ON CLASSIFICATION OF ENVIRONMENTAL ACOUSTIC DATA USING CROWDS

*Shan Zhang** *Aditya Vempaty†* *Susan E. Parks*⋆† *Pramod K. Varshney**

* Dept. of EECS, Syracuse University
† IBM T.J. Watson Research Center
⋆† Dept. of Biology, Syracuse University

ABSTRACT

In this work, we use crowds for acoustic classification of animal species in supervised and unsupervised manners. We demonstrate the effectiveness of the proposed triplet based crowdsourcing systems via actual experiments. Moreover, we propose a generalized 1-bit RPCA algorithm to further improve classification performance. The unique marriage of crowdsourcing and generalized 1-bit RPCA algorithm is shown to yield excellent performance for acoustic data classification.

Index Terms— acoustic data classification, crowdsourcing, generalized 1-bit RPCA, animal specie classification

1. INTRODUCTION

Biological applications for animal signal detection often result in a dataset of sub-optimal exemplars of signals, for example those degraded through propagation or masked with background noise. Automated detection of low SNR sounds of animals has been an on going challenge in signal processing for over two decades [1, 2]. Classification or recognition of detected low SNR sounds can be an even great challenge, with a number of approaches applied to attempt to detect signals from species of interest [3–5].

Crowdsourcing has become a popular paradigm for distributed decision making [6]. By decomposing complex problems into small-scale tasks that utilize human capabilities, crowdsourcing has shown its efficiency for distributed inference tasks [7, 8]. Typically, in a crowdsourcing system, human crowds are employed to solve easy-to-answer tasks, and solutions from the crowd members are combined to obtain a global inference goal by applying a fusion algorithm such as plurality and majority rule [9]. In [10], crowdsourcing and machine learning techniques are both used to classify whale calls and performances comparison is also presented.

In this work, we use crowdsourcing to solve the classification of environmental acoustic data by defining similarity comparisons among objects [11]. Crowds are asked to perform classification tasks using their significant perceptual capabilities, leading to the classification results to achieve a classification goal. Questions of the form "Is a more similar to b than to c" have been shown useful in obtaining similarity comparisons from crowds [12, 13]. Typically, we refer to these questions as triplets. In this work, we design a triplet-based sampling method to create similarity comparison tasks that are provided to crowds who perform the required data classification. Since human workers perform classification according to their own perception of similarity, they may make different classification decisions. To account for different perceptions by crowd members, we propose a generalized binary Robust Principal Component Analysis (RPCA) algorithm to further improve classification performance via the marriage of crowds and machines. To the authors' knowledge, the application of RPCA [14] to further improve the classification performance of crowdsourcing systems is done for the first time.

This work was supported in part by NSF grant EE-1340669.

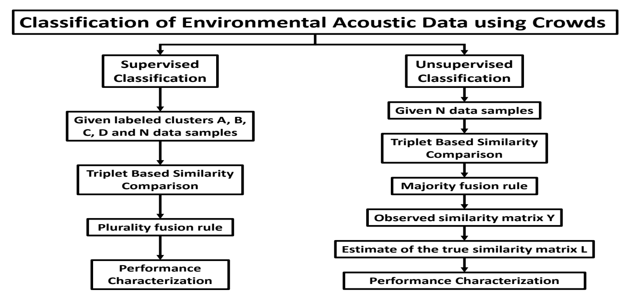


Fig. 1: Classification framework using crowds and 1-bit RPCA.

Supervised and unsupervised crowdsourcing systems for low SNR acoustic data to classify animal species are proposed that are shown in Fig. 1. Given a collection of objects to be clustered, first triplet based similarity comparison tasks are created, and crowds are asked to answer questions based on similarity triplets. Simple plurality rule and majority rule are used to combine the crowds' local decisions and determine the final clusters. For the unsupervised classification system, we propose a generalized 1-bit RPCA algorithm to further improve the classification performance. We make two main contributions. First, we propose triplet based crowdsourcing techniques for the classification of animal acoustic data. Theoretical performance is characterized for both supervised and unsupervised classification systems. Second, we propose the generalized 1-bit RPCA algorithm that is shown to be efficient and robust in improving the classification performance of the crowdsourcing system. Simulation results based on an animal sound based dataset are provided to demonstrate the efficacy of our classification systems. It is expected that the methodology can be scaled to larger datasets.

2. SUPERVISED CROWDSOURCING SYSTEM WITH INDIVIDUAL CROWD PARTICIPANTS

Our supervised crowdsourcing system for classification of animal acoustic data using independent crowd participants is described in this section. Triplet comparisons of the form "Is a more similar to b than to c" are designed. Plurality voting is used to combine partial decisions from individual workers.

2.1. Model

Our crowdsourcing system consists of T independent crowd participants, where crowds are asked to perform the classification tasks based on animal audio clips. We assume that each worker t has

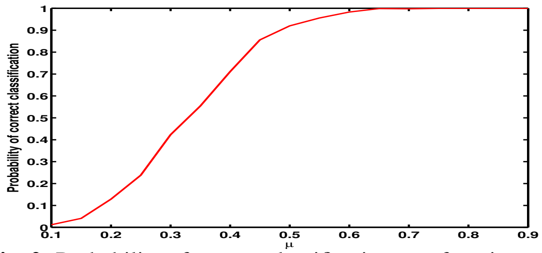


Fig. 2: Probability of correct classification as a function of μ.

a reliability metric p_t for $t = 1, 2, \ldots, T$ which is the probability that he/she perfectly classifies the given audio clip. Therefore, for an M-ary classification task, p_t is the probability that the tth worker decides in favor of class i when the true class is i, for $i = 1, \ldots, M$. We choose the Beta distribution with parameters α and β, Beta(α, β), to characterize the crowd workers' reliability metrics.We assume that we have N samples, and each sample belongs to one of the known M classes.

2.2. Triplet Based Similarity Comparison

Since workers may not be experts in recognizing animals through their sounds, they may not be able to directly classify and so we should ask simpler questions. It is easier for humans to compare two objects and determine whether they are similar or not than determining the exact identity of the object. In our context, it is easier for them to determine if a given audio clip of an animal call sounds similar to another than determining the true species of the animal.

For any set of n samples, there are at most n^3 unique triplets. It would be expensive and time consuming if all the triplets are collected. Generally, we only need to collect several thousand triplets until enough triplets are obtained to achieve a crowdsourcing goal. In this work, we first create triplets and distribute them to workers. Suppose we have N samples, and all the samples belong to one of the M clusters. Triplets are obtained in the following manner (see an illustrative example in Fig. 4): For each of the unlabeled samples $[1, 2, \cdots, N]$, randomly select two labeled samples $[1, 2, \cdots, M]$. There are a total of $\binom{M}{2}$ such triplets for each sample. Due to the desire to maintain high classification accuracy of the crowdsourcing system, it is necessary and important to have redundancy. For each such triplet, we add a redundancy of Q, i.e., Q copies of the triplet are made and distributed to crowd workers. Crowd workers are asked to decide which cluster the unlabeled sample is more similar to.

2.3. Plurality Voting Rule

In our crowdsourcing system, we apply plurality voting to combine local decisions by crowd workers [9]. We assume that independent responses corresponding to R triplets are available regarding one specific unlabeled sample. After combining the decisions from R triplets, a sample is believed to belong to the class that receives the largest number of worker votes. Since each worker has a probability to be wrong, correct classification is made when the fused result is correct. The strength of the plurality voting method is its error correction capability.

2.4. Performance Characterization

Proposition 1. *The expected probability of correct classification using the plurality fusion rule [9, 15] is:*

$$P_c(\mu) = \frac{1}{M}\sum_{i=1}^{M}\sum_{\mathbf{m}}\frac{1}{|\{\max m_i\}|}\binom{R}{\mathbf{m}}\mu^{m_{(M)}}(1-\mu)^{M-m_{(M)}},$$

where $R = \binom{M}{2}(Q+1)$ is the number of questions asked about a single sample, $\mathbf{m} = (m_1, \ldots, m_M)$, $\{\max m_i\}$ is the set of maxima corresponding to each class, $\binom{R}{\mathbf{m}} = \frac{R!}{m_1!m_2!\cdots m_M!}$, $|\cdot|$ is the cardinality, $m_{(M)}$ is the maximum of all m_is, and μ is the mean value of each worker's reliability p_t.

Proof. In a plurality-based approach, $R = \binom{M}{2}(Q+1)$ workers send their decisions regarding a given sample, and the plurality rule with random tie-breaking is used to determine the final class. Let P_c be the probability of a given sample being correct given the reliabilities of the workers that have answered the R triplets. Then, $P_c = \frac{1}{M}\sum_{i=1}^{M} P_{c,i}$, where $P_{c,i}$ is the probability that the true class is i and the final fused class is also i. This value is given as

$$P_{c,i} = \sum_{\mathbf{m}}\sum_{\forall G_m}\prod_{k\in G_{m_{(M)}}} p_t \prod_{k\notin G_{m_{(M)}}}(1-p_t) + \sum_{\mathbf{m}}\frac{1}{L}\sum_{\forall G_m}\prod_{k\in G_{m_{(M)}}} p_t \prod_{k\notin G_{m_{(M)}}}(1-p_t),$$

where $G_{m_{(M)}}$ is a set of $m_{(M)}$ out of R workers who correctly classify the given sample and L is the number of values equal to $m_{(M)}$; the second part of the expression is due to the tie-breaking rule. Since the reliabilities are i.i.d. across workers and triplets, the expected probability of correct decision P_c is:

$$E[P_c] = \frac{1}{M}\sum_{i=1}^{M} E[P_{c,i}],$$
$$= \frac{1}{M}\sum_{i=1}^{M}\sum_{\mathbf{m}}\frac{1}{|\{\max m_i\}|}\sum_{\forall G_m}\prod_{k\in G_{m_{(M)}}}E[p_t]\prod_{k\notin G_{m_{(M)}}}E[(1-p_t)],$$

where the expectation is with respect to the probability of participants' reliability $\mathbf{p}$, $\mathbf{p} = [p_1, p_2, \cdots, p_T]$. Therefore,

$$P_c(\mu) = \frac{1}{M}\sum_{i=1}^{M}\sum_{\mathbf{m}}\frac{1}{|\{\max m_i\}|}\binom{R}{\mathbf{m}}\mu^{m_{(M)}}(1-\mu)^{M-m_{(M)}}.$$

□

From the above proposition, observe that the performance of the supervised crowdsourcing system is higher on an average for a better crowd (higher μ), and for a higher value of R. This implies that higher redundancy (Q) results in improved performance. Fig. 2 shows the classification performance with varying μ when $N = 100$, $M = 4$, $Q = 2$ and $R = 18$. We assume that all individual workers have the same expected reliability of perfectly classifying the given audio samples. For each audio sample, we have 18 individual local decisions with respect to the comparison of this sample with each cluster and then apply plurality fusion rule to obtain the global classification results. As we can see from the figure, the performance improves with increasing μ.

3. CROWDSOURCING EXPERIMENT FOR SUPERVISED CLASSIFICATION

We apply crowdsourcing and the plurality rule to classify acoustic data with known cluster information, i.e., the ground truth and analyze our experimental results.

3.1. Data

In our experiment, we have $N = 100$ acoustic samples along with the spectrum image files, that belong to $M = 4$ clusters (3 birds and 1 mammal): Blue Jay, Black Capped Chickadee, Eastern Chipmunk, and Brown Creeper. For the sake of notational simplicity, we use letters A to D to denote the four species of animals. Fig. 3 shows a sample spectrum image that was provided along with the audio file that is to be classified. Spectrum images were added to provide a visual dimension to the data besides the audio dimension, thereby making it easier for humans to perform the task. Having both visual and audio information makes it easier for humans to perform the similarity comparisons.

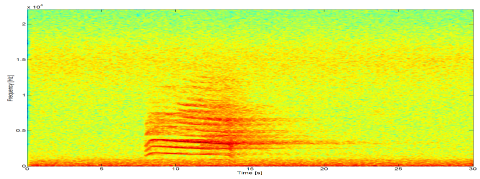
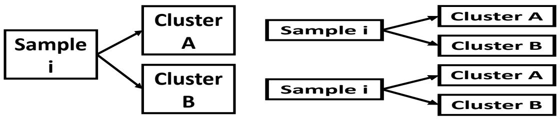


Fig. 3: Spectrum of Blue Jay acoustic sample.



(a) One triplet example. (b) Redundancy of the triplet.

Fig. 4: An example of triplet with redundancy.

3.2. Experiment

$N = 100$ unlabeled audio samples and their corresponding spectra belonging to $M = 4$ classes (A, B, C, D) are used for the experiment. Fig. 4a shows the way to create one triplet for supervised classification with 4 clusters. Sample i can be compared by asking class A vs class B question to determine which class it "sounds" closer to. Similarly, there are 5 other triplets that can be constructed using the clusters $\{(A, C), (A, D), (B, C), (B, D), (C, D)\}$. Addition of redundancy is illustrated in Fig. 4b where two additional copies of the triplet are made resulting in $Q = 2$. For each such combination, triplets were created for each sample. For each triplet, a redundancy of $Q = 2$ was added. Therefore, the total number of triplets were $\binom{M}{2}(Q + 1) = 1800$. We randomly distributed these triplets to $T = 36$ workers ensuring that the same worker does not answer to the same triplet twice. Each participant responded to 50 triplets, with three possible choices. Besides the two labeled clusters, we also provided a "Don't know" option, to address the case when the sample's true class is neither of the given two options and the crowd worker also feels the same. Fig. 5 shows a sample triplet consisting of an unlabeled sample to the left that is compared with two labeled samples on the right. After participants sent their decisions, we applied the plurality rule to get the final classification result. Summary of results is presented in Table 1.

3.3. Experimental Results

Table 1 shows that we have a very good classification performance using crowds. The slightly lower performance in some cases was due to the fact that six people did not finish their subtasks. Also, note that class C (Eastern Chipmunk), has perfect classification as it can be distinguished from others when using similarity triplets but is often confusing otherwise. Also observe that for $M = 4$ and $Q = 2$, the value of $R = 18$ and we have observed the average correct classification probability of $P_c = 0.92$. The confusion matrix is shown in Table 2.

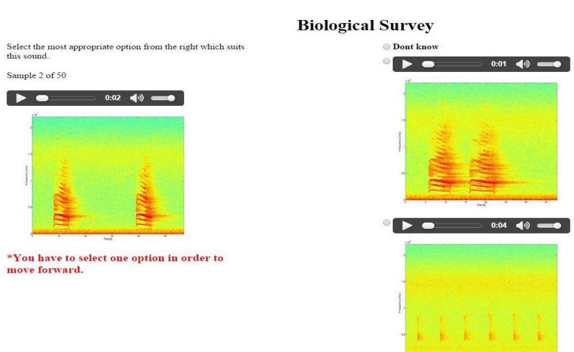


Fig. 5: A sample similarity triplet that a crowd worker responds to.

Table 1: Classification Results using Plurality Rule

Classes	Number of Samples	Errors	Correct Probability
A	25	1	0.96
B	25	4	0.84
C	25	0	1
D	25	4	0.84

Table 2: Confusion Matrix

True \ Classified	A	B	C	D	Don't know
A	24	0	0	0	1
B	0	21	1	3	0
C	0	0	25	0	0
D	1	3	0	21	0

4. UNSUPERVISED CLASSIFICATION WITH INDIVIDUAL CROWD PARTICIPANTS

4.1. Model

Our unsupervised crowdsourcing system consists of N audio clip samples to be classified and T independent crowd participants with associated reliability metric p_t for $t = 1, 2, \ldots, T$. For the unsupervised classification task, p_t is the probability that the tth worker makes a correct decision on the comparison of sample i and sample j, for $i, j = 1, 2, \ldots, N$. In particular, it is the probability that the decision provided by the worker is k, when the true comparison decision of sample i and j is k, $k = 0, 1$. For notational simplicity, let $[N]$ denote the integer set $[1, 2, \ldots, N]$. We also assume that all the clusters are disjoint, and all crowd workers make their own decisions independently. Different from the supervised crowdsourcing system, the number of clusters is unknown. Triplets are obtained in the following way: For each sample i, $i \in [N]$, randomly choose distinct pairs from the entire set of samples $[N]$ (each sample can only be chosen once) till all the samples are chosen. There are a total number of $\frac{N}{2}$ such triplets with sample i and a distinct pair. After obtaining triplets for all the samples $[N]$, we add a redundancy of Q for each triplet and distribute the set of all triplets to crowd workers. For each triplet, workers are asked to decide if the samples j and h are similar to sample i and make comparison decisions $k \in \{0, 1\}$ for sample pairs (j, i) and (h, i). After collecting the decisions from individual workers, the majority voting rule is applied to obtain the global comparison results in an $N \times N$ matrix called the observed similarity matrix $\mathbf{Y}$ where $Y_{ij} = 1$ ($Y_{ij} = 0$) denotes that a majority of the individual workers believe that sample i and j are (not) from the same cluster. Since errors are inevitable in a crowdsourcing system, the generalized 1-bit RPCA approach is applied to improve classification performance.

4.2. Performance Characterization via Crowdsourcing

Suppose H ($H < T$) is the number of participants that respond to the similarity task between samples i and j.

Proposition 2. *The expected probability of correct classification for unsupervised classification using the majority fusion rule [9, 15] is:*

$$P_c(\mu) = \left[\frac{1}{2}\left(1 + S_{H,\mu}\left(\frac{H}{2}\right) - S_{H,(1-\mu)}\left(\frac{H}{2}\right)\right)\right]^{N^2} \quad (1)$$

where $H = Q + 2$ is the number of questions regarding sample $i, j \in [N]$, $S_{H,\mu}(\cdot)$ is the complementary cdf of the binomial random variable $\mathcal{B}(H, \mu)$ and μ is the mean value of each worker's reliability p_t.

Proof. In a majority-based approach, $H = Q + 2$ workers send their local decisions regarding samples i, j in triplets, $i, j \in [N]$.

Let P_c be the probability of a correct decision given the reliabilities of the workers that have answered the H questions. Then, $P_c = \prod_{j=1}^{N}\prod_{i=1}^{N} P_{c,\mathbf{p}}^{ij}$, where $P_{c,\mathbf{p}}^{ij}$ is the probability that the true comparison decision between samples i and j is k and the final fused decision by combining the partial decisions from crowd workers is also k, $k = 0, 1$. This value is given as $P_{c,\mathbf{p}}^{ij} = \frac{P_d+1-P_f}{2}$, where P_d is the probability that a worker declares samples i and j as '1' when the true decision is '1' and P_f is the probability that a worker declares samples i and j as '1' when the true decision is '0'. Under the majority rule for sample i, j,

$$P_d = \sum_{k=\lfloor \frac{H}{2}+1 \rfloor}^{H} \sum_{\forall G_k} \prod_{g\in G_k} p_t \prod_{g\notin G_k} (1-p_t),$$

$$P_f = \sum_{k=\lfloor \frac{H}{2}+1 \rfloor}^{H} \sum_{\forall G_k} \prod_{g\in G_k} (1-p_t) \prod_{g\notin G_k} p_t,$$

where G_k is the set of k out of H workers who determine '1' and p_t is the probability of the tth worker making a correct decision for sample i, j. Since the reliabilities are i.i.d. across workers and triplets, $E[P_c] = \prod_{j=1}^{N}\prod_{i=1}^{N} E[P_{c,\mathbf{p}}^{ij}]$, where the expectation is with respect to the probability of participants' reliability $\mathbf{p}$, $\mathbf{p} = [p_1, p_2, \cdots, p_T]$.

$$E[P_d] = \sum_{k=\lfloor \frac{H}{2}+1 \rfloor}^{H} \binom{H}{k} \mu^k (1-\mu)^{H-k} = S_{H,\mu}\left(\frac{H}{2}\right),$$

$$E[P_f] = \sum_{k=\lfloor \frac{H}{2}+1 \rfloor}^{H} \binom{H}{k} (1-\mu)^k \mu^{H-k} = S_{H,(1-\mu)}\left(\frac{H}{2}\right).$$

Therefore, we get the desired result. $\square$

4.3. Generalized 1-bit RPCA Algorithm

In supervised acoustic animal classification experiments, fairly good classification performance was obtained using crowds. However, since crowds have very different criteria in making decisions, a significant number of errors occur in classifying the audio clips in unsupervised classification. The observed similarity matrix $\mathbf{Y}$ is obtained by applying the majority rule on all sample pairs (i, j) for $i, j \in [1, 2, \ldots, N]$. The errors due to unreliable crowds can be assumed to be sparse corruptions, and the true similarity matrix $\mathbf{L}$ is typically low rank [16]. Inspired by the RPCA algorithm [14], we propose a generalized 1-bit RPCA algorithm to estimate the true similarity matrix from a noisy one.

Note that the similarity matrices $\mathbf{Y}$ and $\mathbf{L}$ are binary. We represent the noisy similarity matrix $\mathbf{Y}$ obtained from unsupervised crowdsourcing classification as $\mathbf{Y} = \mathbf{L} \oplus \mathbf{S}$, where $\mathbf{S}$ is a binary corruption matrix, and $\oplus$ denotes binary addition. In our model, we assume if there is noise ($S_{ij} = 1$), L_{ij} will be flipped from 1 to 0 or 0 to 1. The true similarity matrix $\mathbf{L}$ is defined such that $L_{ij} = 1$ if i, jth sample belong to the same cluster, and 0 otherwise. Moreover, the noisy observed matrix $\mathbf{Y}$ is given by

$$Y_{i,j} = \begin{cases} L_{i,j} & \text{w.p.} \quad P_{c,\mathbf{p}}^{i,j} \\ 1 - L_{i,j} & \text{w.p.} \quad 1 - P_{c,\mathbf{p}}^{i,j} \end{cases} \tag{2}$$

where $P_{c,\mathbf{p}}^{i,j}$ is the probability that the true comparison decision is k of sample i, j and the fused decision given by unsupervised crowdsourcing system is also k, $k = 0, 1$. In other words, the noise matrix $\mathbf{S}$ follows Bernoulli distribution with probability $P_{c,\mathbf{p}}^{i,j}$.

To approximate $\mathbf{L}$, we aim to maximize the log-likelihood function of the optimization variable $\mathbf{Y}$ [17] subject to a set of convex constraints. In our case, the log-likelihood function is given by

$$\mathcal{L}_{\mathbf{Y}} := \sum_{(i,j)\in\mathbb{R}^{N\times N}} \log(\mathbb{I}_{[Y_{i,j}=1]} P(Y_{i,j} \mid L_{i,j}) + \mathbb{I}_{[Y_{i,j}=0]} P(Y_{i,j} \mid L_{i,j})).$$

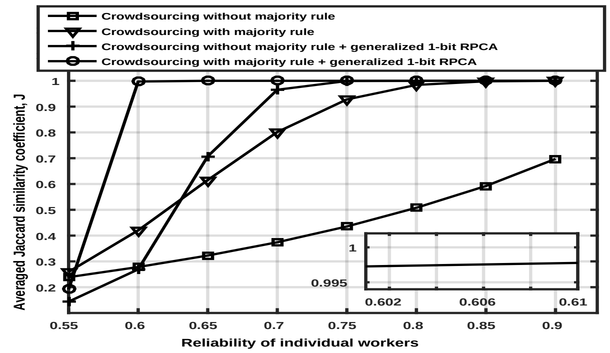


Fig. 6: Averaged Jaccard similarity coefficient as a function of μ.

The problem of recovering $\mathbf{L}$ is cast as

$$\begin{aligned} &\text{maximize } \mathcal{L}_{\mathbf{Y}} \\ &\text{subject to } 0 \le L_{i,j} \le 1 \quad \text{for all } i, j \\ &\qquad \|\mathbf{L}\|_* \le \sigma \end{aligned} \tag{3}$$

where the nuclear norm constraint $\|\mathbf{L}\|_* \le \sigma$ is a relaxation of the low-rankness of matrix $\mathbf{L}$ [18]. Parameter $\sigma > 0$ is used to tune the rank of $\mathbf{L}$ and is determined via cross-validation or AIC/BIC [19]. Binary-valued $\mathbf{L}$ is relaxed as $0 \le L_{i,j} \le 1$, $i, j \in [N]$ [20].

4.4. Simulation Results

Assume we have $N = 100$ samples to be classified, and $H = 18$ individual crowd workers who make similarity decisions of sample i and j, $i, j \in [100]$. As with supervised classification, we assume all individual crowd workers to have the same expected reliability of perfectly classifying the given audio samples. To characterize classification performance of the unsupervised system, we employ Jaccard similarity coefficient to compare the similarity matrices. The Jaccard index of similarity between $\mathbf{L}$ and $\hat{\mathbf{L}}$ is $J(\mathbf{L}, \hat{\mathbf{L}}) = \frac{|\mathbf{L}\cap\hat{\mathbf{L}}|}{|\mathbf{L}|+|\hat{\mathbf{L}}|-|\mathbf{L}\cap\hat{\mathbf{L}}|}$, where $|\cdot|$ denotes the size of the set. When $J = 0$, the recovered matrix $\hat{\mathbf{L}}$ has no intersection with true similarity matrix $\mathbf{L}$. The case $J = 1$ implies that these two matrices are the same.

Fig. 6 shows the classification performance of the unsupervised classification system. We used 100 Monte Carlo runs to obtain the averaged Jaccard similarity coefficient for four classification methods as a function of reliability of individual workers. First, we demonstrate the performance of the proposed crowdsourcing system where majority rule is used to fuse partial solutions from crowds. Second, for comparison, we show the classification performance using crowds without applying majority rule, where only one single crowd worker performs the comparison of sample i and j, $i, j \in [100]$. Then, we present further classification performance by applying the proposed generalized 1-bit RPCA algorithm to the above crowdsourcing methods. As we can see from the figure, the performance of the unsupervised classification system improves with increasing μ which is consistent with that of our supervised model. Unfortunately, the recovery performance when using generalized 1-bit RPCA is poor when $\mu \le 0.6$. However, when $\mu \ge 0.7$, we can see that the generalized 1-bit RPCA algorithm gives good performance in recovering the true similarity matrix from the noisy similarity matrix. Moreover, the recovery performance is pretty good even without applying majority rule in the crowdsourcing system when $\mu \ge 0.7$. Furthermore, the proposed generalized 1-bit RPCA algorithm works nearly perfectly when the majority rule is applied in the crowdsourcing system. This is because the probability of correct classification after applying the majority rule is already pretty high (more than 0.7) and the true similarity matrix can be perfectly recovered with high probability.

5. REFERENCES

[1] David K Mellinger and Christopher W Clark, "Methods for automatic detection of mysticete sounds," *Marine & Freshwater Behaviour & Phy*, vol. 29, no. 1-4, pp. 163–181, 1997.

[2] David K Mellinger and Christopher W Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, pp. 3518–3529, 2000.

[3] Aki Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. IEEE, 2003, vol. 5, pp. V–545.

[4] Arja Selin, Jari Turunen, and Juha T Tanttu, "Wavelets in recognition of bird sounds," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 141–141, 2007.

[5] Forrest Briggs, Balaji Lakshminarayanan, Lawrence Neal, Xiaoli Z Fern, Raviv Raich, Sarah JK Hadley, Adam S Hadley, and Matthew G Betts, "Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach," *The Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4640–4650, 2012.

[6] Jeff Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, June 2006.

[7] Don Tapscott and Anthony D Williams, *Wikinomics: How mass collaboration changes everything*, Penguin, 2008.

[8] Don Tapscott and Anthony D Wiliams, "Macrowikinomics: Rebooting business and the world, 2010," .

[9] Aditya Vempaty, Lav R. Varshney, and Pramod K. Varshney, "Reliable crowdsourcing for multi-class labeling using coding theory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 667–679, Aug. 2014.

[10] Lior Shamir, Carol Yerby, Robert Simpson, Alexander M von Benda-Beckmann, Peter Tyack, Filipa Samarra, Patrick Miller, and John Wallin, "Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 953–962, 2014.

[11] Jinfeng Yi, Rong Jin, Anil K. Jain, and Shaili Jain, "Crowdclustering with sparse pairwise labels: A matrix completion approach," in *Proc. AAAI Workshop Human Comput. (HCOMP'12)*, Jul. 2012, pp. 1–7.

[12] Omer Tamuz, Ce Liu, Serge Belongie, Ohad Shamir, and Adam Tauman Kalai, "Adaptively learning the crowd kernel," in *Proc. 28th Int. Conf. Mach. Learn. (ICML 2011)*, Jun.-Jul. 2011, pp. 673–680.

[13] Laurens van der Maaten and Kilian Weinberger, "Stochastic triplet embedding," in *Proc. 2012 IEEE Int. Workshop Machine Learning Signal Process. (MLSP)*, Sep. 2012, pp. 1–6.

[14] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[15] Pramod K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, 1996.

[16] Ali Jalali, Yudong Chen, Sujay Sanghavi, and Huan Xu, "Clustering partially observed graphs via convex optimization," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1001–1008.

[17] Andrew S Lan, Christoph Studer, and Richard G Baraniuk, "Matrix recovery from quantized and corrupted measurements," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4973–4977.

[18] Maryam Fazel, Haitham Hindi, and Stephen P Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *American Control Conference, 2001. Proceedings of the 2001*. IEEE, 2001, vol. 6, pp. 4734–4739.

[19] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *The elements of statistical learning*, vol. 1, Springer series in statistics Springer, Berlin, 2001.

[20] Samet Oymak and Babak Hassibi, "Finding dense clusters via "low rank + sparse" decomposition," *arXiv preprint arXiv:1104.5186*, 2011.