

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.				
1. REPORT DATE (DD-MM-YYYY) 10-03-2016		2. REPORT TYPE Conference Proceeding		3. DATES COVERED (From - To) -
4. TITLE AND SUBTITLE Fuzzy Clustering of Multiple Instance Data		5a. CONTRACT NUMBER W911NF-13-1-0066		
		5b. GRANT NUMBER		
		5c. PROGRAM ELEMENT NUMBER		
6. AUTHORS Andrew Karem, Hichem Frigui		5d. PROJECT NUMBER		
		5e. TASK NUMBER		
		5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Louisville 2301 S. Third Street Jouett Hall Louisville, KY 40208 -1838			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 63184-CS.14	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.				
14. ABSTRACT We introduce a new algorithm to identify multiple target concepts when data are represented by multiple instances. A multiple instance data sample is characterized by a bag that contains multiple feature vectors, or instances. Each bag is labeled as either positive or negative. However, the labels of the instances within each bag are unknown. A bag is labeled as positive if and				
15. SUBJECT TERMS Diverse Density; Multiple Instance learning; Landmine detection				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU		
			19a. NAME OF RESPONSIBLE PERSON Hichem Frigui	
			19b. TELEPHONE NUMBER 502-852-2009	

Report Title

Fuzzy Clustering of Multiple Instance Data

ABSTRACT

We introduce a new algorithm to identify multiple target concepts when data are represented by multiple instances. A multiple instance data sample is characterized by a bag that contains multiple feature vectors, or instances. Each bag is labeled as either positive or negative. However, the labels of the instances within each bag are unknown. A bag is labeled as positive if and only if at least one of its instances is positive and negative if and only if all of its instances are negative. First, we define a fuzzy Multi-target concept Diverse Density (MDD) metric. The MDD is maximized when the target concepts correspond to dense regions in the feature space with maximal correlation to instances from positive samples, and minimal correlation to instances from negative samples. Then, we develop an iterative algorithm to optimize the MDD and identify K target concepts simultaneously. The proposed algorithm, called Fuzzy Clustering of Multiple Instance data (FCMI), is tested and validated by using it to analyze data of buried landmines collected using a ground penetrating radar sensor. We show that the FCMI algorithm can identify distinct target concepts that correspond to mines of different types buried at different depths. We also show that

Conference Name: IEEE Conference on Fuzzy Systems

Conference Date: August 02, 2015

Fuzzy Clustering of Multiple Instance Data

Andrew Karem and Hichem Frigui

Multimedia Research Lab, CECS Dept.

University of Louisville

Louisville, KY 400292

Email: adkare01@gmail.com or h.frigui@louisville.edu

Abstract—We introduce a new algorithm to identify multiple target concepts when data are represented by multiple instances. A multiple instance data sample is characterized by a *bag* that contains multiple feature vectors, or *instances*. Each bag is labeled as either positive or negative. However, the labels of the instances within each bag are unknown. A bag is labeled as positive if and only if at least one of its instances is positive and negative if and only if all of its instances are negative. First, we define a fuzzy Multi-target concept Diverse Density (MDD) metric. The MDD is maximized when the target concepts correspond to dense regions in the feature space with maximal correlation to instances from positive samples, and minimal correlation to instances from negative samples. Then, we develop an iterative algorithm to optimize the MDD and identify K target concepts simultaneously. The proposed algorithm, called Fuzzy Clustering of Multiple Instance data (FCMI), is tested and validated by using it to analyze data of buried landmines collected using a ground penetrating radar sensor. We show that the FCMI algorithm can identify distinct target concepts that correspond to mines of different types buried at different depths. We also show that FCMI can be used to label individual instances within each bag.

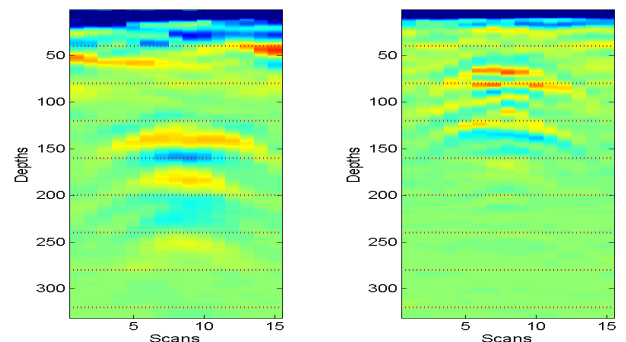
I. INTRODUCTION

Standard machine learning problems characterize an individual data sample by a single representative feature vector. For many applications, such as drug activity prediction [1] and landmine detection [2], each individual data sample may be represented by multiple features, each of which has ambiguous label. Dietterich et al. [1] proposed the Multiple Instance Learning (MIL) framework for identifying and modeling such problems. Under this framework, each data sample is represented by one class-labeled “bag,” that contains an arbitrary number of unlabeled “instances”, each of which is a single feature vector in the feature space. The machine learning task within the MIL framework consists of identifying bags, along with their subset of instances, that can be used to learn a classifier to label new bags.

To illustrate the need for MIL, we consider the application of landmine detection using ground penetrating radar (GPR). The GPR sensor is mounted on a vehicle and collects 3-dim data as the vehicle moves. The first 2 dimensions (down-track and cross-track) refer to the spatial location on the ground while the 3rd dimension refers to the depth. Typically, in labeled training data, the spatial location is known, but the depth is not. To illustrate this data, in figure 1 we display the GPR signatures of the same mine buried at 3 in deep in two geographically different sites. We only show a 2-D view (down-track, depth) of the alarms. First, we note that the actual target signature does not extend over all depth values. Thus, extracting one global feature vector from the alarm may not discriminate between mines and clutter effectively. To

overcome this limitation, multiple features should be extracted from small windows at different depths [3], [4]. For instance, in figure 1 we show 8 windows (typically, more overlapping windows are used). The main challenge in developing a classifier for this application is the selection of the appropriate depth for training. For instance, knowing the burial depth (3in) in figure 1 is not sufficient to identify the best window for training. In addition to soil properties, the true signature depth depends on other factors such as mine type and environmental conditions. In figure 2 we display the GPR signature of a large mine and a small mine. As it can be seen, for the large target, the signature can extend over 3 or 4 consecutive windows while the signature of a small window does not extend beyond one window. In Section (IV), we will show that using an MIL approach, each alarm would be represented by a bag of features extracted from multiple depths. Within each bag, some features would correspond to the mine signature while other features would correspond to background. The label of each instance is not known.

Other applications where the MIL framework has proved to be effective include automated image annotation [5], text document classification [6], speaker identification [7], and many others [8]



(a) Target at site 1

(b) Target at site 2

Fig. 1. Depth of target signature depends on the soil properties of the site. The same mine type is buried at 3in deep in both sites.

Since its formal introduction, MIL research has focused on supervised learning. Existing methods typically rely on two main approaches [8]. The first one concatenates features from all instances of a bag into one feature vector and utilizes highly sparse learning algorithms to learn relevant features/instances [9], [1], [10]. In the second approach, first a collection of *prototypes* is identified. Then, using these prototypes, each bag is mapped to a point in a new feature space and conventional

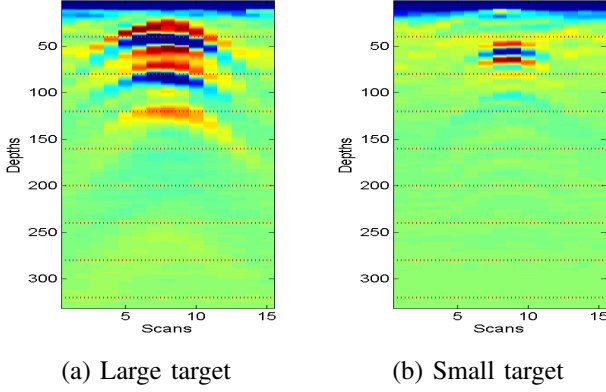


Fig. 2. Size of target signature depends on the target type

classification algorithms are used [11], [12], [13].

In this paper, we focus on unsupervised learning for multiple instance data. Our approach, called Fuzzy Clustering of Multiple Instance data (FCMI), strives to identify dense regions in the feature space with maximal correlation to instances from positive samples, and minimal correlation to instances from negative samples. The proposed FCMI algorithm uses a fuzzy clustering approach [14] to extend the Diverse Density model [9] to identify multiple target concepts simultaneously.

The organization of the rest of this paper is as follows. In Section II, we review related work and highlight the need for our approach. In Section III, we introduce the objective function of the FCMI and derive the necessary conditions to optimize it. In Section IV, we report experimental results and we conclude in Section V.

II. RELATED WORK

Initial contemplation of the need to represent data samples with more than a single feature vector can be traced back to (at the latest) two major applications: the need to predict bonding activity in drug design [15], and the problem of handwritten digit recognition [16]. To the best of our knowledge, Dietterich, et. al [1] were the first to formalize the definition and requirements of the traditional bag-instance Multiple Instance Learning framework. As a solution, they proposed a simple algorithm, called the Axis-Parallel Rectangles (APR). The APR constructs a set of boundaries in the problem feature space that enclose at least one instance from every positive sample in a training dataset, while excluding as many instances from negative data samples as is possible.

The next major step in MIL research was the formulation of the Diverse Density (DD) approach [9]. In [9], the author defines the Diverse Density metric which combines the cumulative probability that the positive bags are correlated with a given point of interest, and the cumulative probability that negative bags are not correlated with it. The DD algorithm seeks to identify the point of interest that maximizes the DD metric. This point is called *target concept*. The DD algorithm spurred several direct variations designed to improve performance or convergence efficiency. For example, the EM-DD algorithm [10] is a variation where an Expectation-Maximization algorithm is used to optimize the DD metric

and identify the target concept. Another research direction has used the DD metric to analyze the relationship between regions of the feature space and bags (collectively or individually) and identify multiple target concepts. The multiple concepts are needed to capture the within-class variations. The learned concepts are then used to perform feature space mapping (similar to kernel space transformation) to convert the multiple instance features to single-vector features, upon which conventional learning methods can be applied. Examples of such methods include DD-SVM [11] and MILES [12]. The above approaches learn the multiple concepts sequentially. First, they repetitively optimize the single concept DD metric using different initialization. Then, a validity measure is used to identify meaningful and diverse target concepts.

Multiple target concept learning in multiple instance data can be viewed as a clustering problem. Within the clustering community, it is well-known that extracting one cluster at a time is not effective. In fact, using this approach, only points within the cluster of interest will be considered inliers. Points in other clusters will be treated as outliers. Thus, when the expected number of clusters is larger than two, even very robust algorithms will break down. A more common practice is to define and optimize an objective function that seeks multiple clusters simultaneously. The K-Means [17] and the EM [18] algorithms fall into this category. Moreover, fuzzy objective functions [14], [19], [20] that allow data samples to belong to multiple clusters with various membership degrees has proved to be more reliable.

In the following, we use fuzzy clustering concepts to define a Multi-target concept Diverse Density (MDD) metric. We show that multiple concepts can be identified simultaneously by optimizing the proposed MDD metric.

III. FUZZY CLUSTERING OF MULTIPLE INSTANCE DATA

Let $\mathcal{B} = \{B_1, \dots, B_n, \dots, B_N\}$ represent the set of data samples. Each bag, $B_n = \{b_{n1}, \dots, b_{ni}, \dots, b_{nI}\}$, has I instances¹ and each instance, $b_{ni} = \{b_{ni1}, \dots, b_{niF}, \dots, b_{niF}\}$, is an F -dimensional feature vector. In MIL, a bag is labeled as positive (class of interest), B^+ , if and only if at least one of its instances is positive. Similarly, a bag is labeled negative, B^- , if and only if all of its instances are negative. We assume that our data has N_{pos} positive bags and N_{neg} negative bags such that $N_{pos} + N_{neg} = N$. Let $\mathcal{B}^+ = \{B_1^+, \dots, B_{N_{pos}}^+\}$ and $\mathcal{B}^- = \{B_1^-, \dots, B_{N_{neg}}^-\}$ denote the subsets of positive and negative bags respectively.

In MIL, each object is represented by multiple instances and no information about the relevance of each feature is unknown. Typically, only one or a few instances are relevant. Thus, this type of data has an additional ambiguity dimension making it more appropriate to analyze with a fuzzy approach as illustrated in figure 3. In this figure, we assume that the data have two true target concepts with centers marked as TC_1 and TC_2 . We display two bags that can belong to either target concept. The first bag, B_1 has five instances $\{a, b, c, d, e\}$ and one of its instances, a , is equally close to TC_1 and TC_2 . This is the same scenario encountered in clustering traditional data. Another scenario, that is unique to MIL data, and that

¹It is not required that all bags have the same number of instances. Here, we assume it is the case only to simplify notation.

requires fuzzy assignment is illustrated with a second bag, $B_2 = \{A, B, C, D, E\}$. In this case, one instance, A , is close to TC_1 while a different instance, B of the same bag is close to TC_2 . In other words, the features that make B_2 similar to one target concept are different from the features that make the same bag similar to a different target concept. The proposed Fuzzy Clustering of Multiple Instance Data (FCMI) algorithm is designed to seek multiple target concepts simultaneously using fuzzy membership assignment of bags to all target concepts to address both of the above scenarios.

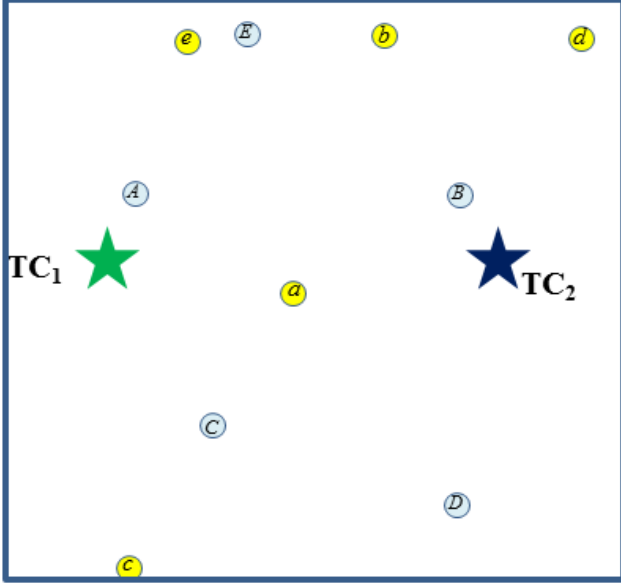


Fig. 3. Two cases that require fuzzy assignment of a bag to multiple target concepts. The first bag, $B_1 = \{a, b, c, d, e\}$ has one instance, a , that is close to both target concepts TC_1 and TC_2 . The second bag $B_2 = \{A, B, C, D, E\}$ has one instance, A , that is close to TC_1 and another instance, B , that is close to TC_2 .

The objective of the FCMI algorithm is to identify K target concepts $\mathcal{T} = \{t_1, \dots, t_k, \dots, t_K\}$, that describe regions in the feature space that include as many positive instances as possible and as few negative instances as possible². Using a fuzzy approach, we assume that each bag, B_n , belongs to each target concept t_k with a membership u_{kn} such that:

$$u_{kn} \in [0, 1], \quad \text{and} \quad \sum_{k=1}^K u_{kn} = 1. \quad (1)$$

Let $\mathbf{U} = [u_{kn}]$ for $k = 1, \dots, T$ and $n = 1, \dots, N$. We define the fuzzy Multi-target concept Diverse Density (MDD) metric as

$$MDD(\mathbf{T}, \mathbf{U}) = \prod_{n=1}^N \prod_{k=1}^K (Pr(t_k|B_n))^{u_{kn}^m}. \quad (2)$$

In (2), m is a fuzzifier that controls the fuzziness of the partition as in the FCM [14]. The MDD in (2) is maximized when the T target concepts correspond to points in the instances feature space such that each target is close to as many instances from positive bags as possible and far from as many instances from negative bags as possible (refer to (10) for the

definition of $Pr(t_k|B_n)$). The proposed FCMI algorithm seeks the optimal $(\mathbf{T}, \mathbf{U})$ that maximize the MDD in (2).

Instead of maximizing (2), we minimize its negative log-likelihood:

$$\begin{aligned} J(\mathbf{T}, \mathbf{U}) &= -\log(MDD(\mathbf{T}, \mathbf{U})) \\ &= \sum_{n=1}^N \sum_{k=1}^K u_{kn}^m \{-\log(Pr(t_k|B_n))\} \end{aligned} \quad (3)$$

subject to the membership constraints in (1).

To minimize (3) with respect to $\mathbf{U}$, we apply Lagrange multipliers and obtain

$$\begin{aligned} J(\mathbf{T}, \mathbf{U}, \boldsymbol{\Lambda}) &= \sum_{n=1}^N \sum_{k=1}^K u_{kn}^m \{-\log(Pr(t_k|B_n))\} \\ &\quad - \sum_{n=1}^N \lambda_n \left(\sum_{k=1}^K u_{kn} - 1 \right) \end{aligned} \quad (4)$$

Assuming that the partial densities $(Pr(t_k|B_n))$, $n=1, \dots, N$ and the columns of $\mathbf{U}$ are independent of each other, we can reduce (4) to the following N independent minimization problems:

$$\begin{aligned} J_n(\mathbf{T}, \mathbf{U}_n, \lambda_n) &= \sum_{k=1}^K u_{kn}^m \{-\log(Pr(t_k|B_n))\} \\ &\quad - \lambda_n \left(\sum_{k=1}^K u_{kn} - 1 \right), n = 1, \dots, N. \end{aligned} \quad (5)$$

Next, we fix $\mathbf{T}$ and set the gradient of J_n to zero, we obtain

$$\frac{\partial J}{\partial u_{qn}} = m u_{qn}^{m-1} \log(Pr(t_q|B_n)) - \lambda = 0 \quad (6)$$

and

$$\frac{\partial J_n}{\partial \lambda_n} = \sum_{k=1}^K u_{kn} - 1 = 0 \quad (7)$$

Solving (6) for u leads to:

$$u_{qn} = \left[\frac{\lambda}{m} \times \frac{1}{-\log(Pr(t_q|B_n))} \right]^{\frac{1}{m-1}} \quad (8)$$

Substituting (8) back into (7), we obtain

$$u_{qn} = \frac{-\log(Pr(t_q|B_n))^{1/(1-m)}}{\sum_{k=1}^K -\log(Pr(t_k|B_n))^{1/(1-m)}} \quad (9)$$

To optimize J_n with respect to the target concepts $\mathbf{T}$, we first need to define the probability of a bag of instances. Recall that a bag is positive if and only if at least one of its instances is positive and is negative if and only if all of its instances are negative. In this paper, we use the NOISY-OR model [9, [21]:

$$Pr(t_k|B_n) = \begin{cases} 1 - \prod_{i=1}^I (1 - Pr(b_{ni} \in t_k)) & \text{if } label(B_n)=1 \\ \prod_{i=1}^I (1 - Pr(b_{ni} \in t_k)) & \text{if } label(B_n)=0 \end{cases} \quad (10)$$

where $label(B_n)=1$ for positive bags (or B_n^+), and $label(B_n)=0$ for negative bags (or B_n^-). In (10), $Pr(b_{ni} \in t_k)$

²Recall that we only know if a bag is positive or negative. Labels at the instance level are not available.

can be regarded as the similarity of instance b_{ni} to target concept t_k . Assuming that each t_k is characterized by a representative feature vector (e.g. centroid), c_k , we let

$$Pr(b_{ni} \in t_k) = e^{-\left(\sum_{j=1}^J s_{kj}(b_{nij} - c_{kj})^2\right)} \quad (11)$$

In (11), s_k is a scaling parameter that weights the role individual features play in defining the overall similarity [9]. Using (11), finding the optimal target concepts reduces to finding their optimal centers c_k and scales s_k for $k = 1, \dots, K$. Thus, we need to solve

$$\frac{\partial J}{\partial c_k} = -\sum_{n=1}^N \frac{u_{kn}^m}{Pr(t_k|B_n)} \times \frac{\partial Pr(t_k|B_n)}{\partial c_k} = 0, \quad (12)$$

and

$$\frac{\partial J}{\partial s_k} = -\sum_{n=1}^N \frac{u_{kn}^m}{Pr(t_k|B_n)} \times \frac{\partial Pr(t_k|B_n)}{\partial s_k} = 0. \quad (13)$$

Since the definition of $Pr(t_k|B_n)$ depends on whether B_n is positive or neagive bag, we rewrite (12) and (13) as

$$\begin{aligned} \frac{\partial J}{\partial c_k} = & -\sum_{n=1}^{N^+} \frac{u_{kn}^m}{Pr(t_k|B_n^+)} \times \frac{\partial Pr(t_k|B_n^+)}{\partial c_k} \\ & -\sum_{n=1}^{N^-} \frac{u_{kn}^m}{Pr(t_k|B_n^-)} \times \frac{\partial Pr(t_k|B_n^-)}{\partial c_k} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \frac{\partial J}{\partial s_k} = & -\sum_{n=1}^{N^+} \frac{u_{kn}^m}{Pr(t_k|B_n^+)} \times \frac{\partial Pr(t_k|B_n^+)}{\partial s_k} \\ & -\sum_{n=1}^{N^-} \frac{u_{kn}^m}{Pr(t_k|B_n^-)} \times \frac{\partial Pr(t_k|B_n^-)}{\partial s_k} \end{aligned} \quad (15)$$

Using (10), it can be shown that

$$\begin{aligned} \frac{\partial Pr(t_k|B_n^+)}{\partial c_k} = & \left\{ \sum_{i=1}^I \frac{1}{1 - Pr(b_{ni}^+ \in t_k)} \times \frac{\partial Pr(b_{ni}^+ \in t_k)}{\partial c_k} \right\} \\ & \times \prod_{i=1}^I (1 - Pr(b_{ni}^+ \in t_k)) \end{aligned} \quad (16)$$

and

$$\begin{aligned} \frac{\partial Pr(t_k|B_n^+)}{\partial s_k} = & -\left\{ \sum_{i=1}^I \frac{1}{1 - Pr(b_{ni}^+ \in t_k)} \times \frac{\partial Pr(b_{ni}^+ \in t_k)}{\partial s_k} \right\} \\ & \times \prod_{i=1}^I (1 - Pr(b_{ni}^+ \in t_k)) \end{aligned} \quad (17)$$

Similar equations can be derived for $\frac{\partial Pr(t_k|B_n^-)}{\partial s_k}$ by substituting s_k for c_k in (16) and (17). Using (11), the partial probabilities in (16) and (17) (and the equivalent equations for the scale) can be computed using

$$\frac{\partial Pr(B_{ni} \in t_k)}{\partial c_{kf}} = 2(B_{nif} - c_{kf}) s_{kf}^2 e^{-\sum_{j=1}^F s_{kj}(b_{nij} - c_{kj})^2} \quad (18)$$

and

$$\frac{\partial Pr(B_{ni} \in t_k)}{\partial s_{kf}} = 2s_{kf}(B_{nif} - c_{kf})^2 e^{-\sum_{j=1}^F s_{kj}(b_{nij} - c_{kj})^2} \quad (19)$$

Equations (12) and (13) have no closed-form solution. Instead, we use approximate solutions based on an iterative line search approach as in [9]. The resulting FCMI algorithm is outlined below.

Algorithm 1 The FCMI Algorithm

Inputs: $\mathcal{B}^+$ and $\mathcal{B}^-$: the sets of + and - bags.

K : the number of target concepts.

Outputs: $\mathbf{C}$: Centers of the K target concepts.

$\mathbf{S}$: Scales of the K target concepts.

$\mathbf{U}$: Membership of all bags in all target concepts.

Initialize $\mathbf{c}^k$ and $\mathbf{s}^k$ for $k = 1, \dots, K$

repeat

 Update $\mathbf{u}_{kn}$ using (9).

 Update $\mathbf{C}$ and $\mathbf{S}$ by performing few iterations of a line search algorithm that minimizes (12) and (13).

until centers do not change significantly or number of iterations is exceeded

return $\mathbf{C}, \mathbf{S}, \mathbf{U}$

IV. EXPERIMENTAL RESULTS

The proposed FCMI algorithm was applied to analyze data of buried landmines collected using a Ground Penetrating Radar (GPR) sensor. The data was collected using a NIITEK vehicle-mounted GPR system [22] from outdoor test lanes at two different locations. The first location was a temperate region with significant rainfall, whereas the second collection was a desert region. The lanes in both locations are simulated roads with known mine locations. All mines are Anti-Tank (AT) mines that can be classified into 2 categories: anti-tank metal (ATM) and anti-tank with low metal content (ATLM). All mines are buried from 0" to 8" under the surface. Multiple data collections were performed at each site at different dates resulting in a large and diverse collection of mine and false alarm signatures. False alarms arise as a result of radar signals that present a mine-like character. Such signals are generally said to be a result of clutter. Each sample, or "alarm," in the dataset has a corresponding datacube with dimensions representing the depth (500 depth bins), down-track (15 frames or scans), and cross-track (15 channels). Using the ground truth, each sample is labeled as mine or clutter. The true depth location is unknown. For our experiment, we use a subset of the data that has 400 mine samples and 400 clutter samples.

Each alarm is divided into 15 overlapping windows along the depth. From each window (50 depths x 15 scans x 15 channels) we extract Edge Histogram Descriptors (EHD) [4]. We extract a 35-dim EHD feature from the (depth, down-track) dimensions at the central channel and another 35-dim EHD feature from the (depth, cross-track) dimensions at the central scan. The 2 EHDs are concatenated to form a 70-dim feature vector. To fit this data into the MIL framework, each alarm is represented by a bag of 15 instances where each instance is represented by a 70-dim feature vector. Each bag is labeled

as positive (mine) or negative (clutter). Labels at the instance level are not available. We only know that a positive bag has one or more instances that exhibit the signature of a mine.

The proposed FCMI algorithm assumes that the number of target concepts is given. In our experiment, we assume that $K=3$. We initialize the centers of the target concepts using the following heuristic. First, using all instances from all positive bags, we select several candidate centers that cover most of the instance feature space. These candidates correspond to instances that are as distant from each other as possible. Next, for each candidate, we identify its 50 nearest neighbors using all instances (positive and negative). Out of all candidates, we select the 3 instances that have the largest ratio of instances from positive bags to instances from negative bags. The scales of each target concept are initialized as the inverse of the standard deviation of all instances identified as its nearest neighbors.

In figure 4, we display a scatter plot of the features of all instances of all bags. Here, for the purpose of visualization, we project the 70-dim data to its 2 principal components. In this figure, we also display the initial centers of the 3 target concepts and the final centers after convergence. We also show the path of each center as the FCMI iterates. For this data, we fix the fuzzifier m to 1.5 and for each iteration, we run the line search (to update the centers and scales) for 5 iterations. First, we note that some features from instances of positive bags are clustered away (top left of figure) from negative instances (on the right side). Typically, these correspond to instances extracted from the “correct” depth. Other instances, on the other hand, are located around instances from negative bags. These correspond to instances extracted from the background part of the positive bags. Second, we note that the 3 centers converged to dense red regions (instances from positive bags) and away from dense blue regions (instances from negative bags).

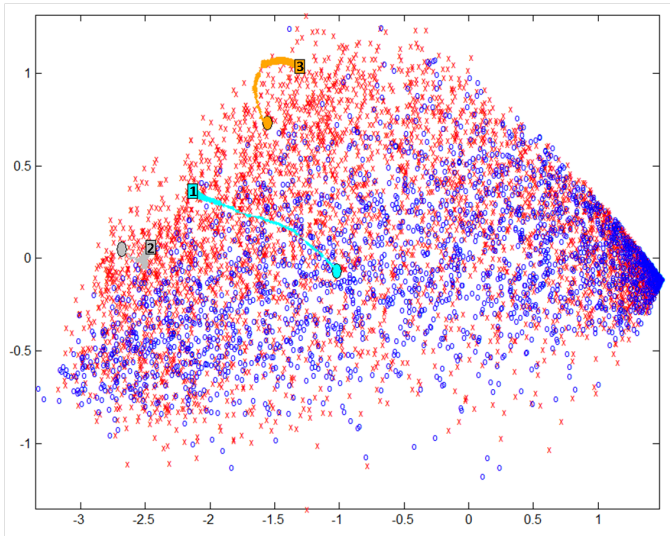


Fig. 4. Scatter plot of the 2 principal components of the instances feature space. Instances of positive bags are displayed as red ‘x’ and instances of negative bags are displayed as blue ‘o’. The location of initial (final) centers is shown by circles (squares).

Recall that labels at the instance level are not available and that positive bags include at least one positive instance.

After running FCMI, we use the following simple steps to identify positive instances within positive bags. Assume that bag B_n is assigned to target concept t_k (i.e. $u_{kn} = \max u_{in}$ for $i = 1, \dots, K$). The likelihood of each instance, b_{ni} , of bag B_n in target concept t_k can be computed using (11). The most likely positive instance is the one that has the largest likelihood (multiple positive instances could be identified using a threshold). To verify that FCMI was able to identify the relevant instances within positive bags, in figure 5, we display few mine alarms where we highlight the window of the most likely instance. As it can be seen, this window corresponds to the strongest part of the mine signature.

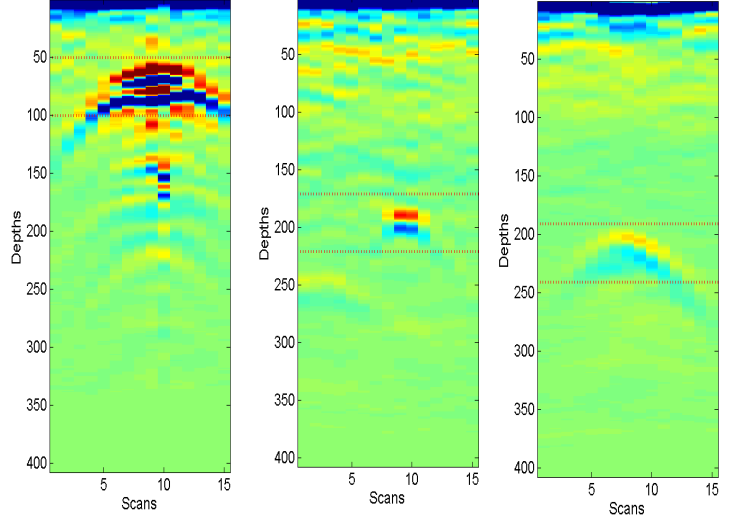
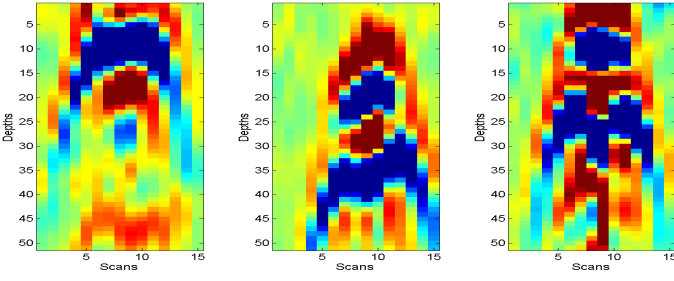


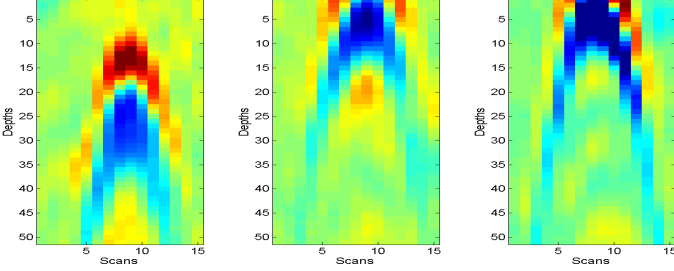
Fig. 5. GPR signatures of three different alarms. Each alarm is represented by a bag of 15 instances extracted at different depths. The most likely instance of each bag is highlighted

To illustrate the need to identify multiple target concepts, in figure 6, we display samples from the 3 target concepts identified by FCMI. For each target concepts, we display 3 typical instances. As it can be seen, target concept 1 corresponds to mines with large and strong energy. Most of the mines assigned to this concept are large and buried no more than 3” deep. Target concept 2 corresponds to large mines with weak energy. These are typically large mines buried deeper than 3”. Target concept 3 corresponds to mines with narrower signatures. These are typically mines of smaller sizes.

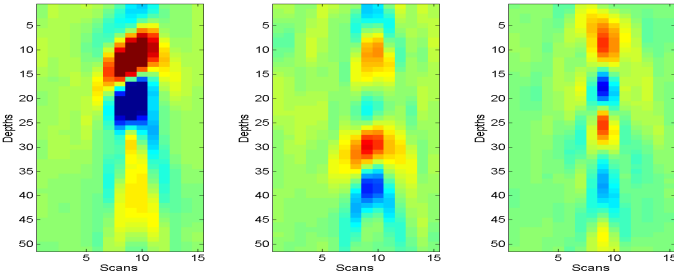
The proposed fuzzy approach has several advantages. First, at the bag level, each bag B_n belongs to each target concept t_k with a fuzzy membership degree u_{kn} . This is the standard advantage that fuzzy clustering methods have over crisp clustering. Second, and more importantly, fuzzy memberships can provide more detailed information at the instance level. Specifically, a bag B_n can have a relatively high membership in concept k_1 , u_{k_1n} , because one of its instances is close to concept k_1 . Similarly, the same bag can have a non-zero membership in another concept k_2 , u_{k_2n} , either because the same instance is close to concept k_2 or because a different instance is close to concept k_2 . Distinction between these two cases can provide a richer description of the data. In figure 7, we illustrate these two scenarios. In figure 7(a), the bag has 0.68 membership in concept 2 (large mines with weak energy, refer to figure 6(b)) and 0.31 membership in concept 3 (small



(a) Sample instances from bags assigned to target concept 1



(b) Sample instances from bags assigned to target concept 2



(c) Sample instances from bags assigned to target concept 3

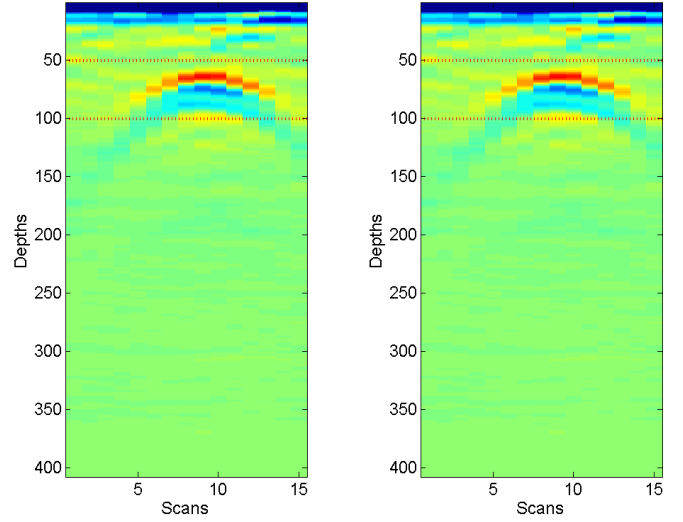
Fig. 6. Representative instances from the 3 target concepts identified by FCMI.

mines with narrow signature, refer to figure 6(c)). In this case, the same instance has the highest likelihood in both concepts. In figure 7(b), the bag has 0.57 membership in concept 1 (large mines with strong energy, refer to figure 6(a)), 0.15 membership in concept 2, and 0.28 membership in concept 3. In this case, different instances have the highest likelihood in the 3 concepts. For instance, the shallower instance is similar to concept 1 while the deeper instance is similar to concept 3.

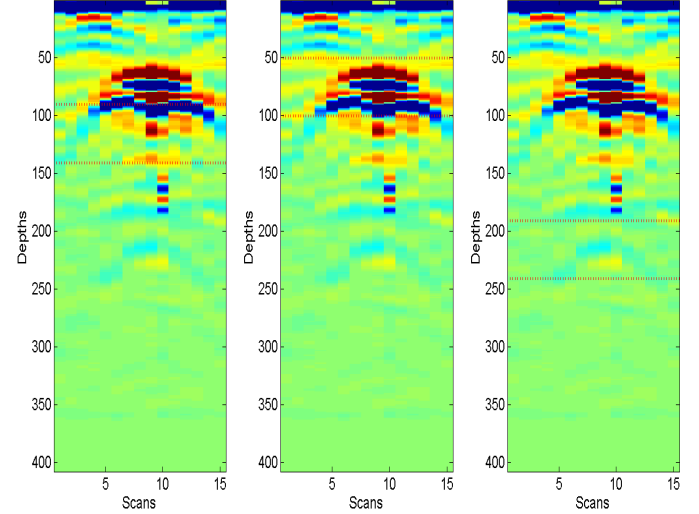
V. CONCLUSIONS

We proposed an algorithm to identify multiple target concepts for multiple instance data. First, we defined a fuzzy Multi-target concept Diverse Density (MDD) metric. Then, we derived the necessary conditions to optimize this MDD and developed the Fuzzy Clustering of Multiple Instance data algorithm. The FCMI algorithm identifies K target concepts simultaneously. Each target concept correspond to a dense region in the instances feature space with maximal correlation to instances from positive samples and minimal correlation to instances from negative samples.

Using data of buried landmines collected with a ground penetrating radar sensor, we showed that the proposed FCMI



(a) an alarm that has 0.68 fuzzy membership in concept 2 (left) and 0.31 membership in concept 3 (right). The instances with the highest likelihood are extracted from the same depth location.



(b) an alarm that has 0.57 fuzzy membership in concept 1, 0.15 membership in concept 2, and 0.28 membership in concept 3. The instances with the highest likelihood and contributing to the above memberships are extracted from different depth locations.

Fig. 7. Sample alarms with high fuzzy memberships in more than one concept.

algorithm can identify distinct target concepts that correspond to mines of different types buried at different depths. Different instances within a bag can be similar to only one target concept or to multiple target concepts. Thus, the FCMI can be used to provide rich description of the data at the instance level.

In this paper, we provided only qualitative evaluation of the proposed FCMI. Quantitative evaluation would require building an additional layer that performs classification. For instance, the identified target concepts could be used to map the multiple instance data as in [12]. We are investigating

this research direction. Another research direction that we are currently investigating is the development of a possibilistic [23] version of FCMI. Since multiple instance data can have a large number of negative bags and even positive bags can have a large number of irrelevant instances, a possibilistic version of FCMI can make it more robust.

ACKNOWLEDGMENT

This work was supported in part by U.S. Army Research Office Grants Number W911NF-13-1-0066 and W911NF-14-1-0589. 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 Army Research Office, or the U.S. Government.

REFERENCES

- [1] T. Dietterich, R. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, pp. 31–71, 1997.
- [2] J. Bolton, P. Gader, H. Frigui, and P. Torrione, "Random set framework for multiple instance learning," *Information Sciences*, vol. 181, no. 11, pp. 2061–2070, 2011.
- [3] P. D. Gader, R. Grandhi, W.-H. Lee, J. N. Wilson, and D. K. C. Ho, "Feature analysis for the nittek ground-penetrating radar using order-weighted averaging operators for landmine detection," pp. 953–962, 2004.
- [4] H. Frigui and Paul, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors," *IEEE Transactions on Fu*, vol. 17, no. 9, pp. 185–199, 2009.
- [5] Z. Xia, J. Peng, X. Feng, and J. Fan, "Multiple instance learning for automatic image annotation," in *Advances in Multimedia Modeling*, ser. Lecture Notes in Computer Science, S. Li, A. El Saddik, M. Wang, T. Mei, N. Sebe, S. Yan, R. Hong, and C. Gurrin, Eds. Springer Berlin Heidelberg, 2013, vol. 7733, pp. 194–205.
- [6] H. Wei and W. Yu, "Text representation and classification based on multi-instance learning," in *Management Science and Engineering, 2009. ICMSE 2009. International Conference on*, Sept 2009, pp. 34–39.
- [7] P. Wohlhart, M. Kostinger, P. Roth, and H. Bischof, "Multiple instance boosting for face recognition in videos," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, R. Mester and M. Felsberg, Eds. Springer Berlin Heidelberg, 2011, vol. 6835, pp. 132–141.
- [8] J. Amores, "Multiple instance classification: Review, taxonomy, and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 2013.
- [9] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," *Advances in Neural Information Processing Systems*, vol. 10, no. 1, pp. 570–576, 1998.
- [10] Q. Zhang and S. A. Goldman, "Em-dd: An improved multiple-instance learning technique," in *In Advances in Neural Information Processing Systems*. MIT Press, 2001, pp. 1073–1080.
- [11] Y. Chen, J. Z. Wang, and D. Geman, "Image categorization by learning and reasoning with regions," *Journal of Machine Learning Research*, vol. 5, pp. 913–939, 2004.
- [12] Y. Chen, J. Bi, and J. Z. Wang, "Miles: Multiple-instance learning via embedded instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 1931–1947, 2006.
- [13] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-i.i.d. samples," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. ICML '09. New York, NY, USA: ACM, 2009, pp. 1249–1256.
- [14] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.
- [15] G. M. Crippen and T. F. Havel, *Distance geometry and molecular conformation*. Research Studies Press Taunton, England, 1988, vol. 74.
- [16] J. D. Keeler, D. E. Rumelhart, and W.-K. Leow, "Integrated segmentation and recognition of hand-printed numerals," in *Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems 3*, ser. NIPS-3. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, pp. 557–563.
- [17] J. MacQueen, "Some methods for classification and analysis of multivariate observations," Berkeley, Calif., pp. 281–297, 1967.
- [18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. pp. 1–38, 1977.
- [19] R. Krishnapuram, H. Frigui, and O. Nasraoui, "Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation. i," *Fuzzy Systems, IEEE Transactions on*, vol. 3, no. 1, pp. 29–43, Feb 1995.
- [20] R. Kruse, C. Döring, and M.-J. Lesot, "Fundamentals of fuzzy clustering," *Advances in fuzzy clustering and its applications*, pp. 3–30, 2007.
- [21] S. Srinivas, "A generalization of the noisy-or model," *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 208–218, 1993.
- [22] K. J. Hintz, "Snr improvements in nittek ground penetrating radar," in *Proceedings of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets IX, Orlando, FL, USA, April, 2004*.
- [23] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *Fuzzy Systems, IEEE Transactions on*, vol. 1, no. 2, pp. 98–110, May 1993.