The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Artificial Intelligence Applied to Assistive Technologies and Smart Environments:
Technical Report WS-16-01

Extracting Generalizable Spatial Features from Smart Phones Datasets

Kevin Bouchard, Lawrence Holder, Diane J. Cook
Center for Advanced Studies in Adaptive Systems
Washington State University
Pullman, WA 99163, USA
{kevin.bouchard, holder, djcook}@wsu.edu

Abstract

This paper is part of the effort to develop assistive smart homes able to monitor the daily life activity of a resident and provide punctual assistance when necessary. One of the limitations of assistive smart homes is the fact that it cannot assist the resident when he is going out. Because of this, many researchers are working on wearable sensors to keep track of the activities outside the home. Our lab proposes to instead focus on smart phones which are a cheap alternative that many persons already carry in their daily life. While most algorithms used in the smart home can be exploited, smart phones generate spatial information from the GPS that do not scale very well. The goal of this paper is to initiate a discussion on spatial features and their exploitation for data mining of smart phones datasets.

Introduction

The aging of the world population is a well-documented problem that most of the developed countries are currently facing (U. Nations 2010). The consequence of world population aging are numerous. The aging of the population results in a higher prevalence of age-related diseases such as Alzheimer's and Parkinson. Older adults also require more services from the healthcare system leading to higher cost and more demand for qualified professionals. It is generally accepted that institutionalization of elders should be used as last resort when the state of the individual cannot allow him to complete the essential Activities of Daily Living (ADLs).

Many researchers envision the use of technology to promote aging in place (Morris, et al. 2013). These technologies can be simple devices to enable monitoring of an individual at a distance or more complex apparatus such as smart homes. The Center for Advanced Studies in Adaptive Systems (CASAS) at Washington State University (Cook, et al. 2013) is working on the development of such assistive smart homes. The goal is to create enhanced housing incorporating various sensing technologies in a non-invasive manner. Therefore, small binary sensors like electromagnetic contacts and infrared motion sensors are prioritized over video cameras and microphones, which can be perceived by the residents as intrusive (Demerism, et al. 2008).

Progress on smart home research has led to the deployment of realistic living places where a resident's ADLs are constantly monitored by this technology (Akl, et al. 2015; Chernbumroong,et al. 2013). One limitation of the current projects is, however, the lack of monitoring when the resident leaves the home. Indeed, the sensors are fixed in the smart home and lose track of the person whenever she gets out of range. To palliate to this problem, the development of wearable sensors blossomed in the last decade (Lara and Labrador 2013). Nevertheless, using smart phones could turn out to be a better alternative. Smart phones are already part of the life of many persons, they are light, have a good battery life and are relatively inexpensive. They are also generally equipped with good sensors including, but not exclusively, accelerometers, gyroscope and proximity.

Our activity learning algorithms are fairly sensor agnostic and can handle data from any type of sensor platform. One notable distinction between smart home and smart phone features, however, is the ability for the phone to collect position data from a GPS chip. A user's GPS-based information is valuable in discerning and predicting ADLs. In a smart home, sensors can be described based on their position in the house (e.g., "motion sensor over couch", "door sensor on kitchen cabinet"). As a result, the learned activity models generalize over multiple users and home settings. Unlike smart home data, the GPS features do not generalize easily over multiple users because individuals live, work, and interact in locations that are unique to each person.

The goal of this paper is to investigate whether spatial features can be learned in an unsupervised fashion from mobile phone data and whether these features can be used to create a generalizable model of activities from smart phone sensor data, but also to initiate a discussion about the various alternative of generalizable spatial features that could be extracted from GPS data. The paper is divided as follows. The research context is described in the next section including a concise description of a new app developed at WSU. Then, the types of spatial features that we could extract are discussed. Finally, a preliminary evaluation of the ability to learn generalizable activity models with and without these GPS data using leave-one-user-out testing on actual phone data collected from study participants is presented.

Research Context

The CASAS lab is focused on designing machine learning algorithms to learn behavioral patterns from sensor data. While we have made progress in smart home settings (Cook 2010), we are now interested in exploring the application of our activity learning methods to mobile platforms. To validate our methods, we collected activity data in naturalistic settings for smart phone users. This section presents the protocol the team has been executing to collect these datasets. With the smart phone data, we aimed to explore spatial features to create models that are generalizable to new user (Renz and Nebel 2007). Our hypothesis is that incorporating generalizable spatial features into an activity model will increase activity recognition accuracy for both cross validation and leave-one-user-out testing.

Collecting the datasets

To create the smart phone datasets, we first designed a mobile version of our activity learning algorithm, called AL (Activity Learner), which runs on IOS and Android platforms. AL is designed to collect 5 seconds of data at fixed intervals (every 10, 15, 30, or 60 minutes) as selected by the user. Whenever the sensor data is collected, AL asks the user what activity he is currently performing. The sensor data and user-supplied label are sent to a compute server to store the data and learn an activity model. Next, the team received IRB approval to collect and use data from real users for research purposes. Students were asked to install the application on a smart phone and collect data for activities of their own choosing. Some students extensively used the application, and other only entered a few ADLs. This led us to obtain 45 datasets of various lengths. Although we provide the user with an initial set of activities, they are free to add their own as well so the number of activities that are tracked by each user varies greatly. The Figure 1 shows a few screen shots of AL.

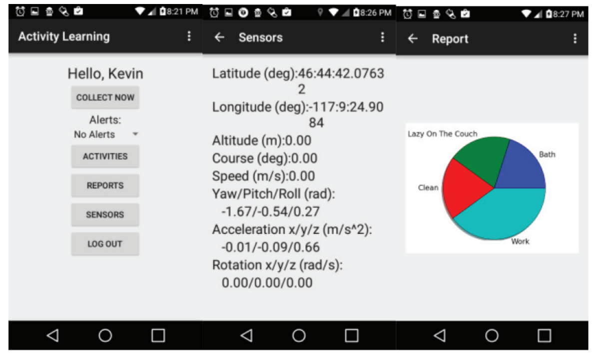

Figure 1. AL activity learner app. From left to right: top-level page, sensors page and an activity report as a pie chart.

Composition of the data

The datasets we collected are composed of 14 basic sensor types. Those sensor types include 3-axis acceleration and rotation, yaw, pitch, roll, latitude, longitude, altitude, course, current speed and date/time. For each sensor type we extract features including maximum, minimum, sum, mean, median, standard deviation, mad, cross-axis correlation, skewness, kurtosis, signal energy, power, and autocorrelation resulting in a total of 245 features. There are few challenging issues with our datasets that have to be taken into consideration. First, there is a distinct possibility that some activity labels may not be accurate since the labels are provided in real time by the users themselves. Second, a single activity may be labeled differently from one user to another. For example, some users labeled *Watch TV* as simply *Watch*, while others used similar labels such as *Watch Television* or *Relax*.

Spatial Information

The previous section provided background for the mobile-based generalized activity learning problem. In this section, we discuss the potential spatial features that can be extracted from mobile data (Koperski,et al. 1996). The sensors embedded in smart phones enable us to obtain rich spatial information. In particular, this data includes the phone movement speed, movement direction and movement acceleration in a three dimensional space. In addition, the GPS chip provides us with latitude, longitude and altitude or, simply, a position in a 2D or 3D space. Such spatial information can enable us to easily discriminate between the ADLs performed by a single user. For example, the activity *Work* will generally take place in a different location than any other ADLs. However, the activity *Work* will not be in the same location for different users. While date/time information can help in the recognition, it might not be sufficient. In this section, we discuss the various spatial features and their properties that we could extract from our datasets in order to achieve a better generalization.

Feature properties

When discussing the basic features that can be extracted from data, we note that there are many ways to represent the raw data and informative descriptors of aggregated data. Sensor data is inherently numeric but corresponding features can also be represented as discrete or qualitative values, each with advantages for activity models (Cohn 1997). Raw GPS data is clearly quantitative, representing the phone's position in a 3D Cartesian space. On the other hand, qualitative features may provide valuable generalized information. Qualitative features can be extracted from quantitative values by partitioning the values into bins (Renz and Nebel 2007).

One possible advantage of a qualitative representation of model features is that the discrete values simplify the model representation. Another advantage is that discrete-valued features can be used to represent spatial features that are actually learned in an unsupervised fashion from the data itself. For example, let's say that two objects A and B are in space at A=<123, 27, 40> and B=<140, 30, 38>. While we can numerically compute the distance between these points, we could also consider learning qualitative relationships between the points such as *same location*, *next to*, or *far apart*. There are also some disadvantages of relying on qualitative representation. The main one is the need to parameterize any qualitative spatial models. In our example, we can clearly see that depending on the method of extracting features, the same quantitative information about the positions of A and B could lead to different qualitative relationships.

Qualitative features

One well-known method of generating qualitative features is through equal-size (or, alternatively, equal-frequency) binning. To do this we obtain the upper and lower bound of the quantitative values represented by the qualitative set S and divide the resulting range into $|S|$ intervals. The binning can be performed either on a per-user basis or globally for all data. Both methods have major implications. If this is on per-user basis then the bin ranges might be different. For example, if we want to represent qualitatively the feature $Walking\ speed$ as the set $S = \{Slow, Normal, Fast\}$ with user #1 ranging from [1, 12] and user #2 ranging from [1, 6], user #1 would be walking *Slow* at 4 while user #2 would be walking at a *Normal* speed. This approach is a good option for features that are intrinsically different from one user to another.

Qualitative models have two other important properties. The first is the feature *granularity*. As we discussed earlier, a bounded quantitative feature can be directly converted into a qualitative set since there is a finite number of possible values. However, it would not be very useful since the granularity would be maxed. The selected granularity is important since it determines the expressivity of the model. However, as argued by Renz and Nebel (2007), selecting the right granularity is challenging. Finally, qualitative models often result in numerous possible entity relationships (Cohn 1997) that are Jointly Exhaustive and Pairwise Disjoint (JEPD). The advantage of this property is that any pair of entities are connected by only one of the basic relations. It is particularly useful for reasoning since it induces definite knowledge with respect to the granularity.

Types of spatial features

The main properties of qualitative spatial features have been discussed briefly in the previous section. This section will describe types of spatial information that can be obtained from mobile data. In particular, we describe the features *distance, position, shape* and *gesture*.

Position

Position may seem like a very simple feature, since it is a concept that we use in our everyday lives without analyzing it. However, there is more to it than may appear at first glance (Clementini, et al. 1997). First, position depends on the point of view that is adopted: deictic, intrinsic or extrinsic (Retzschmidt 1988). Let us consider the sentence *The ball is in front of the car*. A deictic point of view would yield position from the descriptor's origin. Intrinsic describes position from the natural orientation of the relative element (if the element has indeed a natural orientation). Extrinsic combines the position with the movement and describes it relatively to the heading direction. Supposing the letter is the origin of the descriptor, Figure 2 illustrates those three points of view.

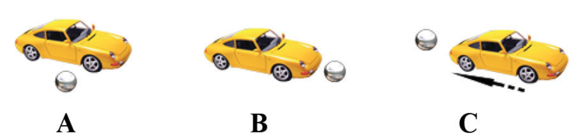

Figure 2. Deictic (A), intrinsic (B) and extrinsic (C).

The position takes an important role in all other spatial features (distance, topology, acceleration, gesture, etc.) (Renz and Nebel 2007; Cohn 1997 Clementini, et al. 1997; Egenhofer 2005). In addition to being expressed as a Cartesian coordinate, it can be expressed qualitatively. It is logical to think about position in a qualitative way (e.g. workplace, home, restaurant). Defining this finite set of qualitative positions is a challenging task. This can be accomplished by using a hierarchical clustering algorithm that would determine the number of elements in the set and the cluster would be the area covered by the qualitative position. There would be two issues. The first one is the inherent variability in user movement patterns (for example in ADLs such as walking, cycling, etc.). A second challenge is how to generate qualitative information from these movements that generalizes across users. One approach is to require all new users to train the model with labeled data for a while, but the long term goal is to be able to use an unsupervised algorithm to accomplish the same task and remove the burden from the user.

Distance

Distance is an important notion that would probably help discriminate between ADLs. Distance is a relative information that can be acquired by comparing two positions. In our dataset, it can take many forms. First, it can be a *traveled distance* during an activity. The *traveled distance* is discriminative in comparing different ADLs as it illustrates the mobility of the person. The *traveled distance* can be represented by several features. Figure 3 illustrates the difference between different *traveled distances*. The corresponding feature can be the max/min distance between any two positions in the data representing one realization of an ADL. It can also be the *absolute traveled distance*, or simply, the sum of the *traveled distance* between all positions in their ordered sequence. This version could help distinguish between ADLs with a similar *max/min traveled distance* but with one involving sedentary behavior versus high mobility. Finally, *traveled distance* can be used to compute other features such as the average *traveled distance* between positions and average variation in *traveled distance.*

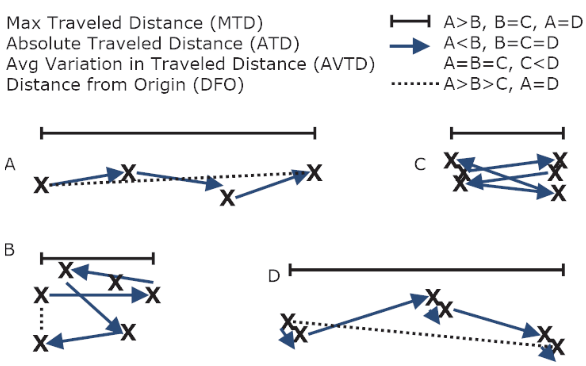

Figure 3. Four datasets having different value for few distance features.

It is worth noting that reasoning with *traveled distance* is related to time. Indeed, comparing different values without the same time frame of reference would not hold any meaningful information. Similarly, a Distance feature can also take on multiple meanings, such as *distance from origin* or *distance between ADLs*. This distance can be computed by using the centroid of the positions across all ADLs of a user. The centroid would probably be a significant location where most of the ADLs occur such as the home. However, the problem is that it widely depends on the proportion of ADLs of each type recorded by the user (e.g. if he is at work for most of the dataset, the centroid will be closer to his workplace). Another option would be to introduce the concept of home in the application. It would simply require the user to enter his home location and then all the measure of *distance to origin* would have a comparable basis across different user. Another challenge would however be that for similar locations (workplace, gym, etc.), the distance from home could be very different. That feature might therefore be more helpful when expressed as a binary value (at home/centroid or not).

Shape

The next class of spatial information we wanted to discuss is the shape. While the precise shaping of a dataset of positions is not especially relevant for the purpose of our research, the general shape might provide a simple qualitative way to distinguish between few ADLs. In particular, some ADLs might generate positions in the shape of a cloud/ disc (e.g.: cooking) because the person is idle or moving around in a certain area. However, for some other it might result in an elongated shape. Obviously, a high number of shapes could represent the positions related to each ADL. However, it is not clear how well they would scale to different user.

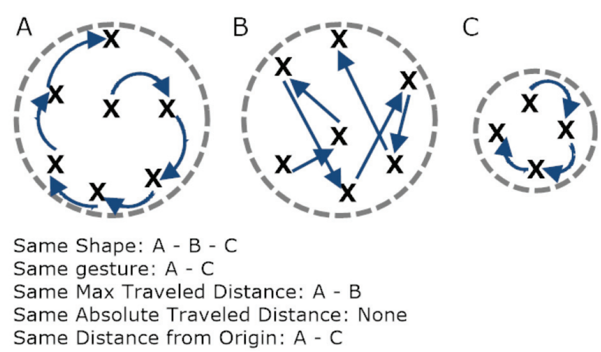

Figure 4. Example of sets of points with similar shape but different gesture. Other properties are shown.

Gesture

A gesture is widely described and recognized as an expressive and meaningful motion that conveys a message or more generally, embeds important information of spatio-temporal nature (Mitra and Acharya 2007). Gestures are ambiguous and incompletely specified, since a multitude of conceptual information can be mapped to one gesture. A single gesture is usually defined as a sequence of movement in space. The distance traveled is usually irrelevant in the gesture, as long as it is not null. In our context, it may seem that gestures and shape carry the same meaning, but they are actually quite different. Gestures are also not related to time. Therefore, while the *traveled distance* for a same ADL could differ across user, the gesture realized could be the same (i.e.: users doing it at a different speed). The Figure 4 illustrates an example where the position datasets result in the same shape but carry a different gesture.

Validations

As we said in the introduction, the goal of this paper is to initiate a discussion about the use of spatial features for data mining smart phone data. The discussion is based on the hypothesis that the GPS data help improve the accuracy of recognition algorithms. While this paper is exploratory, it seemed important to validate our premise, because if we had indication it was false, the discussion would be irrelevant. To do so, we did two sets of experiments. In the first one, the goal was to verify if the spatial data from the GPS improved the accuracy of the recognition. We selected 6 user's dataset to build decision trees with J48 in Weka (Hall, et al. 2009). We tested those decision trees with 3-folds cross validation and not 10-fold because of the size of some of the datasets we had. We also only selected the activities that had enough training data (i.e.: *Personal_Hygiene, Work, Drive, Eat, Errands, Exercise, Cook, Sleep, Watch_TV*). The results we obtained are shown on Figure 5. As we can see, for most users the raw GPS data improve the accuracy. In one case by more than 10%. Those results seems to confirm the significance of spatial information in activity recognition from smart phones. We also tested other algorithms in Weka, but J48 was the one providing the best results.

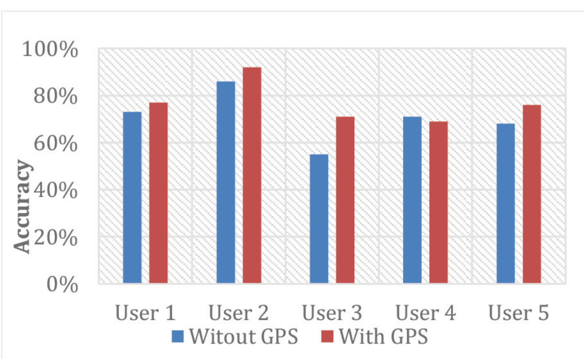

Figure 5. Accuracy of the tests with and without GPS data.

The second set of experiments we did has the goal to confirm our hypothesis that spatial data does not generalize well. To do so, did few very simple leave-one-out tests. We carefully selected classifiers that exploited the GPS data in the decision process. For that reason, we were limited to only few examples. We used classifiers learned from user 1, 3, 5. However, we could not combine the datasets of many users for the learning phase since the resulting classifier would not be using the GPS data (because it does not generalize well). We reproduced the same tests without the GPS data to confirm that the classifier generalize better. The Figure 6 shows the experiments we did and the results of the recognition.

As we can see, in most cases the recognition accuracy with the classifier exploiting GPS data was lower than its counterpart version. These preliminary tests seem to confirm our initial hypothesis. As we mentioned previously, we could not use more than one dataset for the learning phase because the resulting classifier would not exploit the GPS data. We tried different combination, but in none of the cases multiple user datasets used for learning resulted in a tree with GPS based rules. The main limitation of our preliminary experiments is that the current amount of data we have is not sufficient to perform good leave-one-out tests. Moreover, for many users, all the recorded ADLs were done in one place. Therefore, GPS was not discriminatory for these users and not interesting for this paper.

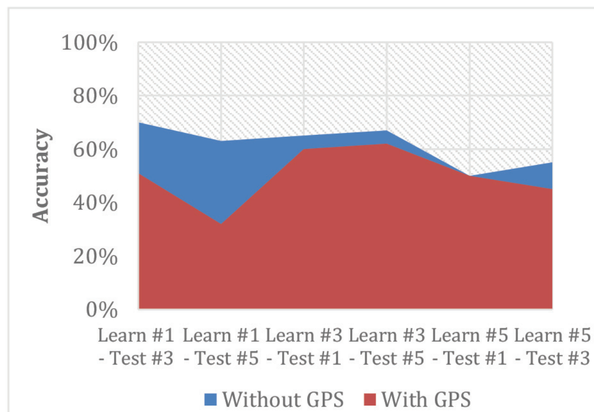

Figure 6. Accuracy of the tests with and without GPS data.

Conclusion

The goal of this paper was to initiate a discussion about the exploitation of spatial features in the process of data mining for activity recognition. It seems interesting to look into other fields of research to improve the methods we use to exploit the spatial information. In particular, it seems that research geographical information systems has make several advance in spatial data mining.

In this paper, we presented an application for smart phones to collect datasets. These data were used to validate our premise that spatial information was helpful in the recognition, but does not generalize well. While our experiments were small due to the limited data we had, the results tend to confirm our hypothesis. As a consequence, in future work, our team will test different spatial features that we discussed in this paper by collecting more data from the smart phone users. The goal will be to verify which one help the most into the recognition and which one generalize well.

References

Akl, A., B. Taati, et al. 2015. Autonomous unobtrusive detection of mild cognitive impairment in older adults. *Biomedical Engineering, IEEE Transactions on* 62(5): 1383-1394.

Chernbumroong, S., S. Cang, et al. 2013. Elderly activities recognition and classification for applications in assisted living. *Expert Systems with Applications* 40(5): 1662-1674.

Clementini, E., P. D. Felice, et al. (1997). "Qualitative representation of positional information." Artificial Intelligence 95(2): 317-356.

Cohn, A. G. (1997). Qualitative Spatial Representation and Reasoning Techniques. Proceedings of the 21st Annual German Conference on Artificial Intelligence: Advances in Artificial Intelligence, Springer-Verlag: 1-30.

Cook, D. J. 2010. Learning setting-generalized activity models for smart spaces. *IEEE Intelligent Systems* 2010(99): 1.

Cook, D. J., A. S. Crandall, et al. 2013. CASAS: A smart home in a box. *IEEE Computer* 46(7).

Demerism, G., B. K. Hensel, et al. 2008. Senior residents' perceived need of and preferences for "smart home" sensor technologies. Cambridge University Press.

Egenhofer, M. J. (2005). *Spherical topological relations.* Berlin, Germany, Springer.

Hall, M., E. Frank, et al. 2009) The WEKA data mining software: an update. *SIGKDD Explorations Newsletter.* 11(1): 10-18.

Koperski, K., J. Adhikary, et al. (1996). Spatial data mining: progress and challenges survey paper. *Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery,* Montreal, Canada.

Lara, O. D. and M. A. Labrador 2013. A survey on human activity recognition using wearable sensors. *Communications Surveys & Tutorials, IEEE* 15(3): 1192-1209.

Mitra, S. and T. Acharya 2007. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on* 37(3): 311-324.

Morris, M. E., B. Adair, et al. 2013. Smart-home technologies to assist older people to live well at home. *Journal of aging science* 1(1): 1-9.

Nations, U. 2010. World population ageing 2009, United Nations, Dept. of Economic and Social Affairs, Population Division.

Renz, J. and B. Nebel 2007. Qualitative spatial reasoning using constraint calculi. *Handbook of spatial logics,* Springer: 161-215.

Retzschmidt, G. 1988. Various views on spatial prepositions. *AI Magazine* 9(2): 95-105.