Fuzzy and Possibilistic Clustering for Multiple Instance Linear Regression

Mohamed Trabelsi
Multimedia Research Lab, CECS Dept.
University of Louisville

Hichem Frigui
Multimedia Research Lab, CECS Dept.
University of Louisville

***Abstract*—Multiple instance regression (MIR) operates on a collection of bags, where each bag contains multiple instances sharing an identical real-valued label. Only few instances, called primary instances, contribute to the bag's label. The remaining instances are noise and outliers observations. The goal in MIR is to identify the primary instances within each bag and learn a regression model that can predict the label of a previously unseen bag. In this paper, we present an algorithm that uses robust fuzzy clustering with an appropriate distance to learn multiple linear models from a noisy feature space simultaneously. We show that fuzzy memberships are useful in allowing instances to belong to multiple models, while possibilistic memberships allow identification of the primary instances of each bag with respect to each model. We also use the possibilistic memberships to identify the optimal number of regression models. We evaluate our approach on a series of synthetic data sets. We show that our approach achieves higher accuracy than existing methods.**

***Index Terms*—Multiple instance regression, Fuzzy clustering, Possibilistic clustering, Multiple model regression.**

I. Introduction

Unlike classical supervised learning where every object is represented by a single feature vector and a label, in multiple instance learning (MIL) [1], [2], an object contains a set of instances, called a bag, with a single label. Labels are available only at the bag level and labels of the individual instances are unknown. The label of each bag cannot be simply propagated to all of its instances as a significant number of instances can be irrelevant to the object they describe. This many-to-one relationship between instances and data labels produces an inherent ambiguity in determining which instances in a given bag are responsible for its associated label. MIL was formalized in 1997 by Dietterich et al. providing a solution to drug activity prediction [1]. Ever since, MIL has increasingly been applied to a wide variety of tasks including drug discovery [3], image analysis [4], [5], [6], content-based information retrieval [7], time series prediction [2], landmine detection [8], information fusion [9], [10], and remote sensing [11].

As in classical learning, data labels in MIL can be categorical or real-valued. Categorical labeling leads to Multiple Instance Classification (MIC), while real-valued labeling leads to Multiple Instance Regression (MIR). Most of the existing work on MIL has focused on MIC. There are different assumptions in MIC. The most common one assumes that a bag is classified negative if all of its instances are negative and positive if at least one of its instances is positive. Many algorithms have been proposed to solve MIC problem in the past two decades. Examples include APR [1], MILES [5], MILIS [12], SVM [13], Diverse Density [3], MI kernels [14], EM-DD [15], Citation-kNN [16] and BP-MIP [17], [18].

In Multiple instance Regression (MIR), bags have real-valued labels and the objective is to learn a regression model that can predict the label of a bag from its content. There is no prior knowledge of the relevant instances within each bag. MIR has received much less attention than MIC and only recently few algorithms have been proposed.

II. Related work

The two simplest approaches to MIR, that are commonly used as baselines (e.g., in [19], [20], [11]), are the Instance-MIR and Aggregated-MIR. In Instance-MIR, every instance within a bag is treated as a relevant examplar and inherits the bag's label. Then, a model is trained using traditional supervised regression techniques. To predict the real-valued label of an unseen bag, the learned regression model is applied to each one of its instances and the final bag label is computed by aggregating (e.g min, max, mean, median) the learned instances' labels. In Aggregated-MIR, the aggregation is applied at the feature level. First, features of all instances within a bag are combined to form a meta-data. Then, a standard regression method is applied to the meta-data. To predict the label of an unseen bag, the aggregation is first applied to form the meta-data and then the learned regression model is used to predict the label of the testing bag based on the meta-data.

Primary instance regression (PIR) [21] is one of the earliest MIR that maintains the bag structure. It assumes that each bag has one true instance, called primary instance, and that the remaining instances are noisy observations. PIR is an iterative algorithm that uses an EM-based approach to alternate between selecting the primary instances and fitting a linear regression to these instances. Like most MIR methods, PIR does not provide a mechanism to predict the label of an unseen bag. Typically, as in the instance-MIR, the predicted value of a test bag is an aggregation of its instances' output.

Motivated by the fact that bags contain items drawn from different distributions, MI-ClusterRegress [19] uses a clustering step to partition the data into a predefined number of clusters. Instances that are relevant to each cluster, called exemplars, are identified and used to learn a regression model

978-1-5090-6020-7/18/$31.00 ©2018 IEEE

for each cluster using traditional simple instance regression techniques. The cluster with the best fitting error is identified as the "prime" cluster and its model is assumed to have generated the bags' labels. The label of a previously unseen bag is an aggregation of the labels of its instances that are assigned to the prime cluster.

In this paper, we introduce a novel MIR framework, called Robust Fuzzy Clustering for Multiple instance Linear Regression (RFC-MILR). We show that, for multiple instance data, regression models can be identified as clusters when appropriate features and distances are used. RFC-MILR uses two types of membership functions for each instance. The first one is a fuzzy membership and is needed to identify the primary instances within each bag. The second membership is a possibilistic function and is needed to identify non-primary instances as noise and outliers to reduce their influence on the learned regression parameters.

The organization of the rest of the paper is as follows. Section III proposes a robust clustering to learn multiple linear regression models for MIR; Section IV illustrates the steps of the proposed algorithm using an illustrative example and compares previous and proposed MIR algorithms using synthetic datasets.

III. Robust clustering to learn multiple Linear regression models

Let $D = \{B_j, j = 1 \ldots N_B\}$ be a collection of N_B bags, where $B_j = \{(b_{ij}, y_j), i = 1 \ldots n_j\}$, $b_{ij} \in \mathbb{R}^d$ is the attribute vector representing the ith instance from the jth bag, y_j is the real-valued target value of the jth bag and n_j is the number of instances in the j^{th} bag. The instances b_{ij} that determine the label y_j, called primary instances, are unknown. The objective of MIR is to identify the primary instances in every bag, learn the regression model, and be able to predict the label of previously unseen bags.

In the following, we propose a new approach, called Robust Fuzzy Clustering for Multiple Instance Linear Regression (RFC-MILR). RFC-MILR performs clustering and multiple linear model fitting simultaneously. RFC-MILR has four main properties. First, instead of using clustering to partition the instances in the feature space regardless of the labels of their bags as in MI-ClusterRegress [19], we combine features and labels and use clustering, with an appropriate distance, to identify multiple local linear regression models. Second, we use a robust clustering approach so that non-primary instances (that incorrectly inherit the label of the bag they belong to) can be treated as noise and outliers to minimize their influence on the learned regression parameters. Third, we use fuzzy clustering so that each instance can contribute to each local regression with a fuzzy membership degree. Finally, we use properties of the possibilistic memberships to find the optimal number of regression models.

Let $x_{ji} = [b_{ji}, y_i] \in \mathbb{R}^{d+1}$ represent the concatenation of the j^{th} instance from the i^{th} bag and the label of its bag. Recall that labels are not available at the instance level and that y_i is valid only for the primary instances of bag i. Thus, many of the x_{ji}'s can have an irrelevant y_i. We combine x_{ji} from all training bags into $D = \{x_{ji}, i = 1 \ldots N_B, j = 1 \ldots n_i\}$. To simplify notation, we assume that all bags have the same number of instances $n_i = n$ for $i = 1 \ldots N_B$, and we rewrite $D = \{x_i, i = 1 \ldots N\}$, where $N = n \times N_B$. Next, we show how clustering could be used to identify the primary instances from all of the N instances and learn the MIR model simultaneously.

A. Robust Clustering for MIR

The fuzzy c-means (FCM) [22] algorithm minimizes

$$J_F = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij}^F)^m dist_{ij}^2 \tag{1}$$

subject to the constraint:

$$u_{ij}^F \in [0,1] \text{ for all } i,\, j; \text{ and } \sum_{i=1}^{C} u_{ij}^F = 1 \text{ for all } j\ . \tag{2}$$

In (1), C is the number of clusters, $dist_{ij}$ is the distance from x_j to cluster i, $m > 1$ is a weighting exponent called the fuzzifier, and u_{ij}^F is the fuzzy membership of x_j in cluster i.

The distance $dist_{ij}$ used in (1) controls the type and shape of clusters that will be identified. Various distances have been proposed in the literature to identify ellipsoidal, linear, and shell clusters such as lines, circles, ellipses, and general quadratics [23], [24]. In this paper, we assume that the underlying regression model is linear and we use (1) to identify multiple linear models. In particular, we use a generalization of the distance in [22], [25] and let:

$$dist_{ij}^2 = \sum_{k=1}^{d+1} v_{ik} ((x_j - c_i) \cdot e_{ik})^2, \tag{3}$$

where c_i is the center of cluster i, e_{ik} is the k^{th} unit eigenvector of the covariance matrix Σ_i of cluster i. The eigenvectors are assumed to be arranged in ascending order of the corresponding eigenvalues λ_{ik}. In (3), v_{ik} is a parameter that controls the contribution of the distance along each eigenvector to the total distance. In this paper, we let

$$v_{ik} = \frac{\left[\prod_{j=1}^{d+1} \lambda_{ij} \right]^{\frac{1}{d+1}}}{\lambda_{ik}}, \tag{4}$$

that is, more importance will be given to distances projected on the eigenvectors associated with the smaller eigenvalues.

Optimization of (1) with $dist_{ij}$ in (3) subject to (2), using alternate optimization, results in an iterative algorithm that alternates between updating the fuzzy memberships using:

$$u_{ij}^F = \left[\sum_{k=1}^{C} \left(\frac{dist_{ij}^2}{dist_{kj}^2} \right)^{\frac{1}{m-1}} \right]^{-1} \tag{5}$$

and the center c_i and covariance Σ_i of cluster i using

$$c_i = \frac{\sum_{j=1}^{N} (u_{ij}^F)^m x_j}{\sum_{j=1}^{N} (u_{ij}^F)^m}, \quad (6)$$

and

$$\Sigma_i = \frac{\sum_{j=1}^{N} (u_{ij}^F)^m (x_j - c_i)(x_j - c_i)^T}{\sum_{j=1}^{N} (u_{ij}^F)^m}. \quad (7)$$

The objective function of the FCM in (1) is known to be sensitive to noise and outliers, and thus, is not suitable for the considered MIR application where we know a priori that the data is very noisy as non-primary instances and their labels will act like noise. Instead, we use the possibilistic c means (PCM) [26], which relaxes the constraint in (2) and minimizes

$$J_P = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij}^P)^m dist_{ij}^2 + \sum_{i=1}^{C} \eta_i \sum_{j=1}^{N} (1 - u_{ij}^P)^m \quad (8)$$

In (8), $u_{ij}^P \in [0, 1]$ is a possibilistic membership degree that is not constrained to sum to 1 across all clusters. It is close to 0 for samples that are considered outliers, and close to 1 for inliers.

Optimization of (8) also results in an iterative algorithm that alternates between updating u_{ij}^P using

$$u_{ij}^P = \left[1 + \left(\frac{dist_{ij}^2}{\eta_i}\right)^{\frac{1}{m-1}} \right]^{-1} \quad (9)$$

and the center c_i and covariance Σ_i as in (6) and (7) respectively. In (9), η_i is a cluster resolution parameter that could be fixed a priori or updated in each iteration using the distribution of the data within each cluster [26].

Since the PCM does not constraint the memberships u_{ij}^P to sum to 1, it can result in several similar or even identical clusters. We use this feature to identify the optimal number of regression models [27]. We simply start with an over specified number of clusters, then identify and merge similar ones. Two clusters i and k are considered similar if the possibilistic memberships of all points in the two clusters are similar. These clusters can be easily identified and merged if

$$\frac{\sum_{k=1}^{N} |u_{ik}^P - u_{jk}^P|}{\sum_{k=1}^{N} |u_{ik}^P| + \sum_{k=1}^{N} |u_{jk}^P|} < \theta_M \quad (10)$$

In (10), θ_M is a threshold constant.
Currently, we assume that the underlying regression model is linear and thus, it can be captured by one linear cluster. Consequently, if the algorithm identifies more than one cluster, say $c' > 1$, we need to select the "optimal" cluster, p. We simply select the cluster that minimizes the fitting errors, i.e,

$$p = arg \min_{i=1,\dots c'} \left\{ \varepsilon_i = \sum_{j=1}^{N} (u_{ij}^P)^m dist_{ij}^2 \right\} \quad (11)$$

The primary instances of cluster p, denoted $\mathcal{P}^p$, are defined as the set of inliers to this cluster, that is,

$$\mathcal{P}^p = \{x_j, j = 1 \dots N \mid u_p^P > \theta_P\} \quad (12)$$

In (12), θ_P is a constant typically set to 0.1. The primary instances of cluster p are also considered the primary instances of the entire data D, i.e $\mathcal{P} = \mathcal{P}^p$.

The linear regression model parameters can be identified from the cluster center c_p and covariance matrix Σ_p. Let $e_{min} = [e_{min}^1, \dots, e_{min}^{d+1}]$ be the eigenvector associated with the smallest eigenvalue λ_{min} of Σ_p and let $x = [x_1, \dots, x_d, y] \in \mathcal{P}$ be a primary instance. The fact that x and c_p belong to the same linear regression model leads to

$$e_{min}.(x - c_p) = 0, \quad (13)$$

or

$$e_{min}.x = e_{min}.c_p. \quad (14)$$

Decomposing x into the instance feature vector $[x_1, \dots, x_d]$ and its label y, we obtain

$$e_{min}^{d+1} y + \sum_{k=1}^{d} e_{min}^k x_k = e_{min}.c_p \quad (15)$$

Solving for y in (15), we obtain the regression model:

$$y = f(x) = \frac{e_{min}.c_{opt}}{e_{min}^{d+1}} - \sum_{k=1}^{d} \frac{e_{min}^k}{e_{min}^{d+1}} x_k \quad (16)$$

The resulting RFC-MILR algorithm is summarized in Algorithm 1.

B. Label Prediction

Primary instances in the training data can be identified using (12) as the inliers, i.e points that have high possibilistic membership. For testing, this process is not as applicable since labels are needed to assign new memberships. Instead we use the following approach.

Let $B^t = \{x_1^t \dots x_N^t\}$ be a test bag with N instances. First, for each $x_i^t \in B^t$, we identify the closest primary instance (from training data) $x_i^P \in \mathcal{P}$. Then, we assume that y_i^P, the label of x_i^P, is a good initial estimate of the label of x_i^t and use $[x_i^t, y_i^P]$ to estimate the possibilistic membership u_i^P of x_i^t in the regression model f. The primary instance of test bag B^t is identified as the instance that has the highest possibilistic membership, i.e.

$$x_{prim}^t = \{x_k^t \mid u_k^P = \max_{i=1\dots N} \{u_i^P\}\} \quad (17)$$

Finally, test bag B^t is labeled using

$$\hat{y}(B^t) = f\left(x_{prim}^t\right) \quad (18)$$

Algorithm 1 The RFC-MILR Algorithm

procedure RFC-MILR(D,C,m)
    Inputs:
D: Training data
C: an overestimated number of clusters
m: fuzzifier
    Outputs:
f: learned regression model
$\mathcal{P}$: set of primary instances

    initialize centers
    % start with fuzzy memberships to get initial models
    for few (10) iterations **do**
        update centers using (6)
        update covariance matrices using (7)
        update fuzzy memberships using (5)
    end
    % Continue with possibilistic membership to refine the models by ignoring noise and outliers
    repeat
        update centers using (6)
        update covariance matrices using (7)
        update possibilistic memberships using (9)
    until centers and covariances do not change
    Merge similar clusters using (10)
    if number of remaining cluster $c' > 1$ **then**
        select "optimal" cluster using (11)
    Identify $\mathcal{P}$, the set of primary instances using (12)
    Identify regression model using (16)

We should note here that it is possible to select multiple primary instances for each test bag (e.g all instances with possibilistic membership above a threshold). In this case, the label of B^t can be taken as the average of the labels of all primary instances.

IV. Experimental Results

To validate the proposed MIR and evaluate its performance, we generate a series of synthetic multiple instance data sets with linear models. We vary the dimensionality of the feature space, the number of instances per bag, and the noise level. We compare the results of RFC-MILR with 4 existing MIR algorithms. These are the MI-Cluster Regress [19], the Instance-MIR and Aggregated-MIR [19], [20], and the Primary-MIR [21].

A. Synthetic datasets

First, we generate the instances features, $b_{ij} \in \mathbb{R}^d$, using

$$\mathbf{b}_{ij} = \mathbf{t}_i + \epsilon_{ij}^F, i = 1, \cdots, N_B, \quad and \quad j = 1, \cdots, n_i, \tag{19}$$

where $\mathbf{t}_i$ is the primary instance of bag, B_i, generated from a d-dimensional Gaussian distribution with zero mean and covariance $=10 I^{d\times d}$. In (19), ϵ_{ij}^F is a noise term added to the features. It is generated using a normal distribution $\mathcal{N}^F(\mu^F{=}0, \sum^F{=}\sigma^F I^{d\times d})$. As the noise level increases, $\mathbf{b}_{ij}$ will divert from being a primary examplar to an irrelevant instance.

The label of each bag , B_i, is generated using

$$y_i = h(\mathbf{t}_i) + \epsilon_i^L, \tag{20}$$

where $h()$ is a linear d-dimensional function. We use

$$h(\mathbf{x}) = \sum_{k=1}^{d} a_k x_k \tag{21}$$

where a_k are constant coefficients. In (20), ϵ_i^L is a noise term, added to the true label. It is generated from a normal distribution $\mathcal{N}^L(\mu^L{=}0, \sigma^L)$.

Using the above strategy, we generate multiple data sets by varying:

1) the dimensionality of the feature space, d from 1 to 10.
2) The noise level added to the features in (19). We let
$$\sigma^F = k_1 \times \sigma_0^F, \tag{22}$$
with $\sigma_0^F{=}0.1$ and k_1 varies from 1 to 100.
3) The noise level added to the bags' labels in (20). We let
$$\sigma^L = k_2 \times \sigma_0^L, \tag{23}$$
with $\sigma_0^L{=}0.05$ and k_2 varies from 1 to 25.
4) The number of instances per bag, n_i, from 5 to 100.

For each set of parameters, we create 10 linear models by generating random coefficients a_k (used in (21)) . For each model, we generate one data collection that includes 100 bags, i.e. $N_B{=}100$.

B. Illustrative Example

First, we use a simple 1-Dim data to illustrate the different steps of the proposed MIR approach. The true model is $h(x) = 6x$, and each bag has 5 instances. For the noise levels, we use $k_1{=}1000$ and $k_2{=}2$.

The data is displayed in figure 1(a) where the x-axis represents the 1-D feature of the instances and the y-axis represents the label of each bag (all instances in one bag have the same y value as they share the same label). All primary instances are displayed as filled blue circles and the remaining ones are displayed as red 'x'. Recall that in MIR this information is not available, and that we use it here for illustrative purposes only. Using $C{=}3$, figure 1(b) displays the 3 initial clusters obtained after running the RFC-MILR for few iterations with fuzzy memberships. Points that belong to different clusters are displayed with different symbols and colors. Figure 1(c) displays the results after switching from fuzzy to possibilistic memberships and running the algorithm for 3 iterations. As it can be seen, RFC-MILR started identifying noisy instances (displayed as black circles) and the 3 linear clusters started converging to the same true model. Figure 1(d) displays the final results after the clusters became identical and got merged into one using (10). Points with high possibilistic memberships (> 0.9) are located along the linear model. These points will be considered the primary instances. All others, will be treated as irrelevant ones.

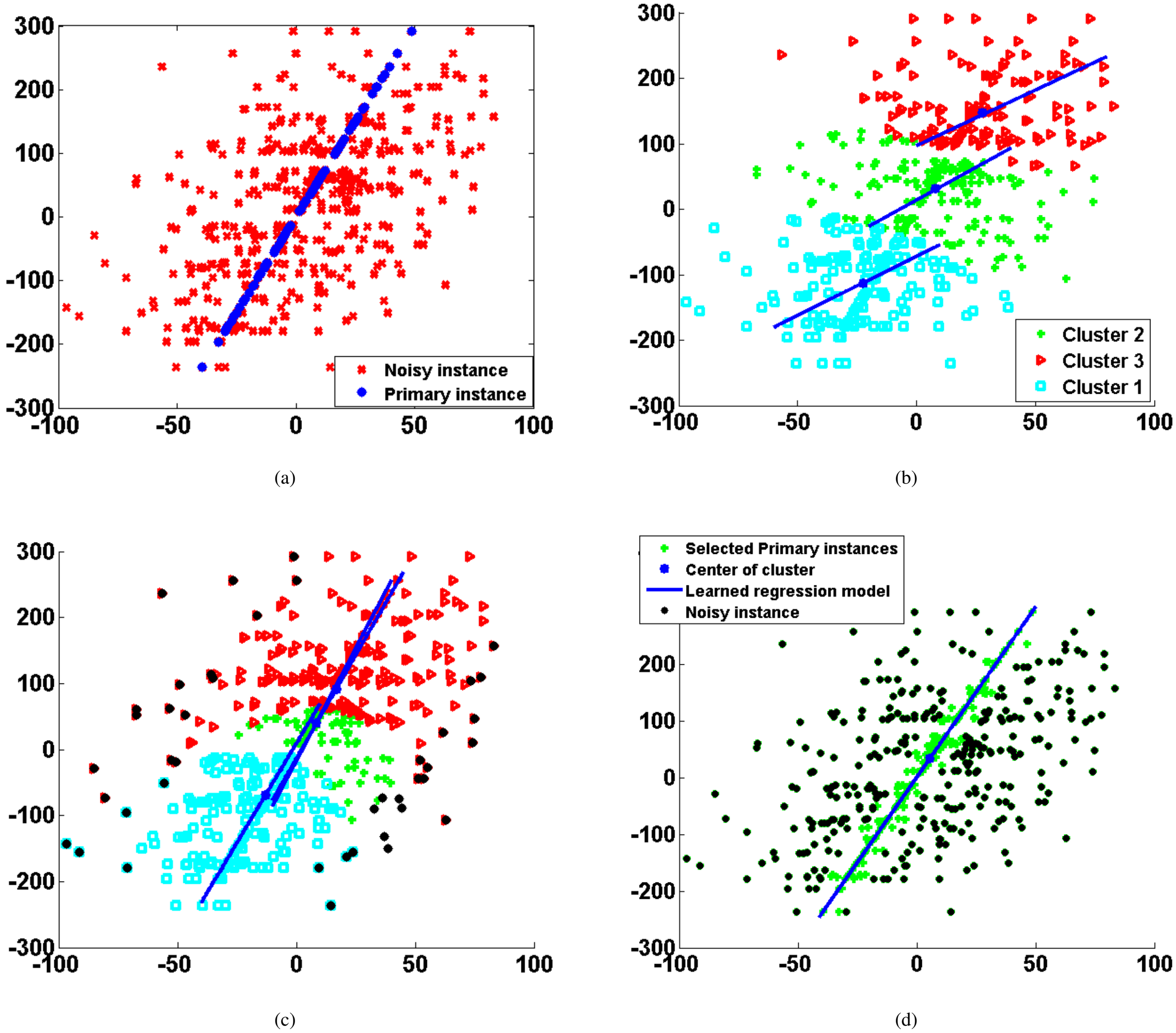

Fig. 1. Illustrations of the steps of RFC-MILR: (a) Example of MIR data, (b) Result of fuzzy clustering, (c) Result of possibilistic clustering after 3 iterations, (d) Result of RFC-MILR after merging similar clusters

C. Results

To compare the performance of the different MIR algorithms, for each data set, we compute the mean square error (MSE) using:

$$MSE = \frac{1}{N_B}\sum_{i=1}^{N_B}(y_i - \hat{y}(B_i))^2, \quad (24)$$

where y_i is the true label of bag i and $\hat{y}(B_i)$ is the label estimated using the different MIR algorithms.

For all data sets, we set the initial number of models C to 10, θ_M in (10) to 0.1, and fuzzifier m to 2. The value of η_i in (9) is estimated using the average fuzzy intra-cluster distance of cluster i as recommended in [26].

In the following experiments, unless stated otherwise, we fix the k_1 value, used to control the level of noise added to the instances (22) to 10. We also fix k_2, used to control the noise added to bags' labels (23) to 10. The number of instances per bag, n_i, and the dimensionality of the instance space, d, to 5 and 1 respectively.

In the first experiment, we vary the noise level added to the bags' labels by increasing k_2 in (23) from 1 to 25. For each value of k_2, we generate 10 data sets using 10 linear models that use random coefficients a_k's (refer to (21)). The results of this experiment are displayed in Figure 2 where for each value of k_2, we display the mean MSE averaged over the 10 random models. We also display the variance of the MSE as a vertical error bar. As it can be seen, RFC-MILR has the lowest error. Moreover, the results of the 10 random

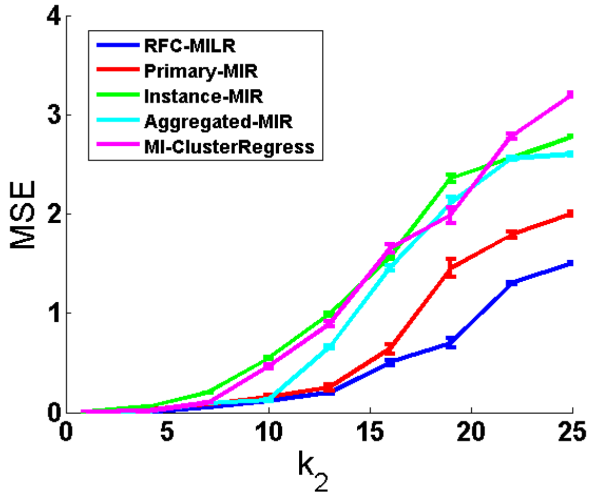

Fig. 2. Comparison of RFC-MILR with previous MIR algorithms when varying the noise level added to the bags' labels in (20)

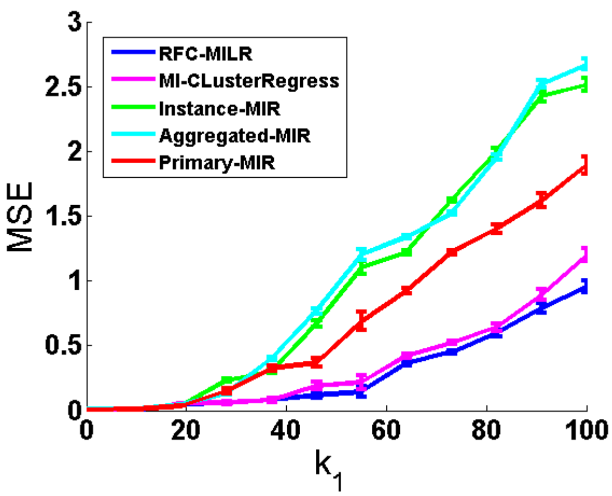

Fig. 3. Comparison of RFC-MILR with previous MIR algorithms when varying the noise level added to the features in (19)

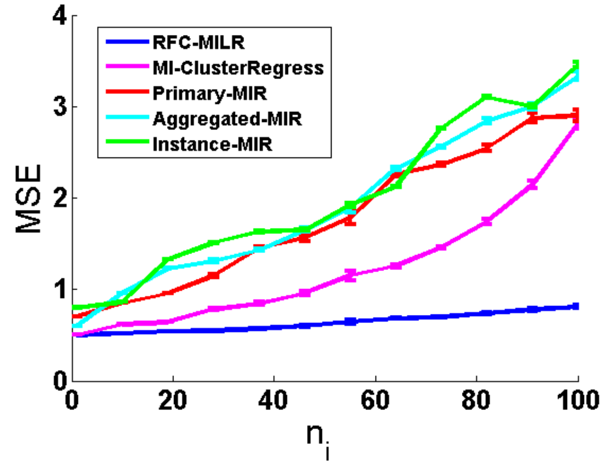

Fig. 4. Comparison of RFC-MILR with previous MIR algorithms when varying the number of instances per bag

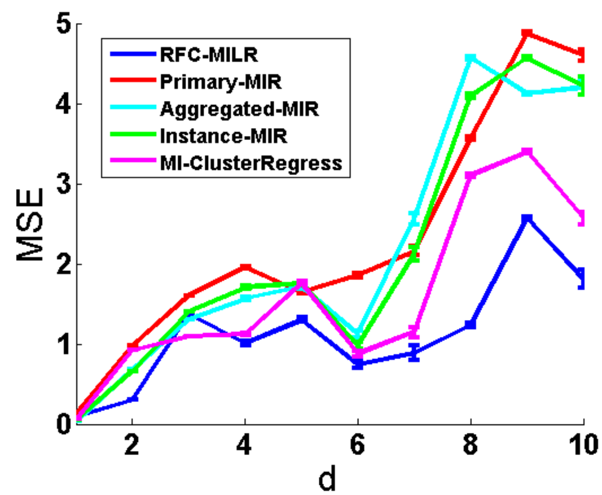

Fig. 5. Comparison of RFC-MILR with previous MIR algorithms when varying the dimensionality of the feature space

models are consistent as indicated by the low MSE variations across the random models. In a second experiment, we vary k_1 from 1 to 100. The results are displayed in figure 3 where the proposed RFC-MILR has the lowest MSE average and variation.

In a third experiment, we vary the number of instances per bag, n_i from 5 to 100. In general, adding more instances increases the number of irrelevant instances and makes the MIR problem more challenging. The results of this experiment are displayed in Figure 4.

As it can be seen, the proposed RFC-MILR algorithm is very robust even in the presence of a large number of irrelevant instances. On the other hand, for all other 4 algorithms the average MSE increases significantly as more irrelevant instances are included in each bag.

In a fourth experiment, we vary the dimensionality of the instances, d, from 1 to 10. The results are displayed in Figure 5.

As for the previous experiments, the proposed RFC-MILR has the lowest MSE values.

V. Conclusions

We proposed a new approach to multiple instance regression based on robust clustering. By combining the bags instances and labels, and using an appropriate distance that measures the deviation of a point from a linear model, we showed that a possibilistic clustering algorithm can be used to estimate the regression model in a MIR setting. More importantly, we showed that the possibilistic memberships can be used to identify the primary instances and the irrelevant instances within each bag. Using several synthetic data sets with known structure and different levels of noise and difficulty, we showed that our approach achieves higher accuracy than state of the art methods.

Currently, we assume that the regression model is linear and after clustering, we identify a single model that has instances from the maximum number of distinct bags and minimizes the fitting error. We are currently investigating two strategies to generalize our approach to non-linear regression. The first one is based on the assumption that a non-linear model can be approximated by multiple piecewise linear models. The second approach modifies the distance measure used within the clustering objective function to represent the fitting error with respect to a non-linear model.

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, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artificial Intelligence 89 (1997) pp. 31–71.

[2] O. Maron, Learning from ambiguity, Ph.D. thesis, Massassachusetts Institute of Technology (1998).

[3] O. Maron, T. Lozano-Pérez, A framework for multiple-instance learning, Advances in Neural Information Processing Systems 10 (1) (1998) pp. 570–576.

[4] R. Rahmani, S. A. Goldman, MISSL: Multiple-instance semi-supervised learning, in: Proceedings of the 23rd international conference on Machine learning, ACM, 2006, pp. 705–712.

[5] Y. Chen, J. Bi, J. Z. Wang, MILES: Multiple-instance learning via embedded instance selection, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) pp. 1931–1947.

[6] C. Yang, M. Dong, F. Fotouhi, Region based image annotation through multiple-instance learning, in: Proceedings of the 13th annual ACM international conference on Multimedia, ACM, 2005, pp. 435–438.

[7] C. Zhang, X. Chen, W. B. Chen, An online multiple instance learning system for semantic image retrieval, in: Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007), 2007, pp. 83–84.

[8] A. Karem, H. Frigui, A multiple instance learning approach for landmine detection using ground penetrating radar, in: 2011 IEEE International Geoscience and Remote Sensing Symposium, 2011, pp. 878–881.

[9] A. Khalifa, H. Frigui, Fusion of multiple algorithms for detecting buried objects using fuzzy inference, in: Proc. SPIE, Vol. 9072, 2014, pp. 90720V–90720V–10.

[10] A. B. Khalifa, H. Frigui, A multiple instance neuro-fuzzy inference system for fusion of multiple landmine detection algorithms, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 4312–4315.

[11] Z. Wang, L. Lan, S. Vucetic, Mixture model for multiple instance regression and applications in remote sensing, IEEE Transactions on Geoscience and Remote Sensing 50 (6) (2012) pp. 2226–2237.

[12] Z. Fu, A. Robles-Kelly, J. Zhou, Milis: Multiple instance learning with instance selection, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 5, 2011, pp. 958–977.

[13] S. Andrews, I. Tsochantaridis, T. Hofmann, Support vector machines for multiple-instance learning, in: In S. Becker, S. Thrun, and K. Obermayer, Eds. Advances of Neural Information Processing Systems 15, Cambridge, MA: MIT Press, 2003, pp. 561–568.

[14] T. Gärtner, P. Flach, A. Kowalczyk, A. Smola, Multi-instance kernels, in: In Proceedings of the 19th International Conference on Machine Learning, Sydney, Australia, 2002, pp. 179–186.

[15] Q. Zhang, S. Goldman, Em-dd: an improved multi-instance learning technique, in: T.G. Dietterich, S. Becker, and Z. Ghahramani, Eds. Advances in Neural Information Processing Systems 14, Cambridge, MA: MIT Press, 2002, pp. 1073–1080.

[16] J. Wang, J.-D. Zucker, Solving the multiple-instance problem: a lazy learning approach, in: In Proceedings of the 17th International Conference on Machine Learning, San Francisco, CA, 2000, pp. 1119–1125.

[17] Z.-H. Zhou, M.-L. Zhang, Neural networks for multi-instance learning. Technical Report, AI Lab, Computer Science and Technology Department, Nanjing University, Nanjing, China, 2002.

[18] M.-L. Zhang, Z.-H. Zhou, Improve multi-instance neural networks through feature selection, in: Neural Processing Letters, vol.19, no.1, 2004, pp. 1–10.

[19] K. L. Wagstaff, T. Lane, A. Roper, Multiple-instance regression with structured data, in: 2008 IEEE International Conference on Data Mining Workshops, 2008, pp. 291–300.

[20] Z. Wang, V. Radosavljevic, B. Han, Z. Obradovic, S. Vucetic, Aerosol optical depth prediction from satellite observations by multiple instance regression, in: Proceedings of the 2008 SIAM International Conference on Data Mining, 2008, pp. 165–176.

[21] S. Ray, D. Page, Multiple instance regression, in: Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01, Morgan Kaufmann Publishers Inc., 2001, pp. 425–432.

[22] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Kluwer Academic Publishers, Norwell, MA, USA, 1981.

[23] H. Frigui, R. Krishnapuram, A comparison of fuzzy shell-clustering methods for the detection of ellipses, IEEE Transactions on Fuzzy Systems 4 (2) (1996) pp. 193–199.

[24] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition, John Wiley and Sons, England, 1999.

[25] R. N. Dave, Use of the adaptive fuzzy clustering algorithm to detect lines in digital images, in: Proc. SPIE, Vol. 1192, 1990, pp. 600–611.

[26] R. Krishnapuram, J. Keller, A possibilistic approach to clustering, Fuzzy Systems, IEEE Transactions on 1 (2) (1993) pp. 98–110.

[27] K. R. Frigui H., A robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Letters 17 (12) (1996) pp. 1223–1232.