Pervasive and Mobile Computing 26 (2016) 103–120

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

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

Recognizing composite daily activities from crowd-labelled social media data

Zack Zhu *, Ulf Blanke, Gerhard Tröster

Wearable Computing Lab, ETH Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 13 October 2015

Keywords:
Web mining
Activity recognition
Crowd sensing

ABSTRACT

Human activity recognition is a core component of context-aware, ubiquitous computing systems. Traditionally, this task is accomplished by analysing signals of wearable motion sensors. While successful for low-level activities (e.g. walking or standing), high-level activities (e.g. watching movies or attending lectures) are difficult to distinguish from motion data alone. Furthermore, instrumentation of complex body sensor network at population scale is impractical. In this work, we take an alternative approach of leveraging rich, dynamic, and crowd-generated self-report data from social media platforms as the basis for in-situ activity recognition. By treating the user as the "sensor", we make use of implicit signals emitted from natural use of mobile smartphones, in the form of textual content, semantic location, and time. Tackling both the task of recognizing a main activity (multi-class classification) and recognizing all applicable activity categories (multi-label tagging) from one instance, we are able to obtain mean accuracies of more than 75%. We conduct a thorough analysis and interpret of our model to illustrate a promising first step towards comprehensive, high-level activity recognition using instrumentation-free, crowdsourced, social media data.

1. Introduction

Human activity recognition (AR) provides the basis for developing context-aware services and applications. Novel applications are recently surfacing to provide just-in-time information. A well known example is the commercial product Google Now,[1] which learns the daily routine of the user to provide relevant information like local weather or driving directions.

Intuitively, two dominant signals for such context-aware services are location and time, which can already provide rough estimates to infer simple and non-specific activity routines like "working" or "staying at home". To detect more fine-grained activities, Inertial Measurement Units (IMUs) are popularly employed. Using such sensor packages, often worn by the user, accelerometers, gyroscopes, and sometimes magnetometers acquire the user's motion and thereby his physical movements. In this way, researchers have been investigating the recognition of various activities, ranging from low-level ones (standing, sitting, or walking) [1] to physical activity (cycling or working out) [2,3] to higher level routines (having dinner or commuting) [4,5] that consist of numerous sub-activities (preparing food, eating, clearing the table).

With the advent of the smartphone and availability of mobile Internet access, users can stay connected with their friends at any time and express themselves and their current situation via status updates or image uploads. Known as "microblogging", users write short on-the-spot updates about their life and publish these to their social circles or interested

* Corresponding author.
E-mail address: zack.zhu@ife.ee.ethz.ch (Z. Zhu).

1 Google Now: http://www.google.com/landing/now/.

http://dx.doi.org/10.1016/j.pmcj.2015.10.007
1574-1192/

followers. Messages are usually short: just 140 characters with optional image attachment in the case of Twitter. According to Twitter's official blog [6,7], Twitter users were generating 340M tweets daily in 2012 with 140M active users, compared to 200M Tweets daily in 2011, 65M in 2010 and 2M in 2009. Therefore, it is to be expected that a large fraction of users posts regularly about their routine life experiences. Investigated by [8], typical content ranges from daily life experiences to special interests and news. In this work, we explore a novel path to conduct activity recognition. Instead of collecting evidence from instrumented sensors, we "probe" users indirectly by picking up implicit signals from their natural mobile phone usage as they post to social media platforms.

As smartphones essentially enable any-time use of social media platforms, relevant properties emerge for collecting evidence about the user's activity. First, content is shared in real-time and focuses on experiences that "happen right now" [9]. Second, "daily chatters" share content multiple times a day [8]. Third, and most importantly, it has been shown that the majority of users focus on themselves, rather than on, for example, sharing plain information or opinions [10]. Moreover, social media usage is widespread geographically and has become a natural part of people's daily lives, much of it taking place on smartphones. As a consequence, an abundance of data revealing a user's activities is generated *implicitly* by the user. Through social media platforms that record such data, we can obtain rich signals for activity recognition without any additional instrumentation. This data is spontaneously-generated and naturally occurring, thereby providing in-the-wild sensing without the restrictions of laboratory environments. Our goal is not to incentivize users to post explicitly about his activities or to post in higher quantities. Instead, we argue that data collected by social media platforms can be directly fed into activity- or context-aware systems. As such, artefacts from user-social platform interaction can be understood as a reflection of the user's daily activities.

However, because users are not constrained beforehand to systematically post specific daily situations or activities for self-reporting, this opens the question as to how to define a common scheme for *activity*. Here, we make use of a standardized activity taxonomy from the American Time-Use Survey (ATUS) [11]. It is defined by the Bureau of Labor Statistics in the United States for investigating time-use of the American population. The taxonomy describes a comprehensive, multi-tier hierarchy of typical activities people perform in everyday life. It has been investigated for ubiquitous computing systems as well in the past [12,13]. We select this taxonomy for its relevancy, comprehensive coverage, and also its overlap with other activity surveys from healthcare [14,15].

Social media instances are short and abrupt in nature, leading to potentially ambiguous and/or manifold interpretations of activity classes. Take the common example of dining out with friends, should it be classified, or even labelled, as "Eating & Drinking" or "Socializing, Relaxing, & Leisure" (categories of [11])? Therefore, it is a challenging task, not only for machine learning techniques, but even for humans to agree on a single activity when examining the expressed content afterwards. We contrast this work with our earlier investigation [16] by allowing multi-labelling and learning of composite activities in hopes of alleviating such ambiguity to comprehensively capture activity implications from social media.

In this paper, we investigate the potential of harvesting and extracting publicly self-reported activities through social media. Using text mining and machine learning techniques, we build statistical models to map user signals to activity classes. The key research questions we aim to answer are, therefore:

- Is it feasible to crowdsource labelling of noisy social media posts to identify human activities?
- Can we automatically estimate the main activity and simultaneously occurring, manifold activities of a user from social media posts?

Towards these two questions, we make the following contributions:

- We present an architecture for gathering and labelling activity reports. In an attempt to comprehensively cover the variety of possible human activities, we rely on social media platforms for large-scale gathering of data and crowdsourcing engines for labelling of data.
- Although noisy and unstructured, we characterize our dataset to reveal the rich variety of activities contained within it and the potential of such data to reflect collective human behaviour.
- Finally, we construct an activity recognition model capable of recognizing the main activity category as well as all applicable categories (for manifold activities) from 10 activity classes, with accuracies of 76% and 75%, respectively. We provide a thorough evaluation and model interpretation comparing the use of single-labelled and multi-labelled training data. We find that, while both approaches perform similarly well for main activity classification, multi-labelled training data is necessary to successfully capture the manifold activities implied within social media instances.

In Section 2, we first review existing work in the area of instrumentation-free approaches for human activity recognition. Then, in Section 3, we present our system architecture and the crowdsourced labelling task. In Section 4, we discuss the collected dataset and challenges that arise from harvesting social media data for activity inference. We describe our model in Section 5 for automatic activity recognition based on a crowd-labelled dataset. We present quantitative results evaluating our approach in Section 6 and discuss the limitations of our approach as well as potential solutions to address these limitations in Section 7. Finally, we conclude and provide an outlook for our work in Section 8.

2. Related work

Since the pioneering work of Bao and Intille [17], activity recognition (AR) research has evolved significantly due to the increasing ubiquity of commercially-available mobile and wearable devices. A summative review by Lane et al. [18]