Prior Knowledge of Human Activities from Social Data

Zack Zhu **Ulf Blanke** **Alberto Calatroni** **Gerhard Tröster**
Wearable Computing Lab, ETH Zurich
{zack.zhu, ulf.blanke, alberto.calatroni, troester}@ife.ee.ethz.ch

ABSTRACT

We explore the feasibility of utilizing large, crowd-generated online repositories to construct prior knowledge models for high-level activity recognition. Towards this, we mine the popular location-based social network, Foursquare, for geo-tagged activity reports. Although unstructured and noisy, we are able to extract, categorize and geographically map people’s activities, thereby answering the question: what activities are possible where? Through Foursquare text only, we obtain a testing accuracy of 59.2% with 10 activity categories; using additional contextual cues such as venue semantics, we obtain an increased accuracy of 67.4%. By mapping prior odds of activities via geographical coordinates, we directly benefit activity recognition systems built on geo-aware mobile phones.

Author Keywords

Web Mining; Activity Recognition; Crowd Sensing

ACM Classification Keywords

H.3.3 Information Search and Retrieval: Information filtering

INTRODUCTION

Automatic activity recognition (AR) is a significant research topic with far-reaching applications in many areas such as healthcare, personal assistance, and targeted delivery of information. Often, sensor data measuring motion and other physical modalities (e.g. sound) provide the basis for recognition. Despite success in recognizing basic activities (e.g. walking, drinking, climbing stairs), the recognition of high-level activities (e.g. making breakfast or socializing at a dinner) remains a difficult task. Large variability in execution complicates the recognition. Social and demographic factors (e.g. age, gender, geography, or cultural background) impact this variability as well. Furthermore, different activities can be expressed similarly (e.g. playing soccer vs. jogging).

Although in-situ measurements provide important signals, another crucial component is location-specific prior knowledge. The key question that we attempt to address in this paper is: how do we comprehensively obtain large-scale context-to-activity mappings in a low-cost way to aid AR systems? For this, we utilize the ever-increasing data of location-based social platforms. In this paper, we propose

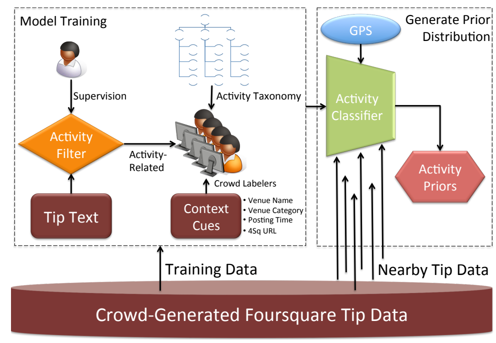

Figure 1. Architecture describing how activity prior distribution is inferred from crowd-generated Foursquare data.

a method to extract, filter, and semantically categorize geo-referenced “tips” from Foursquare. By categorizing geographically nearby tips, we can directly infer the prior odds of a user’s activity given their location. Mining short texts such as “great place for an afternoon jog” or “get a mocha and sit in front of the fire”, we advance towards a comprehensive knowledge-base of to high-level activities from crowd data.

RELATED WORKS

Previous work revealed a strong relationship between pursued activities and time or location. For example, Ye et. al. [6] explore the use of temporal features for classification and [5] shows significant improvement of recognition performance using a daily activity rhythm model. Recently, work of Partridge et al [2], extended by [1], propose the use of government-conducted time-use surveys to generate prior knowledge from thousands of subjects. Although well-annotated, the data is costly to produce and as a result, not freely available for all countries. Foursquare, being used for location “check-ins” with optional text, delivers publicly available data for many parts of the world. Furthermore, Foursquare data is geo-tagged via GPS coordinates. As a result, AR system can directly estimate potential activities at the user’s absolute location, without requiring user-specific, semantic location annotations.

FOURSQUARE TIPS

Using the Foursquare API, we scan a bounding box approximating the city of San Francisco for publicly available venue “tips”. We obtain three relevant components for each tip result: tip text, posting time, associated venue information (e.g. name, semantic category, GPS coordinates). For the purpose of this work, we do not exhaustively crawl all tips in the area, although additional queries can be easily conducted. In Table 1, we summarize the main statistics of the crawled dataset after filtering for incomplete entries.

ISWC’13, September 9–12, 2013, Zurich, Switzerland.

http://dx.doi.org/10.1145/2493988.2494343

Dataset Overview	
# of Tips	16098
# of Users	9280
# of Venues	7217
Geo-Bound (Bot. Lef.)	(37.7099,-122.5137)
Geo-Bound (Top Rig.)	(37.8101,-122.3785)

Table 1. Descriptive statistics of the raw Foursquare tips dataset for the San Francisco region.

ESTABLISHING ACTIVITY PRIORS

To distinguish between activity/non-activity -related short texts, we formulate a binary text classification problem (Figure 1, left). As training labels, we manually distinguish 3284 tips to obtain 1488 activity and 1796 non-activity labels. In accordance with standard text mining techniques, we stem and remove stopwords as part of the pre-processing. Then, unigrams and bigrams are extracted from the text to construct our feature matrix. We apply the Linear SVM of the scikit-learn package [3] with default parameters to distill activity-related tips. Through 10-fold cross-validation, we achieve a testing accuracy of 76%.

Applying this learned filter, we distill the original 16,098 tip instances to 3343 instances that are activity-related. In order to allow AR systems to directly access prior information from crowd-generated activity reports, we need to provide structure to quantify the odds of activity priors. Similar to [2, 1], we make use of the 2011 American Time-Use Study (ATUS) taxonomy [4] to categorize our raw activities data. We label "tip" texts into 17 tier-1 activity categories according to [4] via Amazon Mechanical Turk. As expected, we observe the bias of Foursquare data towards "leisurely" activities while activity categories such as "caring for household members" contain very few instances. Out of the 17 tier-1 categories, we exclude 7 categories with less than 30 instances ($\sim$1% of all instances). We build our model on the remaining 10 tier-1 categories: *Household; Socializing, Relaxing & Leisure; Eating & Drinking; Consumer Purchases; Sports, Exercise, & Recreation; Work-Related; Traveling; Personal Care; Education; and Prof. & Personal Care Services.*

Given the category labels, we learn a "tip" to ATUS category model. Again, we use the same Linear SVM classifier to classify against the 10 ATUS tier-1 categories. In addition to text, we also leverage associated meta-data (i.e. venue information and posting time) as features. Therefore, we have three feature sets, that could be combined by stacking:

1. **Tip Text:** Similar to how we filtered for activity-related texts, we construct the textual feature matrix by extracting uni-grams and bi-grams. After stemming and stop-word removal, we extract 3790 n-gram features.
2. **Venue Semantics:** Since each tip is generated for a specific venue, we extract the semantic category of the venue (e.g. synagogue, home, Mexican restaurant) for each instance. We binarize venue categories to construct an indicator matrix of 267 features (one for each category).
3. **Posting Time:** Associated with each tip is also the posting time. We chunk the time data by hour and binarize to construct 48 features depicting weekday and weekend hours.

By learning the tip-to-category mapping, we can directly calculate the odds of prior activities of a geo-location using nearby tips (Fig 1, right). Although we will conduct field validation of this component in the future, we report on our classifier's ability to make the tip-to-category mapping here.

RESULTS AND DISCUSSION

Overall, we obtain a 10-fold cross-validation accuracy of 59.2%, 61.1%, and 67.4%, using only textual features, only venue semantics, and both, respectively. We find that the posting time feature set results in only 32.6% accuracy, similar to that of the trivial majority-class classifier (32.4%). Incorporating time with other feature sets produces negligible performance change. We believe this is because the posting time does not necessarily indicate the time of an activity, as tips may be posted at any time.

Evaluating the usefulness of the three feature sets, the least indicative feature set is posting time. On the other hand, the combination of textual and venue semantics often yield the best performance, improving on performance achieved by either feature set by itself. This supports the intuition that, although the type of venue constrains what activities are possible, venues of the same category do not necessarily offer the same activities. As a result, textual features are important to make fine-grained adjustments.

CONCLUSION AND OUTLOOK

In this paper, we provide the initial investigation in utilizing Foursquare tip data to generate prior knowledge for activity recognition. We show that this repository can be mined to cover high-level activities from various aspects of life. Furthermore, such data is available for many parts of the world and can be accessed for free. Eventually, we intend to marry our model with sensor-based AR systems and quantify improvement gains via field trials.

ACKNOWLEDGMENTS

We thank the anonymous reviewers. This work has been supported by the Swiss Hasler Foundation project Smart-DAYS.

REFERENCES

1. Borazio, M., and Van Laerhoven, K. Improving activity recognition without sensor data: a comparison study of time use surveys. In *Proceedings of the 4th Augmented Human International Conference* (2013).
2. Partridge, K., and Golle, P. On using existing time-use study data for ubiquitous computing applications.
3. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* (2011).
4. Shelley, K. J. Developing the american time use survey activity classification system. *Monthly Lab. Rev. 128* (2005), 3.
5. Van Laerhoven, K., Kilian, D., and Schiele, B. Using rhythm awareness in long-term activity recognition.
6. Ye, J., Clear, A. K., Coyle, L., and Dobson, S. On using temporal features to create more accurate human-activity classifiers. In *Artificial Intelligence and Cognitive Science*. 2010.