Robust Active Learning Using Crowdsourced Annotations for Activity Recognition

Liyue Zhao[1] **Gita Sukthankar**[1] **Rahul Sukthankar**[2]
lyzhao@cs.ucf.edu **gitars@eecs.ucf.edu** **rahuls@cs.cmu.edu**
[1] Department of EECS, University of Central Florida
[2] Robotics Institute, Carnegie Mellon University

Abstract

Recognizing human activities from wearable sensor data is an important problem, particularly for health and eldercare applications. However, collecting sufficient labeled training data is challenging, especially since interpreting IMU traces is difficult for human annotators. Recently, crowdsourcing through services such as Amazon's Mechanical Turk has emerged as a promising alternative for annotating such data, with active learning (Cohn, Ghahramani, and Jordan 1996) serving as a natural method for affordably selecting an appropriate subset of instances to label. Unfortunately, since most active learning strategies are greedy methods that select the most uncertain sample, they are very sensitive to annotation errors (which corrupt a significant fraction of crowdsourced labels). This paper proposes methods for robust active learning under these conditions. Specifically, we make three contributions: 1) we obtain better initial labels by asking labelers to solve a related task; 2) we propose a new principled method for selecting instances in active learning that is more robust to annotation noise; 3) we estimate confidence scores for labels acquired from MTurk and ask workers to relabel samples that receive low scores under this metric. The proposed method is shown to significantly outperform existing techniques both under controlled noise conditions and in real active learning scenarios. The resulting method trains classifiers that are close in accuracy to those trained using ground-truth data.

Introduction

Human activity recognition in real-world settings, such as the kitchen, has been an significant research topic for household eldercare, healthcare and surveillance applications. However, efficiently labeling a large-scale dataset by hand is still a painful and time-consuming task. Recently, several services such as Amazon's Mechanical Turk (MTurk) allow researchers to divide the large dataset into thousands of simpler annotation tasks and distribute those tasks to different workers. MTurk has thus become a a fast, affordable and effective mechanism for annotating data for supervised learning models (Vondrick, Ramanan, and Patterson 2010; Rashtchian et al. 2010).

Copyright © 2011, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

However, crowdsourcing is not a panacea for data labeling for two main reasons. First, it is difficult for human labelers (who are typically not experts in the field) to interpret noisy data generated by body-worn sensors, such as IMU traces. For this reason, researchers now typically generate an additional time-aligned channel of information, such as video captured from the user's perspective (De la Torre et al. 2008). While such data would not necessarily be available in actual applications, it serves a vital role in the classifier training process. Unfortunately, our studies have shown that even with video data, human labelers generate an unacceptably high rate of labeling errors. In this paper, we propose augmenting the action labeling task with an object recognition task (on images). Our models can then exploit information about which objects in the scene are being manipulated by the user to better infer likely actions (similar in spirit to Wu et al. (2007)).

Second, even with these measures, crowdsourced annotations are observed to generate uneven, subjective and unreliable labels. This presents a fundamental challenge to active learning: how best to obtain labels (at affordable cost) given a distributed set of unreliable oracles? We propose a new active learning paradigm that balances the traditionally myopic selection of instances in active learning (e.g., based on proximity of data to the decision boundary) with a more global term that also considers the distribution of unlabeled data. Additionally, we build a model that calculates the confidence for each of the labels obtained through active learning. Through judicious use of relabeling, we can thus obtain additional labels for instances that have low rank in terms of confidence.

Our experiments show that combining our improved selection criterion with confidence-driven relabeling enables us to affordably train classifiers on noisy crowd-sourced data that are comparable in accuracy to those trained using noise-free training data.

Related Work

Activity recognition using body-worn inertial sensors has been an important research topic for many years, particularly in the context of eldercare and healthcare applications; see Avci et al. (2010) for a recent survey.

Bao and Intille (2004) and Ravi et al. (2005) recognized simple human activities from data acquired using ac-

celerometers. Wu et al. (2007) and Pentney et al. (2006) proposed approaches to estimate complex daily life activities based on object sequences collected using RFID tags attached to objects. To bootstrap the learning process, video data is combined with RFID data to jointly infer the most likely activity. The Carnegie Mellon University Multi-Modal Activity Database (CMU-MMAC) aims to provide a rich source of partially ground-truthed data for research in this area (De la Torre et al. 2008). Promising results have been reported using different machine learning techniques on CMU-MMAC (Spriggs, De la Torre, and Hebert 2009; Zhao, Wang, and Sukthankar 2010) and we also use it as a standard dataset to evaluate our proposed methods.

Tong and Koller (2002) and Tong and Chang (2001) demonstrated the use of Support Vector Machines (SVM) to construct pool-based active learners for text classification and image retrieval, respectively. However, many research issues remain. First, most active learners select the next instance to label in a greedy and myopic manner. Neglecting the global distribution of data can cause the learner to converge only to a locally optimal hypothesis. A second serious problem is that the uncertainty sampling is inherently "noise-seeking" (Balcan, Beygelzimer, and Langford 2006) and may take a long time to converge. Hierarchical clustering approaches such as Dasgupta and Hsu (2008) provide the learner with global information about the data and avoid the tendency of the learner to converge to local maxima.

Crowdsourcing annotation services, such as Amazon's Mechanical Turk, have became an effective way to distribute annotation tasks over multiple workers. Active learning approaches have recently become popular in this context since they provide a budget-conscious approach to data labeling (Vijayanarasimhan, Jain, and Grauman 2010). Sheng, Provost, and Ipeirotis (2008) observed the problem that crowdsourcing annotation may generate unreliable labels. Since the SVM active learner is greedy and myopic, these noisy annotations will lead to error accumulation when selecting the next sample. Proactive learning (Donmez and Carbonell 2008) offers one way to jointly select the optimal oracle and instance with a decision-theoretic approach. Since we cannot select oracles, our proposed approach selects instances using a combination of loss criteria and judiciously resamples potentially incorrect labels when necessary.

Proposed Approach

Figure 1 presents an overview of our approach. Our goal is to accurately label large quantities of IMU data with activity labels, from which we can train activity recognition classifiers. Given temporally-aligned video and IMU data, such as those provided in the CMU-MMAC dataset (De la Torre et al. 2008), we ask MTurk workers to label short, automatically-segmented video clips of cooking activities with the label(s) corresponding to the current action(s). The list of actions is shown in Table 1.

Unfortunately, we observe that the raw crowdsourced action labels are unacceptably inaccurate (see Table 2). Rather than simply voting among a pool of redundant workers (which is expensive), we ask each worker to solve a related task that can clean up the original labels. Specifically, we ask workers to identify which object(s) from Table 3 are visible in the scene. As detailed below, we train a Bayesian Network to infer actions based on the set of observed objects. This approach enables us to significantly improve annotation accuracy (see Table 2).

We then apply this framework in an active learning context to iteratively annotate data. As shown in Algorithm 1, the idea is to select instances according to a mixture of two criteria: 1) MaxiMin (Tong and Koller 2001) and 2) based on the proximity of other labeled data. Each step in our proposed approach is described in greater detail below.

Table 1: List of actions

1. close	6. read	11. twist off
2. crack	7. spray	12. twist on
3. open	8. stir	13. walk
4. pour	9. switch on	
5. put	10. take	

Table 2: Crowdsourced annotation accuracy. The accuracy of annotations improves when the action labeling task in video is supplemented with an object identification task. We further improve on these numbers using relabeling.

Task	Label accuracy
Action only	47.97%
Object only	53.11%
Action+Object	62.52%

Dataset Overview and Feature Extraction

Our experiments use the CMU-MMAC dataset (De la Torre et al. 2008), which consists of data collected from dozens of subjects performing several unscripted recipes in an instrumented kitchen environment. The data corresponding to a given recipe consists of approximately 10,000 samples collected at over a period of about 6 minutes. We focus on two

Table 3: List of objects

1. none	11. cook spray	21. oven
2. brownie box	12. cap	22. counter
3. brownie bag	13. knife	23. sink
4. egg box	14. fork	24. baking pan
5. egg	15. spoon	25. frying pan
6. egg shell	16. scissors	26. measuring cup
7. salt	17. cupboard	
8. pepper	18. drawer	27. big bowl
9. water	19. fridge	28. small bowl
10. oil	20. stove	29. paper towel

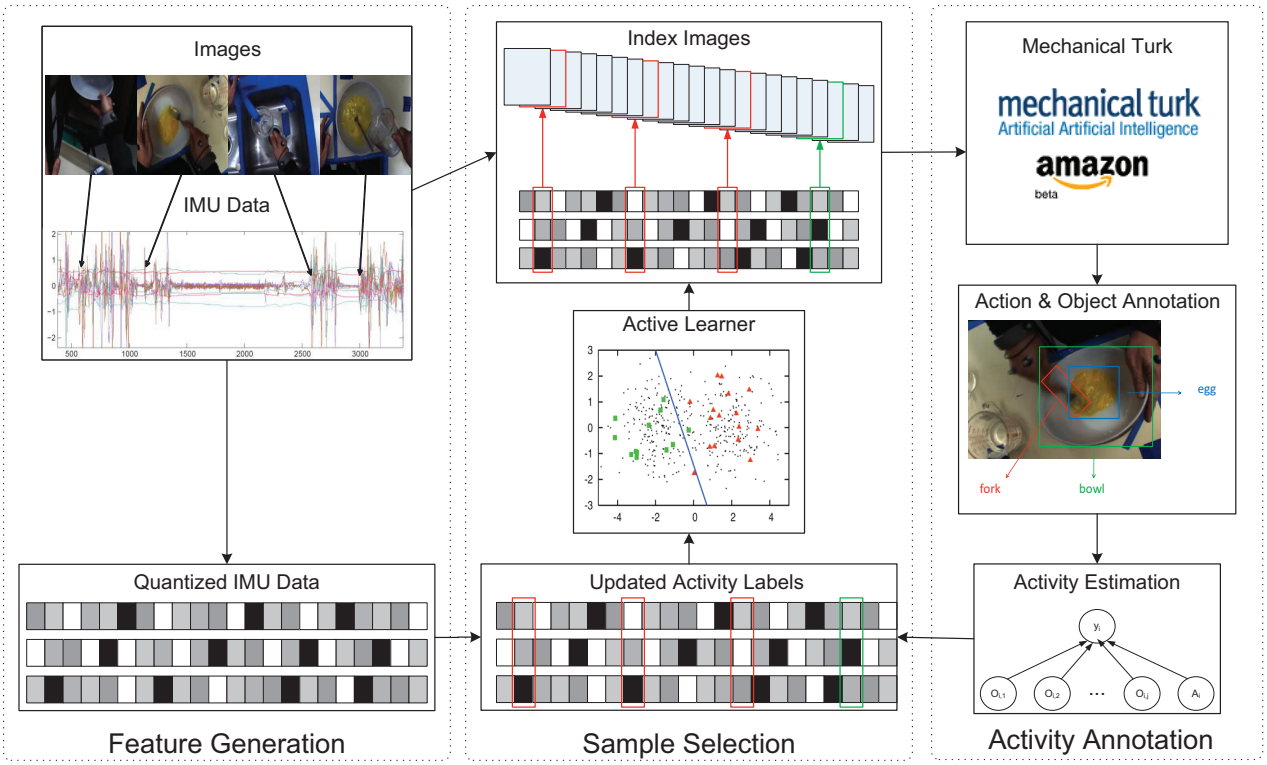


Figure 1: Overview of the proposed method. Since labeling IMU data is difficult for untrained users, we crowdsource short activity labels for temporally-aligned video clips supplemented with object labels for still images. The object labels generate more accurate activity labels. The active learning selects new instances based on a new criterion that combines local and global loss and we request relabels for data that falls below an estimated confidence threshold. The resulting system can generate accurate activity recognition classifiers from unreliable crowdsourced data.

modalities of data: first-person video and IMU.

The egocentric video was collected from a head-mounted video camera that recorded the objects and actions performed by the subject during the cooking activity at 30Hz. The IMU traces were aligned to the video and recorded using 3-axial accelerometers, 3-axial gyroscopes, and 3-axial magnetometers mounted on the subject's wrists, ankles and body, resulting in a 45-dimensional data vector, sampled at 125Hz.

The IMU data was processed using a 256-sample sliding window with 50% overlap between consecutive frames. Thus, each window corresponds to 2.05 seconds of data. Four features are extracted from each axis of the sensor: mean, standard deviation, energy and entropy, resulting in a 180-dimensional feature vector.

Inferring Actions from Visible Objects

As discussed earlier, to compensate for the poor annotation accuracy of raw crowdsourced action labels, we ask MTurk workers to annotate which objects are visible in each scene. We use these as a secondary source from which to infer action labels.

Given a video frame i with true action label y_i, we define $\mathbf{O}_i$ and $\mathbf{A}_i$ to be the object and action labels collected from MTurk, respectively. By assuming that the action and object labels are conditionally independent, we construct the following model:

$$P(y_i, \mathbf{O}_i, \mathbf{A}_i) = P(\mathbf{O}_i|y_i)P(\mathbf{A}_i|y_i)P(y_i). \quad (1)$$

Parameter learning is performed using maximum likelihood estimation and the conjugate prior is modeled as a Dirichlet distribution. Table 2 summarizes the improvements we obtain using this method.

Although this model improves annotation accuracy, our system must still guard against workers that provide random labels and flag unreliable labels. To identify such labels and tag them for relabeling, we compute a confidence score function $\mathrm{CS}(y|\mathbf{O_i}, \mathbf{A_i})$ based on the Bayesian estimate that evaluates the reliability of collected labels:

$$\mathrm{CS}(y|\mathbf{O}_i, \mathbf{A}_i) = H(y, \mathbf{O}_i, \mathbf{A}_i) + c \cdot \log(N(\mathbf{O}_i, \mathbf{A}_i)), \quad (2)$$

where $H(y|\mathbf{O}_i, \mathbf{A}_i)$ denotes the entropy over the labels estimated using Equation 1; $c \cdot log(N(\mathbf{O}_i, \mathbf{A}_i))$ is a term that penalizes workers who give few or no labels for the task on MTurk; $N(\cdot)$ denotes a counting function and c is a weight constant. Annotated instances that score poorly in terms of their confidence score are selected for relabeling.

Active Learning

Active learning seeks to iteratively obtain labels for data by identifying, at each iteration, the most uncertain sam-

ple. More formally, let $\mathcal{T} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_l\}$ be a set of labeled instances, with corresponding labels given by $\mathcal{L} = \{y_1, y_2, \ldots, y_l\}$. We also define a set of unlabeled instances as $\mathcal{U} = \{\mathbf{x}_{l+1}, \mathbf{x}_{l+2}, \ldots, \mathbf{x}_n\}$. The SVM classification problem can be represented as finding the optimal hyperplane with labeled samples that satisfies

$$\begin{array}{ll} \min_{\mathbf{w},b,\epsilon} & C\sum_{i=1}^{l}\epsilon_i + \|\mathbf{w}\|_2, \\ s.t. & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 \quad i = 1, ..., l \end{array} \tag{3}$$

where ϵ_i is a slack term, such that if $\mathbf{x}_i$ is misclassified, C is the penalty for the misclassified samples. All possible hyperplanes that could separate the training data as $f(\mathbf{x}_i) > 0$ for $y_i = 1$ and $f(\mathbf{x}_i) < 0$ for $y_i = -1$ are consistent with the version space $\mathcal{V}$. The most uncertain sample is identified by selecting the sample that halves the version space $\mathcal{V}$. (Tong and Koller 2002) describe three strategies to identify the most uncertain sample with SVM classifier. We apply the *MaxiMin Margin* strategy which selects samples that have the greatest difference for a given positive and negative label. For each unlabeled data $\mathbf{x}_i$, the loss function of the *MaxiMin Margin* approach can be represented as:

$$\text{Loss}_{\text{SVM}}(\mathbf{x}_i) = \min_{\mathbf{x}_i \in \mathbf{U}} (V^+(\mathbf{x}_i), V^-(\mathbf{x}_i)) \tag{4}$$

where $V^+(\mathbf{x}_i)$ is the size of the version space from labeling $\mathbf{x}_i$ as $+$. The loss function is equal to the smaller value of the $V^+(\mathbf{x}_i)$ and $V^-(\mathbf{x}_i)$ defined by these possible hyperplanes. Although Equation 4 applies to a binary classification problem, it can be easily redefined for multiclass settings by computing the product of the loss function for the appropriate classification hyperplanes from all classes.

However, a serious issue with the *MaxiMin Margin* criterion for instance selection is that it is a greedy strategy that can be very sensitive to noise among the existing labels, causing the optimization to fall into a local minimum. To avoid this problem, we propose including a second, more global, term into the loss:

$$\text{Loss}_{\text{cluster}}(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in \mathbf{T}} f(\mathbf{x}_i, \mathbf{x}_j). \tag{5}$$

The function $f(\mathbf{x}_i, \mathbf{x}_j) = h(\mathbf{x}_j, \mathbf{x}_{c_k}) \log(||\mathbf{x}_i - \mathbf{x}_j||)$ evaluates the uncertainty of the sample by computing the distance between the unlabeled sample of $\mathbf{x}_i$ and the labeled sample $\mathbf{x}_j$. $h(\mathbf{x}_j, \mathbf{x}_{c_k}) = 1$ if the label of $\mathbf{x}_j$ is the same as the label of centroid point $\mathbf{x}_{c_k}$, otherwise $h(\mathbf{x}_j, \mathbf{x}_{c_k}) = -1$. $||\cdot||$ denotes the Euclidean distance between two samples. The centroid point $\mathbf{x}_{c_k}$ is obtained by first performing k-means clustering on the unlabeled set and selecting the cluster center closest to $\mathbf{x}_j$.

Finally, we combine both loss functions to estimate the uncertainty of unlabeled samples. The sample that we select is obtained using:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}_i \in \mathbf{U}} \text{Loss}_{\text{SVM}}(\mathbf{x}_i) + \lambda \cdot \text{Loss}_{\text{cluster}}(\mathbf{x}_i), \tag{6}$$

where λ denotes the mixing weight between the two criteria and depends on the expected annotation accuracy. Given *a priori* knowledge about annotation accuracy, one could set λ appropriately — with reliable labels, a lower value for λ suffices. In practice, since the expected accuracy of labels is unknown, we employ cross-validation with a hold-out set to determine λ. The overall algorithm is presented in Algorithm 1.

Input: The dataset $\mathcal{T}$, the unlabeled set $\mathcal{U} = \mathcal{T}$, the labeled set $\mathcal{S} = \emptyset$, and the number of initial samples k.
Output: The fully labeled dataset $\mathcal{S} = \mathcal{T}$

Use k-means to cluster $\mathcal{U}$ into clusters $\{c_1, \ldots c_i \ldots, c_k\}$;
Query labels for $\mathcal{X} = \{x_1, \ldots, x_i, \ldots, x_k\}$ where $x_i \in \mathcal{T}$ is the sample closest to the cluster centroid c_i;
Update $\mathcal{T} = \mathcal{T} \cup \mathcal{X}$ and $\mathcal{U} = \mathcal{U} \setminus \mathcal{X}$;
Train initial classifier $\mathbf{w}_0$ using training set $\mathcal{T}$;
while *the classifier* $\mathbf{w}_j$ *has not converged* **do**
  Use classifier $\mathbf{w}_j$ to calculate the loss $\text{Loss}_{\text{SVM}}(\mathbf{x}_i)$;
  Estimate the loss function $\text{Loss}_{\text{cluster}}(\mathbf{x}_i)$ using Equation 5;
  $j \leftarrow j + 1$;
  Choose $\mathbf{x}_j = \mathbf{x}^*$ via Equation 6;
  Update the labeled set $\mathcal{T} = \mathcal{T} \cup \{\mathbf{x}_j\}$ and the unlabeled set $\mathcal{U} = \mathcal{U} \setminus \{\mathbf{x}_j\}$;
  Train the SVM classifier $\mathbf{w_j}$ with set $\mathbf{T}$;
end

Algorithm 1: Proposed active learning algorithm

Experiments

The CMU-MMAC dataset is an unlabeled dataset. To test our annotation strategies, we use the label set posted on the CMU-MMAC website[1] as ground truth labels. This consists of labels for 16 subjects who baked the brownie recipe, where the annotations were manually generated from the video data alone.

We present results from two sets of experiments. In the first set, we perform a controlled study by corrupting labels with known quantities of noise to evaluate the robustness of our active learning strategies to annotation noise. The second set examines how action annotations improve using the model trained on our auxiliary object recognition task, and how the relabeling strategy boosts the performance of active learning.

Experiment 1: Robustness to Annotation Noise

The first set of experiments examines the robustness of various active learning strategies to annotation noise. Figure 2 compares four active learning sample selection strategies under four scenarios with increasing levels of noise (0%, 10%, 30%, 50%). The selection strategies are: 1) a baseline "random" strategy (green); 2) the standard criterion, Loss_{SVM}, i.e., MaxiMin (Tong and Koller 2001), denoted as "SVM" (blue); 3) our global criterion, $\text{Loss}_{\text{cluster}}$, denoted as "Cluster" (black); and 4) the proposed criterion that combines both losses (Eqn 6), denoted as "Proposed" (red). We also

[1] http://www.cs.cmu.edu/~espriggs/cmu-mmac/annotations/

show the error rate of a gold-standard classifier trained using all of the ground-truth data (which serves as the lower bound on the error), denoted as "Ground" (yellow).

We make several observations. First, in the noise-free case, "Simple Margin" and "Mix" perform best, and "Random" worst. This is consistent with our expectations that the standard criterion is well suited for noise-free scenarios, and that our proposed criterion can match this. Next, in the 10% noise case, we see that "Simple Margin" degrades slightly but "Mix" continues to dominate since it is able to combine both myopic and global information. The trend continues in the 30% noise scenario, but now although neither of the base criteria performs better than random chance, the proposed method continues to do best. Finally, under the challenging conditions of 50% noise, none of the selection strategies are able to outperform random selection. From this, we can conclude that improving the criteria for selection in active learning can effectively counter moderate noise but is not sufficient by itself when dealing with very noisy annotations. This validates our earlier observation that crowdsourced labels may require judicious relabeling in addition to improved active learning strategies.

Experiment 2: Impact of Relabeling

In this experiment, we compare the accuracy of annotations obtained from several sources (see Table 2, discussed earlier). Specifically, we examine the following conditions: 1) action annotations generated by MTurk from video clips alone, denoted "Action only"; 2) action predictions inferred using our model from the secondary task of crowdsourced object annotations, denoted "Object only"; and 3) actions inferred by combining crowdsourced action and object annotations, denoted "Action+Object". We see that the label noise from the raw data is unacceptable (48%). A surprising finding is that we can actually do better by inferring actions from objects alone (53%). More importantly, we see that combining these two noisy sources of annotations enables us to achieve 63% annotation accuracy. Since 63% accuracy falls between the 30% and 50% noise scenarios discussed in Figure 2, we confirm that relabeling is essential.

Figure 3 details the relationship between the relabeling fraction and the resulting improvement in annotation accuracy. For example, to achieve a 95% annotation accuracy ($< 5\%$ noise), we should request about 60% of the data to be relabeled.

Figure 4 shows the error rates for the proposed active learning method, with and without relabeling, on real-world data. The green line uses the raw labels collected from Mechanical Turk. Note that even with the improved criterion for active learning, the error rate barely drops. The blue line uses action labels inferred from both action and object labels. While the error rate is somewhat better, we see that the overall performance is still poor. Finally, the red line shows the impact of relabeling 46% of samples, selected using the confidence score discussed above. Here we see a dramatic difference: the error rates on this dataset are very close to those obtained by active learning on a noise-free ground truth dataset (black line). This experiment confirms that each of our stated contributions (improved selection criteria, inferred labels from multiple crowdsourced annotations, and confidence-based relabeling) is necessary to demonstrate accurate training from noisy crowdsourced data.

Conclusion

Although crowdsourcing annotations using active learning is an attractive and affordable idea for large-scale data labeling, the approach poses significant difficulties. Our study in the domain of wearable sensor-based activity recognition shows that a straightforward approach using the raw annotations obtained from Mechanical Turk in conjunction with standard margin criteria for SVM-based active learning would fail due to the high degree of annotation noise. This paper makes three contributions that enable us to robustly train under these challenging conditions. First, we infer more accurate action annotations by combining objects with actions in a Bayesian framework. Second, we propose a new criterion for selecting instances in active learning that combines local and global measures. Third, we show that relabeling, driven by an automatically estimated confidence score is required to improve the quality of crowdsourced annotations. Our experiments using the CMU-MMAC dataset and Mechanical Turk confirm that the proposed approach improves active learning in noisy real-world conditions. The resulting classifiers are close in accuracy to those trained using ground-truth data.

Acknowledgments

This research was supported in part by DARPA award N10AP20027.

References

Avci, A.; Bosch, S.; Marin-Perianu, M.; Marin-Perianu, R.; and Havinga, P. 2010. Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *International Conference on Architecture of Computing Systems*.

Balcan, M.; Beygelzimer, A.; and Langford, J. 2006. Agnostic active learning. In *Proceedings of ICML*.

Bao, L., and Intille, S. 2004. Activity recognition from user-annotated acceleration data. *Pervasive Computing* 1–17.

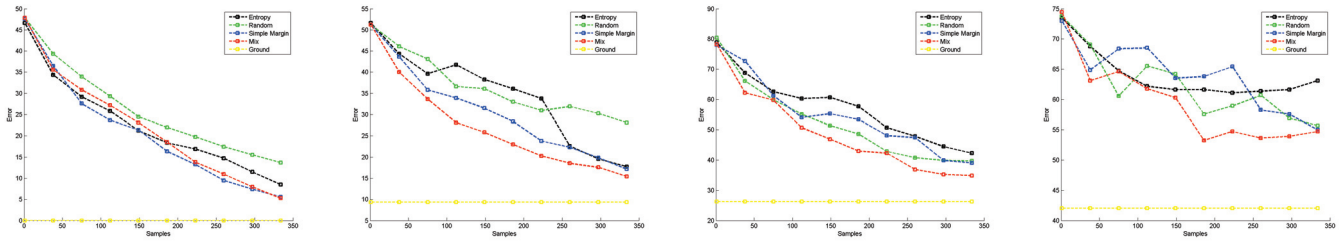
Cohn, D. A.; Ghahramani, Z.; and Jordan, M. I. 1996. Active learning with statistical models. *JAIR* 4:129–145.

Dasgupta, S., and Hsu, D. 2008. Hierarchical sampling for active learning. In *Proceedings of ICML*.

De la Torre, F.; Hodgins, J.; Bargtell, A.; Artal, X.; Macey, J.; Castellis, A.; and Beltran, J. 2008. Guide to the CMU Multimodal Activity Database. Technical Report CMU-RI-TR-08-22, Carnegie Mellon.

Donmez, P., and Carbonell, J. 2008. Proactive learning: Cost-sensitive active learning with multiple imperfect oracles. In *Proceedings of the ACM Conference on Information and Knowledge Management*.

Pentney, W.; Popescu, A.; Wang, S.; Kautz, H.; and Philipose, M. 2006. Sensor-based understanding of daily life via large-scale use of common sense. In *AAAI*.



(a) Noise free (b) 10% noise (c) 30% noise (d) 50% noise

Figure 2: Impact of oracle label noise on classifier error rates for different active learning selection criteria (lower is better). The figure key is as follows: 1) Baseline (green), 2) Loss_{SVM} (blue) 3) Cluster (black) 4) Proposed (red) and 5) Ground truth (yellow). The proposed method (red) performs best with unreliable oracles and generates classifiers that are nearly as accurate as those trained on clean data. However, no method works well once label noise gets too high, motivating our use of judicious relabeling.

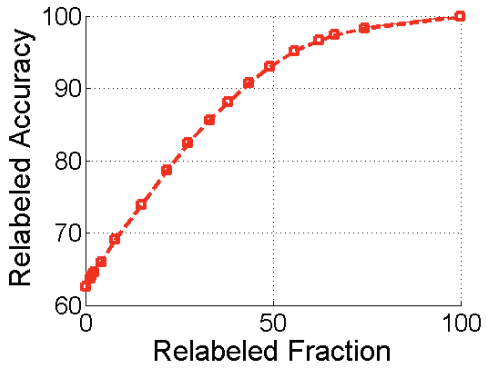


Figure 3: Annotation accuracy can be improved by relabeling instances that have low confidence estimates. For instance, relabeling 50% of the data creates a training set that has $< 10\%$ label noise.

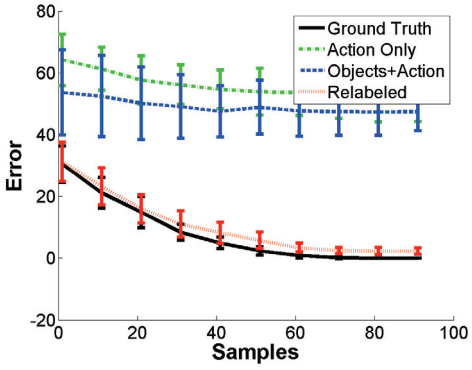


Figure 4: Proposed active learning method with different source of labels

Rashtchian, C.; Young, P.; Hodosh, M.; and Hockenmaier, J. 2010. Collecting image annotations using Amazon's Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*.

Ravi, N.; Dandekar, N.; Mysore, P.; and Littman, M. 2005. Activity recognition from accelerometer data. In *AAAI*.

Sheng, V.; Provost, F.; and Ipeirotis, P. 2008. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Spriggs, E.; De la Torre, F.; and Hebert, M. 2009. Temporal segmentation and activity classification from first-person sensing. In *Workshop on Egocentric Vision*.

Tong, S., and Chang, E. 2001. Support vector machine active learning for image retrieval. In *Proceedings of the ACM International Conference on Multimedia*.

Tong, S., and Koller, D. 2001. Active learning for parameter estimation in Bayesian networks. In *Proceedings of NIPS*.

Tong, S., and Koller, D. 2002. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research* 2:45–66.

Vijayanarasimhan, S.; Jain, P.; and Grauman, K. 2010. Far-sighted active learning on a budget for image and video recognition. In *Proceedings of IEEE CVPR*.

Vondrick, C.; Ramanan, D.; and Patterson, D. 2010. Efficiently scaling up video annotation with crowdsourced marketplaces. In *Proceedings of European Conference on Computer Vision*.

Wu, J.; Osuntogun, A.; Choudhury, T.; Philipose, M.; and Rehg, J. 2007. A scalable approach to activity recognition based on object use. In *Proceedings of the IEEE International Conference on Computer Vision*.

Zhao, L.; Wang, X.; and Sukthankar, G. 2010. Recognizing household activities from human motion data using active learning and feature selection. *Technology and Disability* 22(1–2):17–26.