. In an extensive literature review (Schulz, Hadjakos & Paulheim, 2013) summarized some studies that addressed this challenge of geolocating Twitter users or tweets. Those spatial and temporal data in tweets are helpful for event pattern detection and spatio-temporal queries.

The aim of this paper is therefore to identify and automatically extract criminal events-related spatial and temporal information from tweets.

Reports reported that Arabic language is one of the fastest growing languages with a growth of 2000% in 12 months in twitter’s history. The major task addressed in this paper is the possibility to develop algorithms to detect and extract criminal events and test the applicability of those algorithms to Arabic content published on Twitter.

Arabic is a rich Semitic language which is highly productive, both derivationally and inflectionally (Larkey, Ballesteros & Connell, 2007). It is the fifth language most spoken. The number of Arabic legal words has been estimated to be 60 billion, derived from a closed set of approximately 10,000 roots. In the field of data mining, Arabic language raises many challenges (Darwish & Magdy, 2014). Most of these challenges are due to morphology and orthography. It is true that many other languages share some of these challenges with Arabic language, but the latter shows significant complexity from theoretical to computational linguistics.

Furthermore, Users of microblogging and social networks sites often use vernacular dialects. These dialects can differ among the Arab countries in spelling, vocabulary, and morphology from the standard Arabic which makes language processing more challenging task. The contribution presented in this paper consists on the following points:

- Determining the relationship between Twitter activities and events;
- Supporting the discovery of information that is explicitly and implicitly described in tweets texts;
- The capability to detect criminal events at a given place for a particular time, by identification of spatio-temporal information in tweets;
- Using the Arabic language. The system deals with a challenging task in tweets language processing;
- This approach can estimate earliest happening time and most impacted regions in relation with different criminal events;
- Finally, the proposed approach is validated quantitatively and qualitatively to prove its effectiveness.

The remainder of this paper is organized as follows: Section 2 surveys related work. Section 3 presents the proposed approach. Section 4 includes the presentation of the main elements of the developed system. Section 5 discusses the experiments and the results. Finally, the conclusion and the perspectives for the future work are presented in section 6.

2. RELATED WORK

Research on event detection with Twitter data is growing rapidly. Authors in (Sugitani, Shirakawa & Hara, 2013) developed an approach to detecting events based on their spatio-temporal pattern. In this study, local events, regardless of type and size, were identified by using spatio-temporal clustering technics. Another study by (Schulz, Hadjakos & Paulheim, 2013) proposed the first multi-indicator model for determining location of the user’s residence as well as the location where a tweet was created. The method is based on three types of indicators, including the toponyms in tweet text, dedicated location entries, and information from the user profile.
Sharing Work Practice in the Distributed Organization
Inge Hermanrud (2012). *Journal of Cases on Information Technology* (pp. 46-60).
www.igi-global.com/article/sharing-work-practice-distributed-organization/62862?camid=4v1a

An Empirical Investigation of the Influences of the Degree of Interactivity on User-Outcomes in a Multimedia Environment
www.igi-global.com/article/empirical-investigation-influences-degree-interactivity/1222?camid=4v1a