Doctoral Thesis

Wearable Activity Recognition with Crowdsourced Annotation

Author(s):
Nguyen-Dinh, Long-Van

Publication Date:
2016

Permanent Link:
https://doi.org/10.3929/ethz-a-010714038
Wearable Activity Recognition with Crowdsourced Annotation

A thesis submitted to attain the degree of DOCTOR OF SCIENCES of ETH ZURICH (Dr. sc. ETH Zurich)

presented by

LONG-VAN NGUYEN-DINH

MSc. Computer Science, Purdue University, USA

born on 30.10.1984
citizen of Vietnam

accepted on the recommendation of

Prof. Dr. Gerhard Tröster, examiner
Prof. Dr. Aude Billard, co-examiner

2016
Abstract

The recent ubiquity of wearable sensor technology enables opportunities to monitor daily activities of people. Applications include ambient assisted living, heath monitoring. Activity recognition from wearable sensors is often accomplished by applying supervised learning on sensor signals. An annotated training data set is required to perform supervised learning. So far, most of the existing works on activity recognition require a small number of experts for data annotation regardless of high labeling effort (i.e., time-consuming, tedious). Consequently, labeling by experts provides high quality annotations, but is non-scalable for a large data set.

Crowdsourcing gains popularity recently to distribute data annotation tasks to a crowd of ordinary people. In this work, we investigate the use of crowdsourcing in activity recognition systems to reduce the effort of collecting large-scale training data, but not to sacrifice a high performance from experts’ annotations. This work comprises six scientific publications that leverage the two aspects of crowdsourcing: (1) the use of crowdsourcing to annotate a long continuous recording of activities in which start and end boundaries and labels of activities are explicitly specified (2) the use of online crowd-generated sharing databases (e.g., a sound repository Freesound) where contributors sporadically record each individual activity of interest and upload to the shared repository. Those databases can be retrieved to extract the demanding training set for activity recognition.

In the first aspect, we first conduct a case study to collect annotations for the existing activity data sets by using the crowdsourcing service Amazon Mechanical Turk. The crowdsourced annotations can get as high as 80% sample-based accuracy if multiple crowdsourced labelers are applied. Otherwise, the annotations contain a large presence of noises (52% of instances are labeled incorrectly). Crowdsourced annotations suffer from labeling noises such as mislabeling, or inaccurate identification of start and end time of activity instances. We introduce a taxonomy of annotation noises which possibly occur in a crowdsourcing setting and analyze annotation noises in the crowdsourced annotated data set collected from the case study.

The results show that the noisy annotation can degrade the state-of-the-art activity recognition methods significantly. We propose a novel gesture recognition method - WarpingLCSS as a linear-time template matching method that is robust to annotation noises. The method quan-
tizes signals into strings of characters and then applies variations of the longest common subsequence algorithm (LCSS) to spot gestures. The WarpingLCSS is evaluated extensively on both real and synthetic noisy crowdsourcing scenarios on the three existing data sets with various activity classes (10-17 classes) recorded from accelerometers on arms. The WarpingLCSS achieves better performance than the DTW-based methods and SVM, especially with the large presence of noise. With 60% mislabeled instances, WarpingLCSS outperforms SVM by about 22% F1-score and outperforms DTW-based methods by 36% F1-score. Moreover, WarpingLCSS can tolerate 30%-40% jitter level (i.e., a shift in the annotation temporal boundaries). Additionally, we demonstrate the efficiency of WarpingLCSS in both clean expert-annotated data sets as well as in multimodality settings in which a large combination of different multimodal sensors at different on-body positions is deployed. Given the robustness of WarpingLCSS against annotation noises, we demonstrate that WarpingLCSS can be used as a filtering component to discard noisy-annotated samples and select well-annotated ones for other classifiers like SVM to improve their performance.

We further propose a new annotation technique in which labelers do not have to select the start and end time carefully, but mark a one-time point within the time an activity is happening. This one-time point annotation technique is a special case of annotation noise (the boundary shrinks to a point) and it reduces significantly the labeling burden. However, one-time point annotations cannot be used directly for activity modeling. We propose a preprocessing step to correct temporal boundaries for activities given their one-time point annotations. Specifically, we propose the novel BoundarySearch algorithm to search for temporal boundaries of an activity based on data patterns around their one-time point annotations. We evaluate the method on the three existing data sets with 10-17 classes and the performance on the corrected annotations is just lower than the training on well-annotated annotations by 3% F1-score.

In the second aspect of crowdsourcing, crowd-generated shared repositories capture the diversity in user contexts due to contribution from different people. However, crowd based models fail to capture specific data patterns of targeted users. As a result, it is far to reach user-dependent recognition performance. We focus on adaptation techniques that combine crowd-generated data and user-specific data to achieve high performance similar to a user-dependent recognition system, but still minimize the labeling effort. We investigate
two adapting approaches: 1) a semi-supervised learning to combine crowd-sourced data and unlabeled user data, and 2) an active-learning to query the user for labeling samples where the crowd-sourced based model fails to recognize. We extract audio data from the online crowd-generated audio repository Freesound to train a base model for user daily activities. We test our proposed approaches on 7 users using auditory modality on mobile phones with a total data of 14 days and up to 9 daily context classes. Experimental results indicate that the semi-supervised model can indeed improve the recognition accuracy up to 21% but is still significantly outperformed by a supervised model on user data. In the active learning scheme, the crowd-sourced model can reach the performance of the supervised model with only a few label queries.

Our proposed algorithms enable the opportunities to use crowdsourcing to reduce the labeling effort for activity recognition systems, but still achieve as good performance as experts’ annotation. This work is an important step towards a large-scale activity recognition system in which an effort to collect large number of activities on a large number of users can be distributed to crowdsourcing. Therefore, this work provides the fundamentals for the next generation of wearable assistance scenarios that support activity monitoring for everyone everywhere.
Zusammenfassung


Sogenanntes Crowdsourcing, wobei eine größere Menge an Freiwilligen zur Erhebung von Daten mitwirkt, hat in letzter Zeit an Popularität gewonnen bezüglich Datenannotation. In dieser Arbeit untersuchen wir die Anwendung von Crowdsourcing in der Aktivitätserkennung. Dadurch soll sich der Annotationsaufwand für große Datensätze begrenzen lassen, aber eine vergleichbare Qualität wie bei Expertenannotation erreicht werden. Die vorliegende Doktorarbeit umfasst sechs wissenschaftliche Publikationen, die zwei Aspekte dieser Zielsetzung genauer untersuchen: 1) die Anwendung von Crowdsourcing bei der Annotation von langen, kontinuierlichen Datenauflagen und 2) die Gewinnung von Trainingsdaten aus Online-Datenbänken (z.B. die Tondatenbank „Freesound“), wo Benutzer freiwillig und sporadisch annotierte Trainingsdaten zur Verfügung stellen.


Die Resultate dieser ersten Untersuchung zeigen, dass Fehlannotationen die Aktivitätserkennung für Standard-Algorithmen stark beein-


Bezüglich des zweiten Ziels dieser Arbeit, der Verwendung öffentlicher Datenbänke als Trainingsdaten, zeigen unsere Analysen, dass Crowdsourcingdaten die Diversität an Benutzersituationen gut beschreiben können. Gleichzeitig scheitern die Modelle aber dabei, spezifische, benutzertypische Verhaltensmuster zu erkennen. Als Resultat sind die Erkennungsgenauigkeiten bei Crowdsourcing-
