Exploiting Social Networks for Large-Scale Human Behavior Modeling

The Cooperative Communities (CoCo) learning framework leverages everyday social connections between people to personalize classification models. By exploiting social networks, CoCo spreads the burden of providing training data over an entire community.

Nicholas D. Lane
Microsoft Research Asia

Ye Xu, Hong Lu, and Andrew T. Campbell
Dartmouth College

Tanzeem Choudhury
Cornell University

Shane B. Eisenman
Harris Corporation

The mobile-sensing revolution is coming of age, and we will soon see these systems in everyday use. This progress is accelerated by the development of smartphones as a viable sensing platform. Today, most mobile phones include various sensors, such as GPS, accelerometers, microphones, and cameras.[1] Classification models can exploit such data to allow, for instance, a mobile phone to understand our actions and environment.[2–4] Research into building these models is driving key mobile application domains including mobile heath[2] and green energy awareness.[5] However, significant challenges exist in the real-world use of human activity modeling. For example, a key obstacle is the differences in contextual conditions and user characteristics (for example, age, gender, and lifestyle) encountered in large-scale mobile sensing systems. This leads to the discriminative features in sensor data, used by classifiers to recognize different human activities, varying from user to user. The problem only worsens as the scale of these systems increases, further widening the diversity of contexts and users to which classifiers are exposed.

Training personalized classification models, which continues to attract attention from mobile-sensing researchers,[6–8] can counter such problems. These models are tuned to the particular sensor data encountered by each user. However, training a personalized model for each user requires a large amount of labeled data from each person. The burden of manually collecting and annotating sensor data falls upon the user, making personalization ill-suited for general use in large-scale mobile sensing systems.

The Cooperative Communities (CoCo) framework is a new approach to personalizing classification models by leveraging social networks. The CoCo framework significantly lowers the amount of training data required from each user by sharing training data and classification models within social networks. Unless this is done carefully, however, the resultant classification model will not be personalized to a specific user. In the CoCo framework, only people with strong similarities share training data. The wide range of social networks (for example, collocation, temporal co-occurrence, and friendship) indicates different types of interpersonal similarities. The CoCo framework uses these different networks to determine a social network graph that quantifies the similarity between users, identifying opportunities for effective sharing. We use this social network graph as an overlay network to perform an efficient search of the entire user population without having to resort to a brute force search of all users and their data, which would not scale. We have evaluated CoCo under three common application scenarios and found that by leveraging communities it outperforms conventional approaches to modeling behavior and diminishes the manual effort required by users.

Challenges to Mobile Classification

Mobile sensors, such as smartphones, are exposed to varying environmental conditions (for example, loud outdoor streets and quiet indoor offices), and users carry them differently (stored in bags and briefcases, placed in pockets and on belts, and so on). Accompanying the diverse range of contexts to which a phone is exposed is an equally diverse assortment of users. Users can vary for many reasons, a clear example being physical differences, such as height, weight, or gender. Other categories of dissimilarity include lifestyle and background. People might live and work in different places, and although they might even do the same set of basic activities (for example, work, socialize, and exercise), they can perform these activities in distinctly different ways. The statistical models typically used for mobile sensing are supervised (example-based) and fail to generalize to the diversity common in real-world deployments. We refer to the conditions and problems in classifying human activities and contexts, using mobile sensors, as *mobile classification*. Although a wide variety of statistical models and application scenarios have been investigated,[1] the vast majority share a common approach and build a single supervised classifier that must generalize to all potential environments and users.

Increasing attention is being placed on alternative learning approaches that overcome these barriers to mobile classification. Community-guided learning (CGL)[9] proposes new methods that let otherwise conventional classification models be trained on noisy error-prone labeled sensor data crowd-sourced from users. As we mentioned earlier, another promising direction being actively explored is personalized classification models. The basic concept is to train multiple classification models, one for each user. Each model is tuned—that is, specialized to the discriminative patterns observed in an individual user's sensor data. Tuning typically requires manual input from users, either by them correcting classification errors or providing training examples by labeling sensor data with the ground-truth activity. Such user input lets the model be trained using data specific to a particular user. This approach's effectiveness has been demonstrated in a number of classification domains, such as mobile health[6] and everyday activities.[7,8]

However, these models' success depends heavily on the time and effort invested by all users. How broadly applicable this technique will become in its current form is questionable. The burden on users severely limits it from being used in mass-market consumer systems and application domains where users are not highly motivated to perform time-consuming classifier tuning. Existing approaches to the personalization of classification models are inefficient and focus on an individual providing all the necessary training data. As a consequence, users are frequently forced to collect what is in retrospect redundant training data because a similar person in a similar context might have already collected nearly identical data. To address the problem of user labeling in personalization, CoCo shares training data and classification models within the social networks of users.

Coco's primary objective is to train robust personalized models with a small amount of user time and effort by carefully sharing data between users with whom they share strong social ties. By using the training data of only highly similar users (that is, those who share characteristics and contexts), the personalized models produced by CoCo can closely approximate models personalized by individual users providing all the training data themselves.

Researchers have long recognized the tremendous power of social networks, and recent breakthroughs in mobile sensing technology have made major advances in how these natural phenomena can be studied and leveraged. Social networks exist in many forms, such as those based on friendship, online interaction, proximity, and conversation patterns. Individually each network type indicates a certain type of connection or similarity between people. Collectively, these networks capture a large amount of information about users and their communities. Not only are researchers studying the social networks themselves, but much interest exists in learning how to exploit the information these networks contain for various applications. For example, one approach to routing in delay-tolerant networks is based on user social network centrality measures.[10] However, to the best of our knowledge, applying social networks to boost human classification models remains unstudied.

Model and Data Sharing with Cooperative Social Networks

To remove the need for users to label enough data to train their model entirely by themselves, CoCo uses social connections to recognize opportunities to adopt training data already collected by other users or entire classification models that have already been personalized. As Figure 1 illustrates, CoCo consists of two key phases:

- *Constructing the social network similarity graph.* Different forms of social network information are distilled

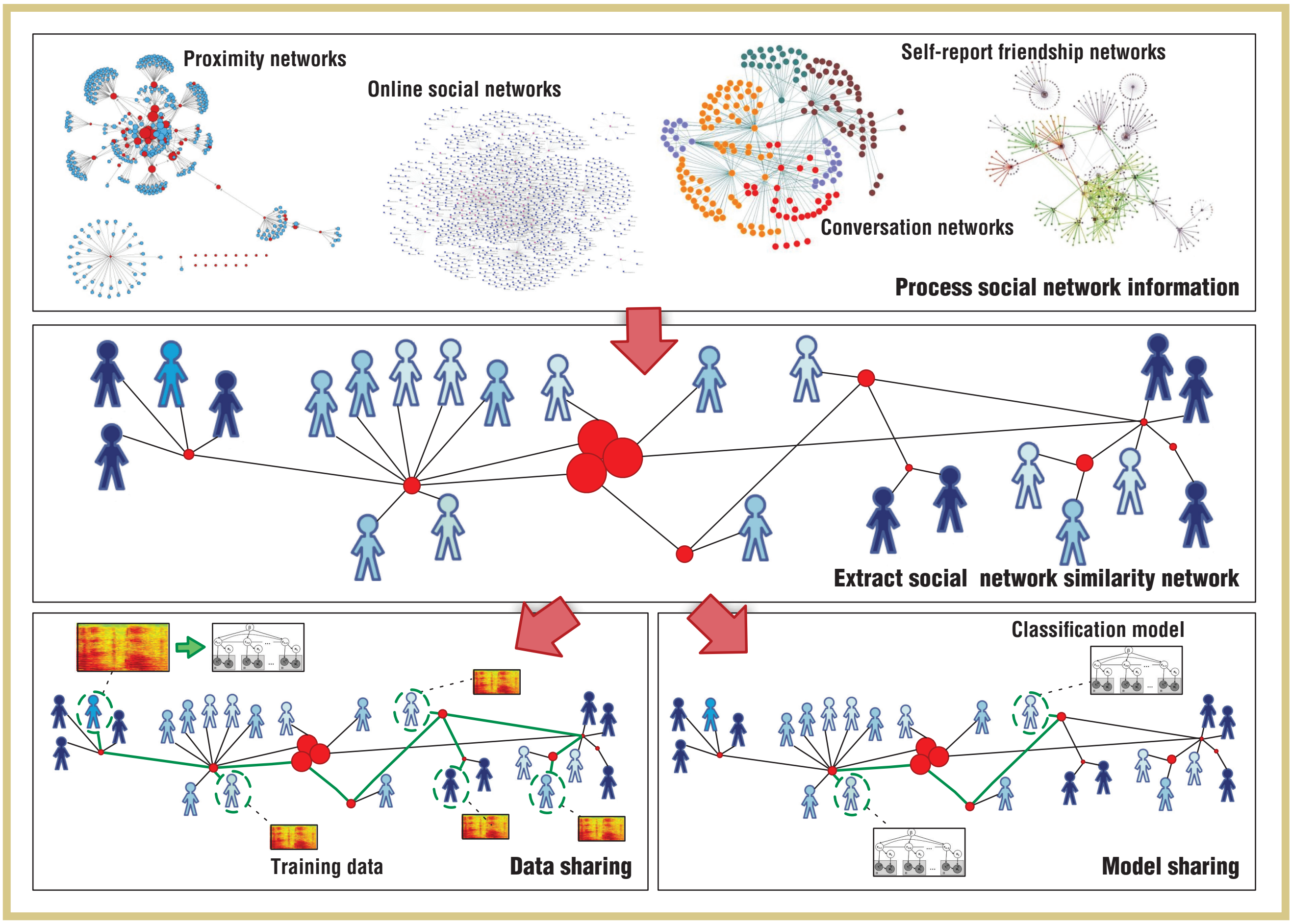


Figure 1. The Cooperative Communities (CoCo) framework. A variety of available social network information is processed to form a weighted graph between users. This graph guides an efficient search, using either data or model sharing, to find a personalized classification model for each user.

into a weighted similarity graph between users. Edge weights indicate the suitability of either training data or a complete classification model to be shared between individuals.

- *Searching for training data or classification models*. For each user model, the entire user population is searched to identify suitable training data or whole classifiers. The similarity graph defined by the social network guides this process, not only identifying which data segments or classification models are suitable, but also determining the order in which individuals are searched.

We propose data and model sharing that uses social network information to identify similarities between users. Such similarities let us train models that are specialized to recognize classes of behaviors using the discriminative features particular to the user's characteristics and environment. The use of social networks is critical not only in training personalized classifiers but also in making the proposed framework efficient and practical. As the user population increases, there will be increasing opportunities to find users who share both contextual conditions and personal characteristics. At the same time, the computational cost of searching for pairwise opportunities to share data or models between users will increase exponentially. Rather than exhaustively search all users within the system, CoCo allows searching within groups of highly similar users, enabling the search process to shift quickly to the next most similar group once the search of the current group is complete.

Diverse Social Network Information

Over the past few years, researchers have studied many forms of social networks (conversation networks, proximity networks, co-occurrence networks, friendship networks, and so on). These social networks capture different interpersonal similarities. CoCo uses strong social ties to determine if a user model will still be personalized to the user's own sensor data after either sharing training data with another user or adopting that user's model. One practical example of this is closely connected members of a conversation network, who often work or live in similar locations and so have overlapping contextual conditions. Sensor data sampled by such individuals will be similarly influenced by this shared context, increasing the suitability of their data or models to be shared. Another example is the presence of social relationships. People who regularly socialize or even groups of people who are densely interconnected within online social networks such as Facebook or Twitter likely share behavioral patterns that can manifest in their sensor data. For example, these groups might participate in similar activities (such as hobbies or entertainment) or even have similar personal characteristics (such as age, income level, and weight[11]). These types of user-centric similarities can influence the sensor data just as much as context (for example, users' age and weight can influence how they run or exercise).

We evaluated the CoCo framework using three types of readily observable social networks—real-world friendship, temporal co-occurrence, and collocation—which we combined to measure the similarity among users. Friendship can be directly determined by self-report surveys completed by the user population, although methods to automatically determine these relations using sensors are maturing.[12,13] Temporal co-occurrence and collocation networks are easier to observe directly from sensor data (for example, GPS, Bluetooth, and device timestamps) than friendship connections. The observation of repeated collocation at certain times or days helps CoCo determine various relationships and thus various types of similarity.[12,13] For example, collocation on a Saturday might indicate a more personal connection between people than collocation that occurs regularly during the week. Other factors, such as a specific location, are also helpful in detecting certain types of behavioral similarity. For instance, collocation at places such as a gym can indicate a shared preference for exercise.

Constructing the Social Network Similarity Graph

To identify suitable candidates for sharing either data or models, we construct a similarity graph based on social network information. This process merges different forms of this information before arriving at a final graph. Although we discuss only three types of social networks here, CoCo can easily be generalized to others.

More formally, in this graph every user is a node. The weight of the edge between two nodes i and j is the similarity score $\mathrm{sim}_{\mathrm{user}}(i, j)$. We compute this value independently, once for each type of social network available, before applying a weighted average of these values to arrive at a final similarity value. Depending on the types of activities being classified, we apply different weights to each variety of social network. For example, connections based on online social networks might indicate shared behavioral patterns but not necessarily shared contextual conditions, as many virtual friends rarely meet and spend time together. Similarly when classifying types of activities that rely on a microphone, for which environmental conditions (such as a quiet office or a noisy bar) can be critical, the weight of online social network information should be reduced when computing the final similarity score.

Precisely how we compute the similarity for each variety of social network will vary based on information type. For friendship social networks, we adopt a simple binary score; if two users are friends, the score is 1, otherwise it is 0. For temporal co-occurrence and collocation social networks, we first divide the range of location or time values into m intervals, where each interval is regarded as a bin. For each user, we then construct a histogram from these m bins, which reflects the distribution of the user's location or temporal patterns. Using the histogram, we compute the temporal co-occurrence or collocation similarity score between two users i and j as follows:

$$\mathrm{sim}_{\mathrm{user}}(i, j) = \sum_{k=1}^{m} e_i^{(k)} e_j^{(k)} \qquad (1)$$

in which $e_i^{(k)}$ is the frequency of user i's location or time values that are assigned to histogram bin k.

Searching the Social Network Similarity Graph

CoCo's search policy remains the same irrespective of the framework variant (that is, data or model sharing). This policy tries to find either specific data segments collected by another user or a complete model. Our similarity graph acts as an overlay network that sits above the entire user population. Instead of searching directly within this user population, the similarity graph guides the process. By using the graph's edge weights, the search process can safely ignore many potential candidates, moving quickly by simply finding the next most similar user.

Before the search starts, users are clustered together into single logical search units called *cliques*. In large-scale deployments, it is impractical and unnecessary to search and train models directly for individuals. Thus, CoCo trains personalized classifiers for each clique. CoCo clusters the user population into cliques using the k-means clustering algorithm based on the social network similarity scores. However, the use of cliques is entirely optional, with the framework operating identically regardless of whether cliques are used. In what follows, the same description can apply without the use of cliques, if we assume a maximum clique size of one.

CoCo initially assumes that a classification model is trained for every clique i using any training data the clique members provide up to this point. Next, for each clique a search proceeds whereby users in any clique with high similarity to clique i shares either data or models. We define the between-clique similarity score as follows:

$$\mathrm{sim}_{\mathrm{clique}}(C_i, C_j) = \frac{\sum_{a \in C_i} \sum_{b \in C_j} \mathrm{sim}_{\mathrm{user}}(a, b)}{|C_i| \cdot |C_j|} \quad (2)$$

Therein, $\mathrm{sim}_{\mathrm{user}}(a, b)$ is the similarity score between two users a and b (computed using all three social networks). The value of $\mathrm{sim}_{\mathrm{clique}}(C_i, C_j)$ indicates the similarity score between two cliques C_i and C_j. This search proceeds until the stopping criterion is reached, specifically, if the number of searched cliques is larger than a threshold σ.

Different actions are taken after the search is complete, depending on which variant of the CoCo framework is used (that is, data or model sharing). If data sharing is employed, all the users in the cliques visited during the search prior to reaching the stopping criteria are exploited. Each of these users provides the entire training dataset they have collected. These data, along with the data collected by the users belonging to clique i, are used to train a classification model. All members of a clique share this same model.

Alternatively, if model sharing is used, all cliques visited during the search are evaluated one by one. If any of the clique classification models exceeds the performance of clique i's current model, we replace this model, and the procedure of sequentially testing clique models continues. We check all cliques in case a later clique provides a better model. We evaluate models relative to clique i's current model based on their performance using any labeled data collected by clique i members. Any one of several existing approaches for judging classification model quality can be used within the CoCo framework.

Evaluation

We evaluate the CoCo framework using three datasets that represent common types of mobile classification. Each dataset not only offers different types of social network information, it also requires classifying different human activities. Using these datasets allow us to examine our technique's effectiveness and versatility.

We benchmark our framework's performance against two baselines that are representative of conventional approaches to mobile classification in use today. The first baseline, **single model**, trains the same generic model for all users based on the available labeled data. Under the second baseline, the **isolated model**, we train a different classification model for each user, using the labeled training data collected by the users themselves. No cooperation or sharing of training data occurs among users during any phase of these two benchmarks.

Datasets

The *everyday activities* dataset contains both simple activities, {walk, run, stationary}, and high-level behaviors, {meeting, studying, exercising, socializing}. We perform an experiment over 19 days with 20 people, each carrying a Nexus One smartphone sampling from the accelerometer, microphone, and GPS. We ask participants to label their own data on the phones directly or later during an offline Web-based data-collection phase, which we verify manually through random sampling. Two types of social ties between participants are available: binary self-reported friendship associations from surveys completed by participants, and collocation and temporal co-occurrence information based on GPS location and time stamps.

Significant places is a commonly studied problem in pervasive computing.[14] A temporal sequence of sensor data, often based on GPS location estimates, is used to compute locations that are personally significant to a user (for example, gym, home, and work). The *significant places* dataset includes data collected over a 12-day period. We ask 13 participants to carry Nokia N80 or N95 mobile phones that capture data from the Wi-Fi, Bluetooth, and GPS sensors. We ask participants to label the times when they find themselves in locations they consider to be personally significant, entering this information directly on the phones as the ground truth. Post-collection validation via exit interviews occurs to clean data entry errors. For this dataset, we only collect participant social network ties based on the results of self-report surveys.

We externally source a dataset[4] containing transportation modes, specifically, {bike, bus, car, walk}. The *transportation* dataset consists of 51 people, each carrying a personal GPS device for three months. Participants provide ground truth data of their transportation mode, which is independently verified while the dataset is collected. This dataset only includes collocation and temporal co-occurrence network information.

Results

For each experiment, we train classification models using the model and data sharing variations of the CoCo framework, using the **single model** and **isolated model** benchmarks for comparison. We use a maximum clique size of one, searching and training on each individual. We repeat the training of each of these four models assuming different amounts of labeled data provided by users to show the sensitivity of the model accuracy to the availability of training data. During experiments we use a randomly assigned 20 percent of labeled data as test data (used to evaluate accuracy) and the remainder as a pool of training data. Although our results indicate CoCo is effective across all datasets, each dataset is based on between 13 and 51 people and lasts between 12 days and 3 months. Given that CoCo is intended to overcome the challenges of long-term large-scale deployments, we intend to further evaluate the framework in future experiments with much larger numbers of participants over multiple months.

Comparing data and social network similarity. We begin with a simple experiment

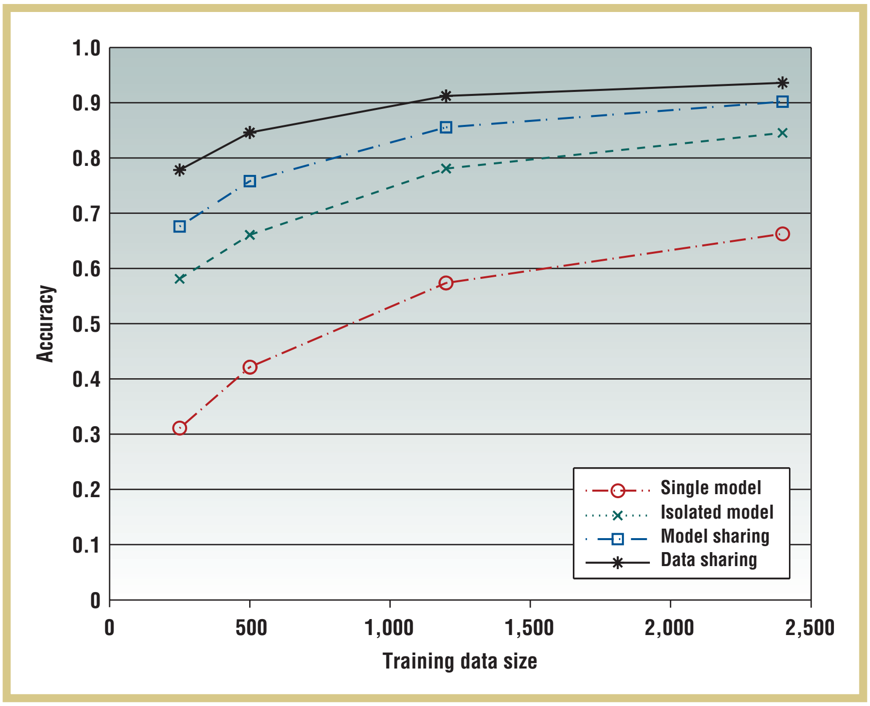


Figure 2. Classification accuracy under different amounts of training data on the everyday activities dataset. For each dataset size tested, the CoCo framework techniques of model and data sharing outperform conventional approaches to classifier training.

to test the intuition that underpins the CoCo framework. Our approach implies that people connected by social ties have similarities in their sensor data because of shared behavioral patterns and environment. If this is true, when groups of people are tightly connected by close social ties, training data can be merged safely between them, and classification models trained by one person can be used by others. An alternative means to identify such similarities is to search the raw sensor data directly. We could compute pairwise similarity between the data contributed by users. Although such an approach would be effective, it would not scale to large-scale mobile sensing systems with large volumes of data and people.

To evaluate our hypothesis, we use the *everyday activities* dataset. We compare how closely the edge weights of our social-network-based similarity network correspond to a network based on pure raw sensor data similarity. To compute the similarity based on raw data, we adopt a commonly used approach from computer vision.[15] We compare the two similarity graphs by calculating the Frobenius norm of the difference of their respective matrix representations.

The result of this comparison reflects the average difference of each corresponding pair of similarity measurements present in the two matrices. We find that the norm is small, 0.014, with the average difference in edges being only about 1.4 percent. Our results show that the similarity computed by leveraging social networks is closely aligned to the one we compute using a purely data-driven method, which has a far greater computational cost.

We coarsely quantify this cost difference by comparing the computational time for both approaches.

Our *everyday activities* dataset contains more than 400 Gbytes of raw sensor data. Computing the data-based similarity requires around 20 hours using a single desktop Linux machine. In contrast, on the same machine and dataset, social-network-based similarity can be obtained in under 10 seconds.

Everyday activities. Figure 2 presents the accuracy of model sharing, data sharing, and the two baselines using the *everyday activities* dataset. This experiment uses previously proposed features and classification models.[3] It varies the quantity of training data between 250 and 2,500 labeled feature vectors collected per user. This quantity of training data corresponds to between 20 and 200 minutes of human labeling time. The figure indicates that model and data sharing can achieve better performance than the single model and isolated model, regardless of how much training data is available. Importantly, the CoCo framework's benefit relative to the closest baseline is largest when users provide smaller quantities of training data. Figure 3a illustrates the social network graph we use in this experiment, which is based on merging friendship, temporal co-occurrence, and collocation information.

Significant places. Our experiment uses features and a classification model validated in previous work as being appropriate for recognizing significant places.[16] For this experiment, the training set size ranges from 100 to 1,500 feature vectors. These extremes correspond to approximately 1 and 12 days of semiregular labeling by users. In this experiment, a primary difficulty in collecting labeled data is the slow rate at which people visit locations significant to them. As Figure 4 shows, when the training dataset is small, both data sharing and model sharing versions of our framework outperform the single model and isolated model. As the training set grows, model sharing continues to outperform all other approaches, although eventually the isolated model achieves higher accuracy than data sharing. Furthermore, a counterintuitive result occurs in which classification model accuracy drops although training data increases.

Upon carefully reviewing the data, we find this is due to disagreement among participants when labeling locations. Consequently, when certain users add labeled data, it actually weakens the existing models of other users due to inconsistencies in class labeling. We do not see this effect in the *transportation* or *everyday activities* datasets because their classes are more clearly defined and agreed upon between all users compared to the loose categories present in *significant places* dataset.

Figure 4 suggests that model sharing is also relatively unaffected by this complication due to label disagreement. We find the reason for this is the presence of some carefully labeled classifiers provided by a few users with whom the majority of people agree. Accuracy under model sharing is not affected because the models themselves are shared, rather than training data.

The issue encountered in this dataset highlights the more general problem of data quality when sharing within user communities. Data quality can vary based on, for instance, the amount of user consensus for definitions of each activity. Malicious users could even intentionally "pollute" shared data. CoCo does not currently address this issue, although it might be partially mitigated by sharing only within social networks in which members know and trust each other. Existing work investigating reputation systems might be applicable to this problem, as might CGL techniques,[9] which can "clean" inconsistent and noisy crowd-sourced labels prior to training otherwise conventional classification models.

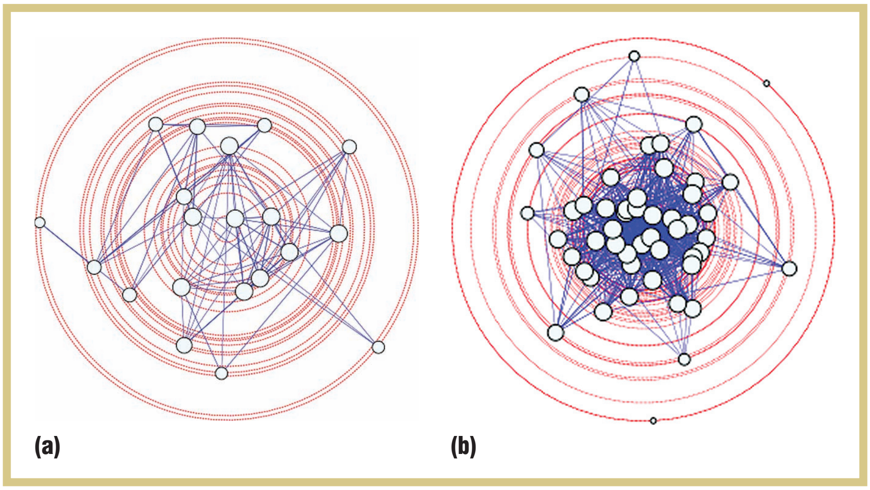


Figure 3. The social network graph for the everyday activities and transportation datasets. (a) The everyday activities dataset, with 20 people, combines friendship, temporal co-occurrence and collocation information. (b) The transportation dataset, with 51 people, uses only temporal co-occurrence and collocation information.

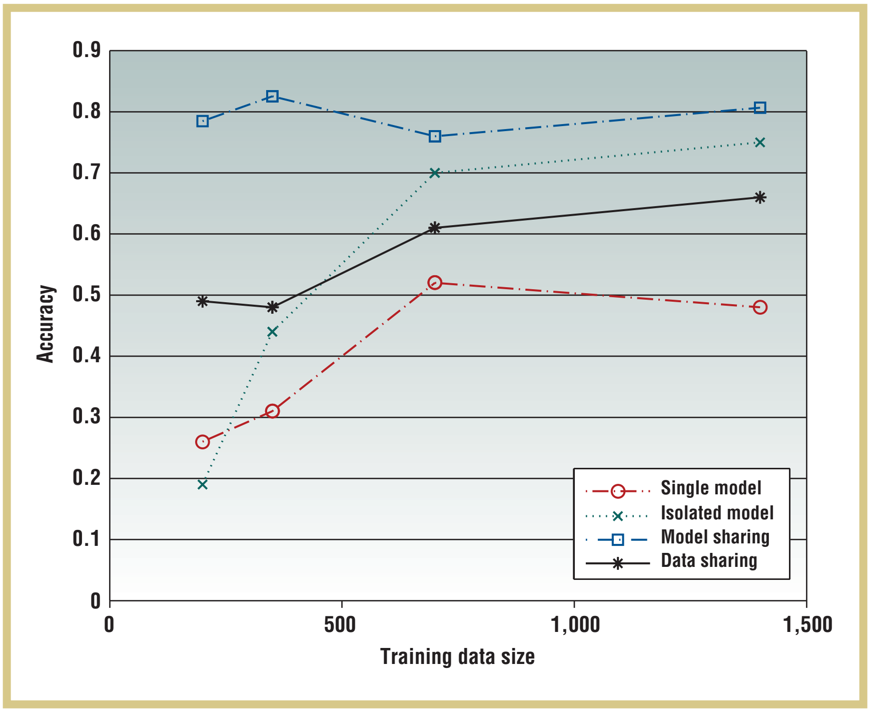


Figure 4. Classification accuracy under different amounts of training data on the significant places dataset. This graph indicates classifier accuracy actually falls, at times, although the amount of training data is increased. This is the result of disagreement as to label semantics between study participants.

Transportation. We concluded our experiments by examining the accuracy of transportation mode inference. In this experiment, both the features and classification model are identical to those used in already published work.[4] We test classifier accuracy with training datasets as small as 400 and as large as 3,200 feature vectors per user. Training data of this size can be labeled by an individual over approximately 13 to 100 hours, respectively. Figure 3b shows the social network graph that forms based on solely temporal co-occurrence and collocation information from the 51 people in this experiment.

Figure 5 indicates that our model and data sharing variations can achieve

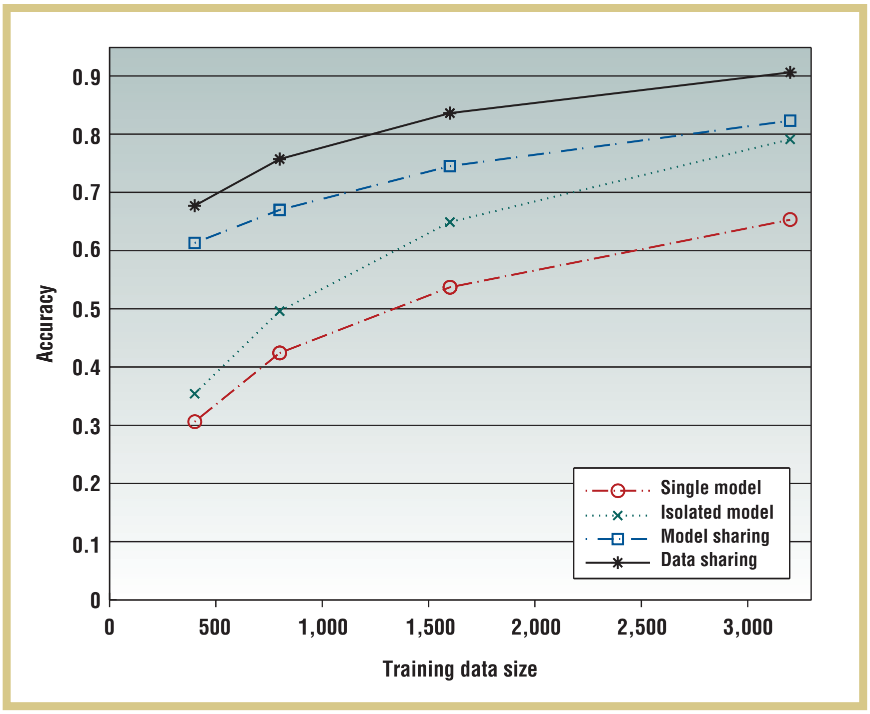


Figure 5. Classification accuracy under different amounts of training data on the transportation dataset. We find that the CoCo framework is more accurate than any of the representative classification baselines, across all sizes tested of training data.

about 30 percent better accuracy than the single model, and 20 percent better than the isolated model, for all sizes of the training set we test. Our experiment demonstrates that data sharing can perform with similar accuracy to the isolated model when trained with just 800 labeled feature vectors per user, whereas the isolated model requires more than 3,200.

As the field of mobile sensing matures, researchers are increasingly encountering the limits of conventional approaches to activity recognition. Today's de facto standard practices for modeling human behavior rely on supervised learning and experiments to collect carefully controlled training data. However, these techniques struggle to cope with diverse user populations and noisy real-world deployment conditions. The CoCo framework is contributing to a promising new alternative direction for activity recognition—one that studies how not just individual users, but the communities in which they live, can be leveraged to better model human behavior[17]. In this article, we described how social networks can be exploited, but anticipate this is only one of the many opportunities that exist to incorporate networks of users within sensing systems. Ultimately, we believe hybrid sensing systems that can intelligently exploit user communities in a variety of ways will overcome many of the obstacles to human modeling that currently prevent widespread usage of mobile sensing during everyday life.

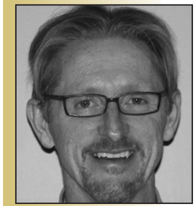
REFERENCES

1. N.D. Lane et al., "A Survey of Mobile Phone Sensing," *IEEE Comm.*, vol. 48, no. 9, 2010, pp. 140–150.
2. S. Consolvo et al., "Activity Sensing in the Wild: A Field Trial of Ubifit Garden," *Proc. Conf. Human Factors in Computing Systems* (CHI 08), ACM Press, 2008, pp. 1797–1806.
3. H. Lu et al., "The Jigsaw Continuous Sensing Engine for Mobile Phone Applications," *Proc. Embedded Networked Sensor Systems* (SenSys 10), ACM Press, 2010, pp. 71–84.
4. Y. Zheng et al., "Understanding Mobility Based on GPS Data," *Proc. Conf. Ubiquitous Computing* (Ubicomp 08), ACM Press, 2008, pp. 312–321.
5. M. Mun et al., "PEIR, the Personal Environmental Impact Report, as a Platform for Participatory Sensing Systems Research," *Proc. Conf. Mobile Systems, Applications, and Services* (MobiSys 09), ACM Press, 2009, pp. 55–68.
6. B. Longstaff, S. Reddy, and D. Estrin, "Improving Activity Classification for Health Applications on Mobile Devices Using Active and Semi-Supervised Learning," *Proc. Pervasive Health*, IEEE Press, 2010, pp. 1–7.
7. H. Lu et al., "SoundSense: Scalable Sound Sensing for People-Centric Applications on Mobile Phones," *Proc. Conf. Mobile Systems, Applications, and Services* (MobiSys 09), ACM Press, 2009, pp. 165–178.
8. M. Stikic, K. Laerhoven, and B. Schiele, "Exploring Semi-Supervised and Active Learning for Activity Recognition," *Proc. 2008 Int'l Symp. Wearable Computers*, (ISWC 08), IEEE CS Press, 2008, pp. 81–88.
9. D. Peebles et al., "Community-Guided Learning: Exploiting Mobile Sensor Users to Model Human Behavior," *Proc. Conf. Artificial Intelligence* (AAAI 10), AAAI Press, 2010, pp. 1600–1606.
10. E. Daly and M. Haahr, "Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs," *Proc. Mobile Ad Hoc Networking and Computing* (MobiHoc 07), ACM Press, 2007, pp. 32–40.
11. N.A. Christakis and J. H. Fowler, "The Spread of Obesity in a Large Social Network Over 32 Years," *New England J. Medicine*, vol. 357, no. 4, 2007, pp. 370–379.
12. N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure Using Mobile Phone Data," *Proc. Nat'l Academy of Sciences* (PNAS), vol. 106, no. 36, 2009, pp. 15274–15278.
13. J. Cranshaw et al., "Bridging the Gap between Physical Location and Online

Nicholas D. Lane is a researcher in the Mobile and Sensing Systems Group at Microsoft Research Asia. His research interests revolve around mobile sensing systems that incorporate scalable and robust sensor-based computational models of human behavior and context. Lane has a PhD in computer science from Dartmouth College; his dissertation pioneered community-guided techniques for learning human behavior models better able to cope with real-world conditions containing diverse user populations. Contact him at niclane@microsoft.com.

Ye Xu is a graduate student in the computer science department at Dartmouth College, where he works with the Mobile Sensing Group. His research interests include machine learning and ubiquitous computing, with a particular interest in Bayesian learning, multi-instance learning, dimensionality reduction and feature selection, and human behavior modeling. Xu has an MS in computer science from Nanjing University, China. Contact him at ye@cs.dartmouth.edu.

Hong Lu is a PhD candidate in the computer science department at Dartmouth College and a member of the Mobile Sensing Group and the MetroSense Project. His research interests include ubiquitous computing, mobile sensing systems, and human behavior modeling. Lu has an MS in computer science from Tianjin University, China. Contact him at hong@cs.dartmouth.edu.



Andrew T. Campbell is a professor of computer science at Dartmouth College, where he leads the Mobile Sensing Group and the MetroSense Project. His research interests include mobile phone sensing systems. Campbell has a PhD in computer science from Lancaster University, England. Contact him at campbell@cs.dartmouth.edu.

Tanzeem Choudhury is an associate professor of information science at Cornell University. She directs the People-Aware Computing Group, which develops systems that can reason about human activities, interactions, and social networks in everyday environments. Her research interests include using wearable sensors to capture and model social networks automatically on the basis of face-to-face conversations. Choudhury has a PhD from the Media Laboratory at the Massachusetts Institute of Technology. Contact her at tanzeem.choudhury@cornell.edu.

Shane B. Eisenman is currently with the Networking and Advanced Development group at Harris Corporation. Eisenman has a PhD in electrical engineering from Columbia University, where he was a member of the Control, Management and Telemedia Group and the Armstrong Project. He was also a member of the Sensor Lab and the MetroSense Project at Dartmouth College. His research interests include wireless networking and mobile computing, with a particular interest in human-centric sensing, computation, and communication. Eisenman has an MSc in electrical engineering from Columbia University. Contact him at seisenma@harris.com.

Social Networks," *Proc. Conf. Ubiquitous Computing* (Ubicomp 10), ACM Press, 2010, pp. 119–128.

14. D. Ashbrook and T. Starner, "Using GPS to Learn Significant Locations and Predict Movement across Multiple Users," *Personal Ubiquitous Computing*, vol. 7, no. 5, 2003, pp. 275–286.
15. W. Dong et al., "Efficiently Matching Sets of Features with Random Histograms," *Proc. Conf. Multimedia* (SIGMM 08), ACM Press, 2008, pp. 179–188.
16. N.D. Lane et al., "Cooperative Techniques Supporting Sensor-Based People-Centric Inferencing," *Proc. Conf. Pervasive Computing*, Springer, 2008, pp. 75–92.
17. N.D. Lane, "Community-Guided Mobile Phone Sensing Systems," *PhD thesis*, Dartmouth College, 2011.