Activity Recognition with Coarse Labels Obtained from Virtual Data Sources

Le T. Nguyen, Ming Zeng, Patrick Tague, Joy Zhang
Carnegie Mellon University, Moffett Field, CA
le.nguyen@sv.cmu.edu

To recognize user physical activities using mobile/wearable sensors, one needs to first obtain activity labels to train an activity recognition (AR) model. Instead of explicitly asking users to label their activities, we study how to leverage the implicit labels inferred from virtual sources such as calendar or social media posts [3]. Real-time social network sites such as Foursquare or Instagram encourage users to share their activities (e.g., having lunch) as they happen. The key idea is to use this information as labels for training an AR system assuming that we have an access to both user virtual data streams and the sensing device (as illustrated in Figure 1). While information from virtual data streams are relatively sparse (i.e., a user only shares her activities from time to time), the sensing device can collect user's physical activity data continuously over the whole day. We propose to use the sparse virtual data as an incomplete set of labels for training the AR system. The system can then be used to recognize user's activities even when no virtual data is available.

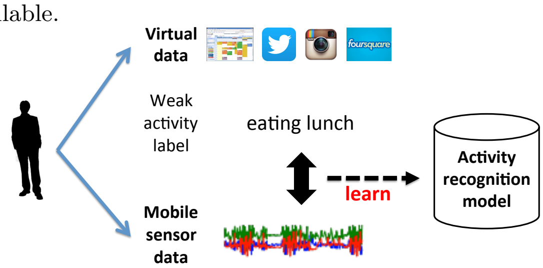

Figure 1: Virtual data sources provide information about user's current activities. This information is then used in combination with collected sensor readings from mobile phones to train an AR model.

Labels obtained from virtual data sources are inherently unreliable. Figure 2 illustrates an example of user's calendar entry with the actual activities performed by the user. For a "lunch" we assume that a user will be eating and drinking, however, she will be likely performing many other activities not associated with the lunch, such as talking on the phone. In this work, we frame this problem as an issue of *coarse labels*, where we assume that the labels extracted from the virtual source do not completely describe the activities performed by the user. In the above example, the user is having a lunch only 15 out of the 60 minutes entered in the calendar. Thus, 75% of time the activity label is incorrect (this is further referred to as a noise level of 0.75).

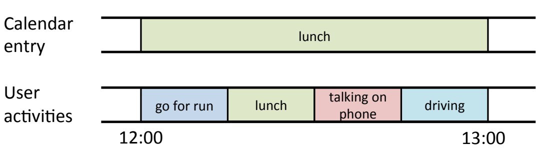

Figure 2: Coarse labeling: A calendar entry (corresponding to activity label) indicates a user having a lunch, while in reality, the user is performing many activities not related to the activity label.

To address the coarse label problem we use two Multi-Instance Learning (MIL) approaches: mi-SVM and MI-SVM [1]. In traditional supervised learning each instance is assigned one label, whereas MIL handles the cases when a label is assigned to a group of instances, but the labels of each individual instances are unknown. MIL is then used to infer the labels of the instances. Both *mi-SVM* and *MI-SVM* are built on top of a traditional SVM [1]. They are iterative approaches consisting of two steps in each iteration: estimation step and update step.

mi-SVM starts with assuming that all instances in positive bags are positive and all instances in the negative bags are negative. First, a traditional SVM classifier is learned from the initial label configuration. The trained classifier is used to estimate the labels of the instances in positive bags. In the update step, the estimation is assumed to be correct and used to retrain the SVM model. The whole training procedure is repeated until the SVM predictions are stabilized, i.e., until there are no changes of label predictions.

In **MI-SVM**, the estimation step remains similar to mi-SVM. However, in the update step, instead of considering all instances in the positive bags, each positive bag is represented by one "most positive instance". These representatives are then used for retraining the SVM.

We conduct a preliminary experiment on a public dataset [2] and show the F1 score for varying noise level in Figure 3. The traditional supervised approach (SVM) performs well when there is no noise. However, with the increasing noise level, MIL outperforms supervised learning, since it was designed to handle the noise.

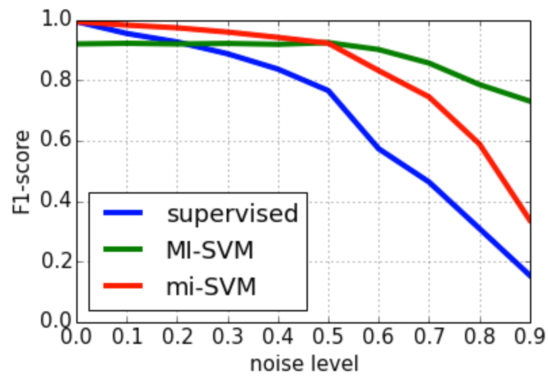

Figure 3: The traditional supervised SVM performs poorly in the cases of noisy labels, especially when the noise level is high. On the other hand, MIL approaches are designed to handle such cases.

[1] S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems*, 2002.

[2] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 2011.

[3] Z. Zhu, U. Blanke, A. Calatroni, and G. Tröster. Human activity recognition using social media data. In *Conference on Mobile and Ubiquitous Multimedia*, 2013.