Audio Analysis as a Control Knob for Social Sensing

Dinesh Verma
IBM T J Watson Research Center
Yorktown Heights, NY 10598, USA
dverma@us.ibm.com

Bong Jun Ko
IBM T J Watson Research Center
Yorktown Heights, NY 10598, USA
bko@us.ibm.com

Shiqiang Wang
IBM T J Watson Research Center
Yorktown Heights, NY 10598, USA
wangshiq@us.ibm.com

Xiping Wang
IBM T J Watson Research Center
Yorktown Heights, NY 10598, USA
xiping@us.ibm.com

Graham Bent
IBM United Kingdom Ltd
Hursley Park, Hants, UK
gbent@uk.ibm.com

ABSTRACT

While humans can act as effective sensors, human input is subject to a high degree of error and highly dependent on the context. Furthermore, extracting the signal from the noise for social sensing is a difficult challenge. One approach to improving the accuracy of social sensing is to use physical sensors as a control knob for social sensing algorithms. In this paper, we present an architecture for using audio sensors as a way to control an algorithm used for social sensing of interesting events. We present various use cases where the architecture is applicable, and go into the details of one specific use case, namely using crowd behavior in a golf-course to identify and control social media feeds related to the course.

CCS CONCEPTS

•Human-centered computing →Collaborative and social computing; •Computer systems organization →*Distributed architectures;* **•Computing methodologies →**Machine learning approaches;

KEYWORDS

Audio analytics, control, social sensing, machine learning

ACM Reference format:
Dinesh Verma, Bong Jun Ko, Shiqiang Wang, Xiping Wang, and Graham Bent. 2017. Audio Analysis as a Control Knob for Social Sensing. In *Proceedings of The 2nd International Workshop on Social Sensing, Pittsburgh, PA USA, April 2017 (SocialSens 2017),* 6 pages.
DOI: http://dx.doi.org/10.1145/3055601.3055616

1 INTRODUCTION

In social sensing applications, advance knowledge of the context of sensing can provide useful insights for selecting the appropriate sensing and analytics algorithms for the task. This can have a significant impact on the relevance and quality of the collected data and the effectiveness of the analytics. For example, suppose one is interested in finding out the times and the locations of severe traffic accidents in a large city through a social/crowd sensing task utilizing visual data from people's mobile phones and other sensory data (e.g., accelerometer reading) from vehicles around the city. Social sensing using a setup similar to [5] would be much more efficient and economical than setting up a physical sensing infrastructure similar to [8] for a city. The collective set of data that can be obtained from all the sensing devices will contain a vast amount of data completely irrelevant to sensing objective (in this case, data from devices not in the vicinity of any accident), while only a small portion will be useful for the task at hand. Sourcing from only a subset of sensing devices that are potentially relevant to the particular situation of interests would make it easier to extract true signals from the noises, and also avoid unnecessary consumption of computing and network resources. Furthermore, if the same set of the data is to be used and shared across multiple sensing tasks, each with a distinct goal (e.g., one for traffic accidents detection and the other for road surface condition monitoring), different types of analytics algorithms can be deployed according to the particular context. This will improve the quality of the output.

A common challenge in social sensing applications is the difficulty of determining the relevance of the data being collected. A typical approach widely adopted in many social sensing applications is to use the location of the sensing device as a type of context. Such prior knowledge on the locations of and the relationship between the devices can be easily obtained (most mobile phones today are equipped with GPS sensor). It can be useful in some types of social sensing applications, such as those for finding the general trends of environments and overall conditions of various kinds in a specific region [2, 7] and those utilizing the social relationship between people to perform collaborative tasks [3, 6]. However, it is not useful in detecting interesting events because, by definition, the events occur at unspecified, unpredicted times and locations, and the "interesting" events are rare (otherwise they are not interesting any more), so it is difficult, if not impossible, to pre-determine the geo-spatial region to define the relevancy of the data.

In order to determine the context of the sensor devices, one needs to collect and analyze at least "some" data anyway from the sensing field, either directly from the sensor devices to be used in the social sensing tasks or from other sensors that either perform the tasks of other application or are specifically purposed to help drawing the contextual information. The question then is what types of data and how much of them ought to be collected and used to determine the contextual information of the targeted sensing field. While the answer is application-dependent, a basic principle would be to strike

SocialSens 2017, Pittsburgh, PA USA

DOI: http://dx.doi.org/10.1145/3055601.3055616

the good balance in the cost-benefit tradeoff between the resource consumption and the accuracy of contextual information; Resource consumption is the amount of network bandwidth, storage, computational and energy consumed in analyzing the data; Accuracy and precision is defined with respect to analyzing the right amount of data to get good utility from the sensing task. Analyzing very little amount of data, or doing elementary analysis can lead to high false positives or false negatives. On the other hand, analyzing too much data may lead to computationally expensive approaches which do not add utility.

To address this challenge, we propose that social sensing architectures can gain significantly if they are augmented with control knobs, i.e. signals and sensors which influence the interaction between different sensors generating data and the analytics performed upon them. We also argue that the use of the sound, specifically ambient sound, is a good control knob for many social sensing situations.

There are several benefits of using sounds as a control knob:

- Audio sensors are ubiquitous and generally inexpensive (e.g., all mobile phones have microphones by default).
- Sound propagates in all directions and can be heard at a fair bit of distance from the source of the sound.
- Relatively simple processing of the sound signals can reveal useful contextual information.

The first two, that is, the ubiquity of the sounds and sound sensors, in particular make the audio analytics an excellent candidate for realizing an "always-on-everywhere" sensing capability. What is remaining then is how to minimize the overhead of collecting and processing the sound data. In the remainder of the paper, we present our architecture that addresses this issue in using audio analytics as the control knob for social sensing tasks in Section 2, and its use cases in Section 3, one of which is used to validate the feasibility of our architecture through experimental analysis. Section 4 concludes the paper.

2 CONTROL ARCHITECTURE FOR SOCIAL SENSING

We begin with a description of the basic architecture of the proposed social sensing control plane that utilizes the ambient sound from a sensing field and controls the sensor selection and analytics task deployment for the specific social sensing tasks. We then present a few improvements that address the challenges of scalability and managing resource consumption.

2.1 Base Architecture

Figure 1 illustrates the base architecture of the proposed control plane, where a shaded box on the right-hand side shows the internal components of the control plane. We assume the sensing field (on the left-hand side of the figure) can be monitored by various types of sensor devices that can potentially participate in social sensing tasks, many of which are capable of capturing the sound signal from the sensing field; these include mobile phones, video cameras with microphones, and stand-alone microphone devices. We also assume that the collective set of the sound sensors can sufficiently cover the sensing field both in space and in time–we are exploiting the ubiquity of the sound sensor devices we argued in Section 1.

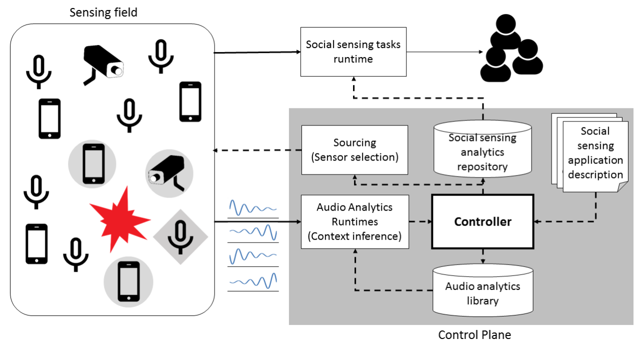


Figure 1: Base architecture of social sensing control plane based on audio analytics.

We are primarily interested in detecting interesting events that occur at unspecified times and locations within the sensing field. The detection of the events itself is performed by the social sensing tasks that utilize various types of the sensors in the field (not necessarily the sound sensors, though sound sensors can also be used in the event detection tasks). The main function of the control plane is threefold:

- Understand the requirements of the social sensing application in terms of what contexts extracted from the sounds shall trigger the active social sensing task.
- Deploy the audio analytics tasks suited for the given social sensing task, and extract the contexts out of the captured sounds in the sensing field.
- Upon detecting a specified context, deploy the social sensing tasks (sensing and analytics) for a set of sensor devices deemed to be in the detected context.

As an illustrative example, consider a social sensing application that detects incidents of gun violence in busy streets of a city. The application would utilize video and image footages collected from various camera devices (e.g., smart phones of people participating in the social sensing, surveillance cameras) to analyze the scene of the incidents and identify potential suspects in them through visual analytics. Since collecting and analyzing video footages all the time in the entire area is very expensive, in terms of sensing, collecting, and processing, the application in our architecture would instead first detect possible incidents via audio analytics that is able to detect the sounds of the gunfire (this gives the "context" of possible gunfire), and then deploy the video-based social sensing tasks for only a subset of camera devices near the location that the gun sound is detected.

In this scenario, the architecture shown in Figure 1 is applicable in the following manner:

(1) The Controller first interprets the requirements of sound-based context detection from the social sensing application description,
(2) The Controller deploys a set of audio analytics algorithm suitable for detecting the specified context,
(3) The deployed audio analytics detects the sound indicating the specified contexts from one or more microphones (marked by "diamond" shape) in the field,

(4) The Controller is notified of the occurrence of the specified context and the associated meta-information (e.g., time and location), and
(5) The Controller deploys the sensing and analytics tasks for the sensor devices (marked by "circles") near the location of the detected incident.

While our focus here is on the architectural aspects of using audio analytics as a control mechanism for social sensing applications and the specifics of the audio analytics algorithms are beyond the scope, we want to point out that various mechanisms for inferring the contextual information from the sound data exist in literature. These methods use various combinations of audio feature extractions (e.g., FFT, DCT, MFCC, etc.) and machine learning algorithms for general scene detection and specific event detection problems [1].

The description of the social sensing application description and how to translate it into the deployment action can be done in a variety of ways as well. For example, a formal, schematic language can be used to govern precisely how the controller shall interpret the requirements by the application. Another approach is to utilize NLP (natural language processing) techniques, such as Latent Semantic Analysis (LSA), on unstructured, free-form description of the application's intent and requirements, and draw "approximately appropriate" action. We plan to explore the latter in the future.

In this way, our architecture enables the targeted deployment of the social sensing and analytics tasks in which the tasks requiring expensive operations can be executed in an on-demand fashion, triggered by a ubiquitous, less expensive, background sensing and analytics of ambient sounds. In what follows, we describe several improvements that can be made in this base architecture in order to address issues related to the scalability and resource consumption of this background activity.

2.2 Scalability and Resource Consumption

Our architecture uses ambient sounds and audio analytics as a less expensive mechanism for detecting the context in which further social sensing tasks can be planned and controlled. This analytics does consume system resources for collecting, transmitting, and analyzing sounds, especially when it is performed continuously across the entire sensing field. We present here several approaches, which in combination are effective in reducing the resource consumption of such sensing.

2.2.1 "Shallow" Audio Analytics Models. In our architecture, the audio analytics algorithms that perform the context inference from the ambient sounds do not always need to produce very precise analytical results. Since they are only used as a pre-filtering mechanism, they can afford to be "approximate", essentially trading the accuracy for the resource consumption. In the context of commonly used machine learning approaches, this means that we do not need to employ "deep" models (e.g., deep neural networks such as Convolutional Neural Network, Recurrent Neural Network, etc.) that requires extensive use of memory and processing power. More traditional "shallow" models that requires significantly less computing resource (such as logistic regression, support vector machine, and various tree-based methods) can be effective enough as the control-plane analytics methods.

2.2.2 Edge Analytics. Edge analytics generally refers to a class of analytics techniques that can be executed at the edge of the network, e.g., IoT gateways, mobile devices, sensor devices. Performing the analytics at the edge has a number of advantages over conventional approaches of running the analytics at a central place like in a cloud computing environment [11], including:

- It can save a substantial amount of network bandwidth that would be spent should the data collected in the field be sent to the central servers for processing.
- It can minimize the delay in producing the analytics outcome as the data are locally processed.
- It can alleviate the privacy concern of data as the raw data are processed locally and only the results of the analytics are to be transmitted for further processing.

A challenge in edge analytics is that the edge devices generally have limited computing resources. This is not a significant concern in our case since the shallow models can be used effectively in such resource limited environment. Using this principle, the Audio Analytics Runtime module in our architecture (Figure 1) are physically placed at the edge devices, though nothing prevents from running the actual social sensing analytics tasks at the edge as well. Note that modern smart phones, which are widely used in social sensing applications, are excellent edge audio analytics platform because they have microphones and adequate computing resources.

2.2.3 Multiplexing of Sound Sensing. Capturing the ambient sound continuously throughout the entire sensing field by all audio sensing devices would consume significant energy and computing resources of the devices. Luckily, the combination of sound propagation characteristics (i.e., omni-directional) and the remote sensing capability and ubiquity of the sound sensors effectively removes the need of having all of them active all the time. Instead, a more conservative approach can be effective for our purpose, that is, letting the sound sensors "take turns" based on either a pre-determined schedule or at random. The significance of this approach is that it does not require all the devices to capture specific events indicating a specified context, but only a small subset of them in the vicinity of the event is good enough to trigger the social sensing task in the region. As an example, consider the event occurring in the sensing field in Figure 1. Though there are a few microphone-enabled devices near the location of the event, only one of them hearing the sounds of the event will be enough to trigger the deployment of the sensing task for nearby devices. Note, however, that there exists another spectrum of tradeoff between the accuracy and the resource savings: With more microphones detecting the sound-based events, it becomes gradually possible to pin-point the precise location of the sound source (through, e.g., triangulation), yet at the expense of energy/processing.

In what follows, we will present some use cases of this architecture in the context of using the audio analytics to detect interesting events in social sensing scenarios. We'll also present preliminary experimental evaluation results that demonstrate the feasibility of our approach in one of the use cases.

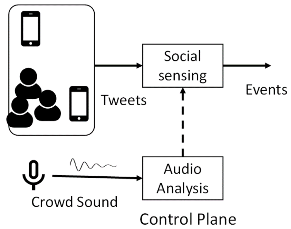


Figure 2: Sports tournament event analysis using sounds.

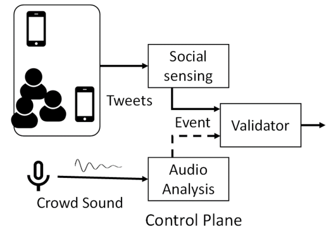


Figure 3: Validation of public event social sensing using audio analysis.

3 USE CASES

In this section, we present several use-cases where the use of audio analytics as a control knob can help improve the effectiveness of social sensing.

3.1 Sporting Events

In many golf tournaments, and indeed in many modern sports, it is common for organizers to use social media to provide highlights into interesting events. In the context of a golf match, it could be a player making an very long putt, a "hole-in-one" or other such interesting actions that happen during the game. When such an interesting event happens, social media is used by many members in the audience to send details of such events out in the field.

If we want to detect the interesting events from the social media feed, we do face the challenge that the feeds from different individuals tend to be erratic and highly subject to their individual preferences. It could be difficult to determine interesting events just from the analysis of the social media feed themselves.

On the other hand, if audio sensing is used as a control knob for social sensing, i.e. they are used to filter the relevance of a social media stream relevant to a sporting event, a significant improvement in the quality of social sensing can be obtained. In almost all sports, an interesting event is accompanied by a reaction from the crowd: applause, a roar or other reaction which shows that something interesting has happened. An analysis of the crowd reaction, i.e. the ambient sounds from the crowd, can provide a good estimation of the ground truth – whether or not something of interest has happened.

We can therefore propose a control architecture where social analysis on the tweets is coupled with a control knob that is controlled by the ambient sounds in the crowd (Figure 2). When there is a roar or clapping from the sound, the analysis of tweets about the event for some time interval is commenced. When the crowd is silent, or only normal background noise is heard, the system can ignore the tweets that are happening, anticipating that many of those tweets may not be relevant to anything of interest.

The control architecture used for this use case assumes that people would tend to tweet about the most recent interesting event within some given time threshold after the event happens. Assuming that aspect of human behavior is valid, controlling the analysis and processing of social media by using the crowd sounds as a control knob can significantly reduce the amount of irrelevant information that needs to be processed.

3.2 Public Demonstration Analysis

Social sensing can be used to track the status of public demonstrations, protests and similar expressions of public opinion. Usually, any significant event of this nature is accompanied by an active feed in social media. This would lead to the natural assumption that social sensing can provide a good mechanism to identify the effectiveness and estimate the size of a public demonstration. However, the ground truth may be quite different from the information that can be gleaned from social sensing. Social sensing is very susceptible to false positives in the case of demonstrations resulting from overactive social media participants.

An interesting example where social sensing could be led astray emerges from the pilot of OSCAR project in Wales, UK. OSCAR is a social media analysis program that was piloted by a group of universities in Wales along with the local police department to track and analyze public sentiment [9]. During a high profile meeting being held at Cardiff, UK, the OSCAR system analysis of social tweets reported a large protest group gathering around the meeting. Concerned by the analysis, the police went over to the protest site only to find a single tweeter with a computer who was inundating the social media feeds with reports of a fictitious large protest group using multiple handles.

Such type of manipulations of social media is not unusual, given the ease with which multiple handles can be established, and the ability to automate the generation of realistically-sounding tweets using modern computers. Using audio analysis as a control knob to validate the results of the social sensing can significantly reduce such manipulation of the social media.

Unlike the sports tournament where the control knob was used to reduce the workload required for social sensing, this example uses the audio control knob as a validation mechanism for social sensing (Figure 3). Using the control knob for event validation means that the ambient sounds in the environment are checked against the detection of the attributes that are reported in the social media. If the ambient sounds in the environment does not match what social sensing analysis is indicating, the social sensing system can discard those events. This cross-validation helps reduce the number of false positives that can be generated in social sensing by those who would try to manipulate the social media in this manner.

3.3 Event Localization

One of the common use cases for social sensing is that of accident localization. The analysis of information from many different cars, their collective velocity as captured by the phones of the drivers and passengers, and the motion of different people can be used as

Table 1: Performance factor definitions

	Original sound has event	Algorithm detects event
True positive (TP)	Yes	Yes
True negative (TN)	No	No
False positive (FP)	No	Yes
False negative (FN)	Yes	No

a mechanism for detecting accidents and other congestion points in real-time, and for mapping them out.

In the localization process of extracting information from such incidents, audio analysis can be used as a trigger to determine when the information pertinent to an accident may be collected from the various social sensing devices. In addition to the monitoring of the speed and location of various phones involved in the social sensing experience, one can also enable acoustic monitoring on the different phones. While the sounds do not directly contribute to information such as velocity and position of the car, the sounds that are typical of an accident can be used to trigger the collection of the vehicle positioning information from the different devices that are in the vicinity of the different physical sensors.

The same accident can trigger a sound analysis in many different devices. Assuming that at least some number of phones are within the hearing range of the accident, the use of acoustics can reduce the amount of traffic that needs to be sent out to detect and identify accidents in a significant manner.

The above three are but some of many examples where a control signal mechanism, specifically audio-based ones, can be used to control and improve the effectiveness of social sensing. Despite the relative inaccurate estimation of sounds using shallow audio analytics, the system can still provide tremendous value.

4 EXPERIMENTAL RESULTS

In this section, we present some experimental results on example use cases including the golf tournament scenario described above. We focus on the automatic detection of events from sounds using shallow models that are feasible for running at the edge.

The pipeline for sound classification and event detection includes feature extraction and classification. The feature extraction step extracts useful features from raw audio waveforms, such as FFT coefficients and MFCC coefficients [4], which are then used by the classifiers in the classification step. The classifiers are first trained on the training data and then evaluated on testing data.

4.1 Performance Metrics

We evaluate the performance of event detection with commonly used metrics. For a particular event of interest, the definitions of true/false positives and negatives are summarized in Table 1, which depend on whether the original sound contains an event and whether the algorithm detects that event.

Based on the above, the *precision* and *recall* are defined as follows:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

Intuitively, the precision means among all sounds where the algorithm says that there is an event, how many of them really have an event; the recall means among all real events, how many of them are detected by the algorithm.

If the focus is on the classification of different types of sounds, instead of detecting one type of sound out of all other types of sounds, one can also define the *accuracy* as the number of correctly classified sounds divided by the total number of them considered in the classification. We study the precision, recall, and accuracy for two different scenarios in the following.

Table 2: Results for golf event detection

Training data	Precision	Recall	Accuracy
Background + applause/cheering	90.6%	91.3%	90.7%
Background only	80.0%	83.7%	80.2%

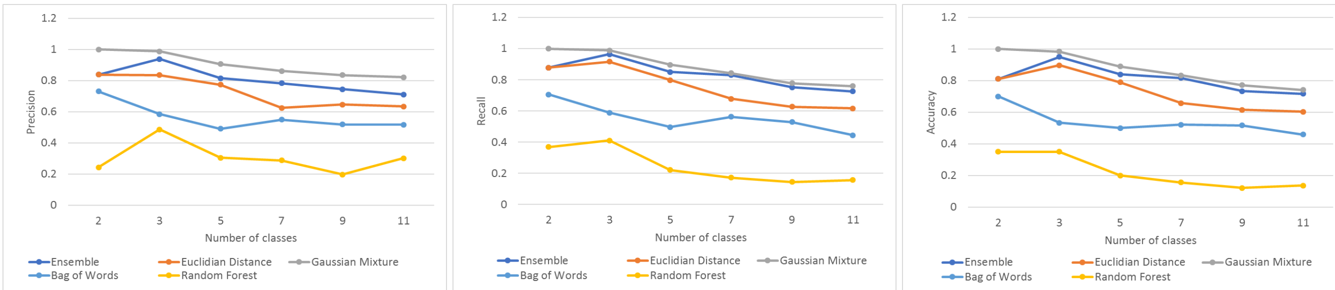
4.2 Golf Tournament Sounds

Several real golf competition videos are used to train the classification model used for crowd behavior detection. The videos are from different golf competitions and are rich in sound effects. The sound track of each video is decoded into the wave format and chopped into 5 second segments (or "sound clips"), and each segment is then labeled manually for training. For the purpose of golf crowd behavior detection, we simply divide sound clips into two categories: background sound and crowd behavior sound. The background sounds consist of regular background noise in the field and golf competition commentators' speech that do not include an event, and the crowd behavior sounds include events such as applause, cheering, and roaring sound by crowds (possibly in addition to the background sound). The sound dataset that we use in the experiment has 43 clips with an event, and 43 clips without event.

For this particular application scenario, we found that an MFCC feature extractor combined with a nearest neighbor classifier gives good performance. The nearest neighbor classifier saves all (subject to an upper limit) of the features of clips in the training dataset. In the classification phase, each new sound clip is classified as the same category as its closest neighbor in the feature space. The "closeness" is measured according to a distance metric. For the golf scenario, we found that using $\mathcal{L}^p$ norm with $p = 0.5$ as the distance function gives the best performance.

The nearest neighbor classifier can also be extended to support the detection of an unknown sound. If the distance exceeds a given threshold, we classify this sound as unknown, indicating that some interesting event might have happened. The threshold value has a direct impact on the trade-off between precision and recall; it can be tuned such that the precision and recall are approximately equal.

We study the performance of two different cases in our experiments. In the first case, we use both types of sounds (i.e., with and without event) for training. Here, we use K-fold evaluation with K=4 [10], where 75% of the total amount of sound clips are used for training and 25% are used for testing. There are four different ways of partitioning the training and testing dataset and the average results are shown. In the second case, we only use 5 background sound clips (without event) for training, and all the remaining sound clips are used for testing. This corresponds to the case where one has limited amount of training data, which does not contain any event, and the events will be detected using the unknown sound detection capability described above.



(a) (b) (c)

Figure 4: Performance results of multi-class event detection and classification: (a) precision, (b) recall, (c) accuracy.

The results are shown in Table 2. As one would intuitively expect, the first case performs better than the second case. However, it is interesting to note that the performance of second case is quite promising as well, particularly since only 5 clips are used for training the model. Because the number of each type of sounds are approximately equal in both cases, a random guess would give approximately 50% of precision, recall, and accuracy.

4.3 Multi-Class Office Sounds

In more complex environments, one may want to have the capability of detecting the presence of different types of sounds. To evaluate the performance of such cases, we study the performance using the DCASE 2016 dataset [1], which contains sounds of clearing throat, coughing, door slam, drawer, keyboard, keys dropping, knocking, human laughter, page turning, phone, and speech. The precision and recall for this scenario are defined for detecting one particular class of sound out of all other classes of sounds. In addition to considering the entire dataset, we also consider subsets that have a smaller number of classes.

Because this scenario is more complex, we use more advanced (but still shallow) classifiers, including the nearest neighbor with Euclidian distance, Gaussian mixture model, bag of words on audio features, random forest, as well as the ensemble of them. The results are shown in Figure 4 , which are obtained from K-fold evaluation with K=4 and multi-class training data. We see that the best classifier and the ensemble give promising results (around 80% of precision, recall, and accuracy). The ensemble is attractive due to its ability to always select the best set of classifiers, which may vary under different scenarios.

5 CONCLUSIONS

In this paper, we argue that using audio as a control knob for influencing social sensing applications is beneficial, and discuss several use cases where this abstract model can be applied. We evaluate the efficacy of the approach in one use case, specifically for identifying the interesting events that are happening in a golf tournament. While the concepts of the high level architecture are helpful in improving social sensing applications, the primary constraint in getting the desired benefits is the accuracy of audio sensing. Despite the inaccuracy, the audio analysis based control knob can be useful in reducing the volume and computational needs of social sensing applications in many different use cases.

ACKNOWLEDGMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

REFERENCES

[1] 2016. DCASE 2016 Challenge. (2016). http://www.cs.tut.fi/sgn/arg/dcase2016/challenge

[2] Maged N Kamel Boulos, Bernd Resch, David N Crowley, John G Breslin, Gunho Sohn, Russ Burtner, William A Pike, Eduardo Jezierski, and Kuo-Yu Slayer Chuang. 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International Journal of Health Geographics* 10, 1 (2011), 67.

[3] Bernd Girod, Vijay Chandrasekhar, David M Chen, Ngai-Man Cheung, Radek Grzeszczuk, Yuriy Reznik, Gabriel Takacs, Sam S Tsai, and Ramakrishna Vedantham. 2011. Mobile visual search. *IEEE signal processing magazine* 28, 4 (2011), 61–76.

[4] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy, and Kong-Pang Pun. 2006. An efficient MFCC extraction method in speech recognition. In *Proc. of IEEE International Symposium on Circuits and Systems, 2006.* IEEE.

[5] Hong Lu, Wei Pan, Nicholas D Lane, Tanzeem Choudhury, and Andrew T Campbell. 2009. SoundSense: scalable sound sensing for people-centric applications on mobile phones. In *Proceedings of the 7th international conference on Mobile systems, applications, and services.* ACM, 165–178.

[6] Anmol Madan, Manuel Cebrian, Sai Moturu, Katayoun Farrahi, and others. 2012. Sensing the health state of a community. *IEEE Pervasive Computing* 11, 4 (2012), 36–45.

[7] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Péter Boda. 2009. PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th international conference on Mobile systems, applications, and services.* ACM, 55–68.

[8] Rohan Narayana Murty, Geoffrey Mainland, Ian Rose, Atanu Roy Chowdhury, Abhimanyu Gosain, Josh Bers, and Matt Welsh. 2008. Citysense: An urban-scale wireless sensor network and testbed. In *2008 IEEE Conference on Technologies for Homeland Security.* IEEE, 583–588.

[9] Alun Preece, Colin Roberts, David Rogers, William Webberley, Martin Innes, and Dave Braines. 2016. From open source communications to knowledge. In *SPIE Defense + Security.* International Society for Optics and Photonics.

[10] Mervyn Stone. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)* (1974), 111–147.

[11] Shiqiang Wang. 2015. *Dynamic service placement in mobile micro-clouds.* Ph.D. Dissertation. Imperial College London.