WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia FUZZ IEEE

FUMIL-Fuzzy Multiple Instance Learning for Early Illness Recognition in Older Adults

Abhishek Mahnot
Electrical and Computer Engineering
University of Missouri
Columbia, Missouri

Mihail Popescu
Health Management and Informatics
University of Missouri
Columbia, Missouri
PopescuM@missouri.edu

Abstract— **Many important applications in Health Sciences and Biology have underlying datasets that have ambiguous class membership, that is, individual labels are difficult to establish. In such cases, many times, the training examples are easier to label as a group rather than at the instance level. Multiple Instance Learning (MIL) is a supervised learning strategy that addresses this labeling difficulty by employing training example given as positive and negative bags of instances. In this paper we describe a fuzzy variation of the MIL Diverse Density framework (FUMIL) based on ordered weighted geometric operator (OWG) and fuzzy complement operators. We apply FUMIL for early illness recognition of elderly living alone in their home. The available data consists of wireless non-wearable sensor values aggregated at hour level (instance) and ground truth (medical data) available at day level (bag). In our preliminary experiments FUMIL performed better than the traditional MIL framework.**

Keywords- ***multiple instance learning;Fuzzy operators; eldercare;pattern recognition***

I. INTRODUCTION

MANY older adults in the US prefer to live independently for as long as they are able to, despite the onset of conditions such as frailty and dementia. Solutions are needed to enable independent living while enhancing safety and peace of mind for their families [3, 4]. Ageing adults may sometimes purposefully mask any decline in abilities to avoid outside intervention or concern held by their children [4]. Elderly patients are particularly at-risk for late assessment of cognitive changes due to many factors: their impression that such changes are simply a normal part of ageing, their reluctance to admit to a problem, their fear of being institutionalized and even the failure of physicians to fully assess their cognitive function due to belief that no intervention is possible [5].

The above observations suggest the need for automatically detecting early signs of illness and alerting the health care provider in a timely manner [6]. It has been shown that diseases such as cardiac arrhythmia, congestive heart failure and pneumonia, among others, may produce sudden onset of anxiety [2]. Signs of anxiety such as restlessness, insomnia, frequent urination or diarrhea [7] translate in observable behavior changes such as abnormal sleep or room motion patterns. Our early illness approach is based on the assumption that the abnormal behavior patterns can be captured by the environmental sensors (e.g. movement and bed sensors) that we currently have deployed in Tiger Place [8].

We note that the algorithms we develop in this paper attempt to model behavior rather than physiology. While physiology is very similar among humans, the behavior is not. This implies that, while we can train classifiers with data from large amount of different patients, the same is not true for behavioral data. Only sensor data from a given patient can be used to predict his/her behavior. As a result, in behavior prediction experiments, the amount of temporal data is more important than the number of available patients (sample size).

Our sensor data capture external information (behavioral) about residents living alone in their home. Multiple motion and bed sensors are used to track resident's behavior [8]. The sensor data is aggregated at hour level. In the same time, we can asses resident's health status based on the available medical records and self-reported diaries. We constructed training sets by manually classifying every day as abnormal (if an abnormal health event was found) or normal. The goal of our work is to classify a day as "normal" or "abnormal" ("the resident doesn't feel good") based on the collected sensor data during that day. As we can see, the data is more granular (available hourly) than the available ground truth (only daily reports). Aggregating data at day level is not desirable due to averaging effects and decision latency. In previous work [9] we employed several classification strategies such as traditional two class classifiers, one class classifiers and a MIL framework in conjunction with data provided by unobtrusive sensors deployed in the living environment to detect early signs of illness based on health status extracted from nursing visit reports. In this paper we explore the possibility of modifying the traditional MIL framework using ordered weighted geometric operators (OWG). The intent of our inquiry is to make MIL less sensitive to outliers both at the instance and at the bag level. Section I gives an introduction to our research and methodologies. Section II provides an overview of MIL and Section III of FUMIL. Section IV describes our datasets, methodologies and results on the pilot dataset. In Section V we give conclusions and future research possibilities.

II. MULTIPLE INSTANCE LEARNING

Multiple Instance Learning (MIL) [1, 9, 10, 11] is a supervised learning approach in which individual labels for each training example are either hard to assign (e.g. labeling objects of interest in an image) or not available (e.g. in which hour of the day the resident didn't feel well). Instead, class memberships for sets of objects (called bags) are obtained, for example, by labeling the whole image of interest as a "positive" example or labeling the whole day as "abnormal". MIL has been successfully employed in applications such as scene recognition [11], image retrieval [12], drug-target interaction [10] and early illness recognition. A brief introduction in MIL is given in this section followed by Fuzzy Multiple Instance Learning (FUMIL) description in the next section.

MIL classifiers are trained with labeled sets of instances called "bags". Each positive bag, B_i^+, contains at least one positive instance. The individual labels of the instances in each bag are not known at training time. A negative bag B_i^-, contains only negative instances. In our case, a bag consists of 24 vectors of sensor data that correspond to the 24 hours from the day of the nurse report. The days in which the nurse report revealed a concerning health event were labeled "positive" (i.e. they contain some abnormal behavior). The days in which the nurse report didn't mention any health problems were considered "negative". Hence, a positive bag consists of 24 instances (sensor data for each hour) corresponding to a day that contain some abnormal behavior or in other words a day labeled as "positive" make up a positive bag whereas a day labeled "negative" was made a negative bag. We mention that it is possible that the resident had some abnormal behavior due to illness even in the "negative" days. Our proposed FUMIL framework intends to account for this possibility by removing these outliers from the optimization process.

There are many MIL implementations. In this paper we used the diverse density (DD) [1] and its fuzzy modification. The DD of a point $\boldsymbol{x}$ in feature space, $\boldsymbol{x} \in R^p$, is proportional to the number of positive bags with instances close to $\boldsymbol{x}$ and to number of negative bags with instances far from $\boldsymbol{x}$.

If we denote $B_{ij}^+ \in R^p$ the j^{th} instance of the i^{th} positive bag and B_{ij}^- the j^{th} instance of the i^{th} negative bag we can find the point $\boldsymbol{x_{opt}}$ that maximizes DD as:

$$argmax_x \prod_i P(\boldsymbol{x}|B_i^+) \prod_i P(\boldsymbol{x}|B_i^-), \quad (1)$$

where, $P\left(x\middle|B_i^{\pm}\right)$ are computed as:

$$P(\boldsymbol{x}|B_i^+) = 1 - \prod_j (1 - (P(\boldsymbol{x}|B_{ij}^+)) \quad (2)$$

and

$$P(\boldsymbol{x}|B_i^-) = \prod_j (1 - (P(\boldsymbol{x}|B_{ij}^+)) \quad (3)$$

Eq. (3) above expresses the fact that all instances (24 in our case) of a negative bag should be dissimilar to $\boldsymbol{x}_{opt}$ (the prototype of "abnormal behavior"). $P(\boldsymbol{x}|B_{ij}^+)$ measures, for example, the similarity between the point $\boldsymbol{x}$ and instance B_{ij}^+ and can be computed using a weighted Euclidean distance as:

$$P\left(\boldsymbol{x}\middle|B_{ij}^+\right) = \exp(-\sum_{k=1}^{p} w_k (B_{ijk}^+ - x_k)^2) \quad (4)$$

where w_k is a set of scaling factors related to the relevance of each feature k that are also learned in the process of finding the optimal point, $\boldsymbol{x_{opt}}$ (which can be seen as the prototype of positive examples).

III. FUZZY-MULTIPLE INSTANCE LEARNING

In FUMIL, we propose to use fuzzy operators to compute diverse density. Standard fuzzy operations like intersection, complement and union are generalizations of the corresponding classical set operations [13].

In this version of FUMIL we use a combination of aggregation and complement operators at instance level, but, in general, any number of and/or combination of fuzzy operators can be applied depending on the type of dataset FUMIL is being applied on.

The main class of aggregation operators that we use in our application is called *ordered weighted geometric operator* (OWG) [15]. We used OWG because there is an AND operator between bags (i.e. multiplication) in equation (2) and (3). Let $\boldsymbol{a} = \{a_1, \ a_2, \dots \dots, a_n\}$ be a set of vectors, $a_i \epsilon R^p$, and

$$\boldsymbol{w} = \{w_1, w_2, w_3, \dots \dots, w_n\} \quad (5)$$

be a decreasing set of weights such that $w_i \epsilon [0,1]$ for all $i \in [1, n]$ and

$$\sum_{i=1}^{n} w_i = 1 \quad (6)$$

Then, an *OWG* operator associated with $\boldsymbol{a}$ and $\boldsymbol{w}$ is the function

$$OWG(\boldsymbol{a}, \boldsymbol{w}) = {a_{(1)}}^{w_1} * {a_{(2)}}^{w_2} * \dots * {a_{(n)}}^{w_n}$$

where $\{a_{(i)}\}$ is a decreasing permutation of $\boldsymbol{a}$. Effects of different choices of weight vector $\boldsymbol{w}$ on OWG can be found in [15]. In this paper we apply OWG at instance (hour) level, that is, in (3) we use OWG instead of product to emphasize the instances most dissimilar to the target concept. Consequently, for example, (3) becomes:

$$P(\boldsymbol{x}|B_i^-) = \prod_j (1 - (P(\boldsymbol{x}|B_{ij}^+))_{(j)}^{w_j}. \quad (7)$$

In a similar way we modify (2).

To transform the instance similarity $P(\boldsymbol{x}|B_{ij}^+)$ in dissimilarity (see (2) and (3)) we replaced the (1-(.)) instance level operator by a Sugeno fuzzy complement operator [13] given by:

$$c_\lambda(a) = \frac{1-a}{1+\lambda a}, \text{ with } \lambda \in (-1, \infty). \quad (8)$$

For each value of the parameter λ, we obtain one particular *involutive* fuzzy complement. Shape of the function for different values of λ in equation (8) can be seen in [13]. For $\lambda = 0$, the function becomes the classical complement operator used in original MIL. Since there is no specific choice for value of λ and w, we experimentally choose the value of λ and w as described in the next section. We mention that although it is possible to use the Sugeno complement and OWG operator at bag level (i.e. in equation (1)) we didn't explore this alternative in the current paper due to the lack of space. The implementation of the MIL algorithm is described in Fig. 1.We employed a leave-one-out training-testing

Input: ε- a threshold for deciding a positive hour
- N day labels, l_i
$l_i = 1$ if the day was positive (feel bad)
$l_i = 0$ if the day was negative (feel OK)
- $N * 24$ sensor data vectors, $x_k \in R^5$
For each day $i = (1, N)$
Step 1 (train).
Step 1.1 Use data from $N-1$ days, i.e. $24 * (N-1)$ data points, to compute the positive prototype, $\boldsymbol{x_{opt}}$, and feature weights, $w \in [0,1]^p$, using (1);
Step 1.2 Use data from negative bags B_{ij}^- to compute K negative prototypes $\{x_{n1}, \ldots, x_{nk}\}$. The prototypes were obtained as the center of the clusters computed with k-means;
Step 2 (test):
For each hour $j \in [0,23]$ in day i
- compute $d_j = ||\boldsymbol{x_j} - \boldsymbol{x_{opt}}||$ and $d_{jn} = min_{k\in(1,k)}\{||\boldsymbol{x_j} - \boldsymbol{x_{nk}}||\}$
- if $\frac{d_j}{d_j + d_{jn}} < \epsilon$ label day i, $l_i^* = 1$ (positive);
end
End
If no positive hours found, set $l_i^* = 0$ (negative);
End
Step 3. Compare $\{l_i^*\}$ to $\{l_i\}$ for each $\boldsymbol{\epsilon}$ to compute ROC

Figure 1. MIL and FUMIL implementation

approach. Different ways to obtain weight vectors for OWG like *fuzzy quantifier* are reviewed in [15].

IV. DATASETS AND EXPERIMENTS

A. *Experiment 1. MIL-FUMIL comparison on a synthetic dataset*

To compare our FUMIL approach to the one proposed in [1] we generated a synthetic dataset (similar to [1]) comprising of 5 positive and 5 negative bags with 50 instances each (see Fig. 2). Each instance represented a 2D point obtained using a random distribution in the $D = [0\ 100] \times [0\ 100] \in R^2$ domain. The "concept" (the prototype we want to learn) is a 5×5 square in the middle of the domain. A bag was labeled positive if at least one of its instances was drawn from within the square and negative if none did. A plot of the diverse density (DD) surface given by eq. (1) across the domain D is shown in Fig 3(a). We see that the maximum DD is found in the center of the domain D which is within the desired "concept" rectangle. The surface given by FUMIL ((7) and (8)) is shown in Fig. 3(b). The surface was obtained with $\lambda = 5$ and weights $\boldsymbol{w} = \{0, 0, \ldots. 1\}$ for OWG. As we can see from inspecting the two figures, both surfaces present a sharp maximum in the target area. However, the signal to noise ratio (SNR) is much higher for the FUMIL surface (the maximum in Fig. 3(b) is ~0.35, where as in Fig. 3(a) is only ~0.0008).

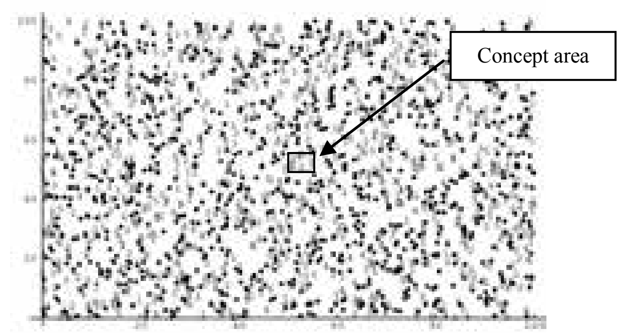

Figure 2. Negative and positive bags are drawn from the same uniform distribution but labeled according to their intersection with the middle square. The instances from negative bags are marked by dots, the ones from positive bags by numbers. The target square contains at least one instance from every positive bag and no negative instances (dots).

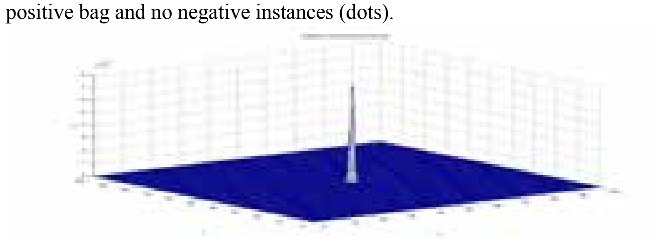

Figure 3(a): Surface using MIL

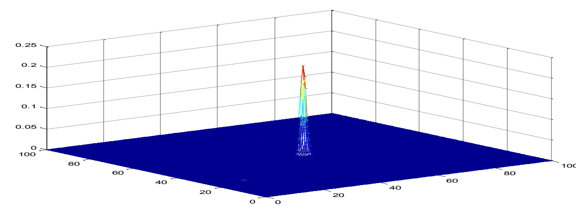

Figure 3(b): Surface using FUMIL

The SNR improvement is due both to the removal of outliers (the points from negative bags "close enough" to the target concept) and the reduced number of instance multiplications.

B. *Sensor Dataset and Methodology*

We now test proposed FUMIL framework for predicting abnormal behavior patterns in elderly. In this test we use sensor data collected in an independent living facility called Tiger Place [8, 14] situated in Columbia, Missouri. The primary goal of Tiger Place is to help the residents not only manage their illness but also stay as healthy and independent as possible. Each resident included in the study has a data logger in his or her apartment that collects data from wireless sensors. The data logger date-time stamps the data, and logs them into file that is sent to a database on a secure server via a wired network connection. Forty seven networks (10 with video) have been installed in Tiger Place apartments; the video part of the network is currently under development. The sensor network consists of several types of sensors mounted in different places throughout the resident's apartments, including motion sensors, bed sensors, and stove temperature sensor. The motion sensors are placed in various places, such as bathroom, bedroom, kitchen, living room, etc. As previously mentioned [9], our early illness recognition approach is based on the intuitions that if the resident does not feel well, his/her sleep and motion patterns are altered. In this study, we used five features ($p=5$) to represent the resident behavior: the total number of motion sensor firings ,bed restlessness, low pulse and low breathing sensors, respectively, for each hour of the day before the nursing report (considered at 12 pm). The fifth feature is represented by the hour of the day when the sensor readings were made. This feature is required in order to differentiate the night time behavior from the day time one.

Although each resident lives alone in his apartment, some extra motion hits were possible due to housekeeping or occasional visits. Although our group has developed algorithms for detecting these events we did not use them here since MIL as well as FUMIL should be able to account for them. Visits are likely to occur in both negative ("feel good") and positive ("feel bad") days, hence feature vectors with an abnormally high motion values generated by a visit will be treated as "negative" instances. The dataset consisted of sensor hits from a 5 year period for a Tiger Place resident. The number of days for the Tiger Place resident considered in this study is shown in Table I.

TABLE I. THE SENSOR DATASET USED IN THIS PAPER

Total Records (N)	Negative(feel good) days	Positive (feel bad) days
744	709	35

C. *Results on the Sensor Dataset*

1) *Choice of Sugeno parameter λ*

In the first set of experiments we studied the influence of the Sugeno complement parameter λ on FUMIL performance. The performance was evaluated using receiver operator characteristic (ROC) curve and area under the ROC curve (AROC). FUMIL performance for various lambda values is shown in Fig. 4.

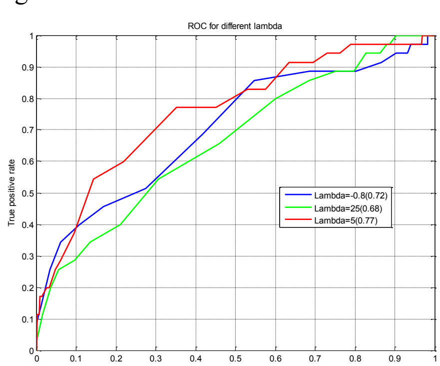

Figure 4. FUMIL performance for various values of λ

As we can see from Fig.4 the best value of λ seems to be 5 (AROC=0.77). We will use $\lambda=5$ thorough out this paper.

2) *Choice of OWG weights, **w***

Next, we tried to determine the best set of weights to use in FUMIL. We tried $\mathbf{w}_1=\{0,0,\ldots,1\}$, $\mathbf{w}_2=\{0, 0, 0, \ldots,1/2,1/2\}$ and $\mathbf{w}_3=\{0, 0, 0, 0,\ldots,1/3,1/3,1/3\}$.

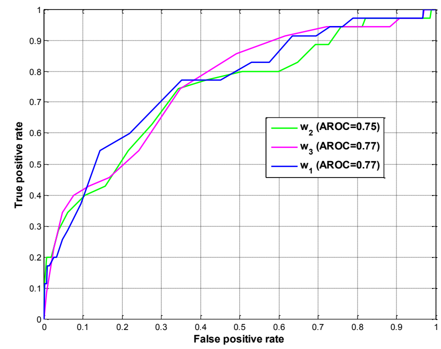

Figure 5. FUMIL performance for various weights, ***w***

From above figure, it seems that both $\mathbf{w}_1$ and $\mathbf{w}_3$ perform equally well. We choose w_1 for further experiments. It should be noted that even though in this case effectively taking a minimum seems to be working the best when choosing weights, that might not always be the case depending on the dataset FUMIL is being applied on.

3) *Comparison between MIL and FUMIL*

The results of the MIL-FUMIL comparison are shown in Fig. 6.

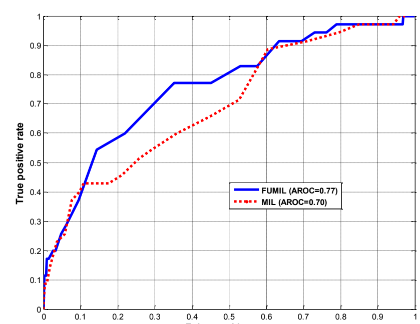

Figure 6. MIL-FUMIL comparison

As can be seen in the above figure FUMIL outperforms MIL by about 10% (FUMIL AROC=0.77 vs. MIL AROC=0.7). This is a significant improvement that is due firstly to the removal of outliers. The most likely source of outliers in our case is the presence of abnormal hours in the normal days (negative bags). Secondly, the improvement is due to the SNR improvement obtained by reduction of the number of multiplications in (2) and (3).

V. CONCLUSION

In this paper we describe, FUMIL, a novel fuzzy logic MIL framework. FUMIL employs ordered weighted geometric (OWG) and Sugeno complement operators to replace traditional noisy-or MIL operators. We compared these two frameworks for detecting early signs of illness in elderly. The experiments were conducted on a pilot sensor dataset obtained from a Tiger Place resident. The detection of early signs of illness may help nursing staff provide interventions that might prevent serious clinical events such as heart attacks or strokes. On our pilot dataset FUMIL outperformed MIL by about 10%.

Our study has several limitations. First, the sample size and datasets are small. Second, the labeling of data (ground truth, i.e. normal vs. abnormal day) was based on medical records but was subjectively performed by the authors. We hope to address both above problems in the future by employing telehealth devices. Also, we hope to improve FUMIL results by testing some optimization technique to determine weight values for OWG. The FUMIL in this paper is applied only at instance level, in future work we plan to apply it on bag level as well.

AKNOWLEDGEMENT

This work has been supported in part by the NSF SHWB grant number IIS-1115956.

REFERENCES

[1] Oded Maron , Tomás Lozano-Pérez, "A framework for multiple instance learning," Proc. of the 1997Conf. on Advances in Neural Information Processing Systems 10, p.570-576, 1998

[2] Chen JP, Reich L, Chung H, "Anxiety disorders," West J. Med. 2002 176:249-253.

[3] Cuddihy P, Weisenberg J, Graichen C, Ganesh M, "Algorithm to automatically detect abnormally long periods of inactivity in a home," Proc. of the 1st ACM SIGMOBILE Intl. Workshop, pp 89-94, New York, 2007.

[4] Rowan J, Mynatt ED, "Digital Family Portrait field trial: support for Aging in Place," Proc. SIGCHI Conf. on Human Factors in Comp. Sys., pp 521-530, New York: ACM Press, 2005.

[5] Hayes TL, Pavel M, Kaye JA, " An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline," Proc. of the 26th Annual Intl. Conf. of the IEEE EMBS, pp 2480-2483, San Francisco, CA, 2004.

[6] Morris M, Intille S, Beaudin JS, " Embedded assessment: overcoming barriers to early detection with pervasive computing," Proc. of PERVASIVE 2005, H. W. Gellersen, R. Want, and A. Schmidt, Eds., pp 333-346, Springer-Verlag, 2005.

[7] Gale CK, Millichamp J, "Generalized anxiety disorder. Clinical Evidence," 2011; 10:1002.

[8] Skubic M, Alexander G, Popescu M, Rantz M, Keller J, "A Smart Home Application to Eldercare: Current Status and Lessons Learned," Tech. and Health Care, 17 (3), 2009, pp 183-201.

[9] Popescu M, Mahnot A, "Early illness recognition in older adults using in-home monitoring sensors and multiple instance learning," IDAMAP-10, Washington DC, Nov 12, 2010, pp 21-25.

[10] Dietterich TG, Lathrop RH, Lozano-Perez T, "Solving the Multiple-Instance Problem with Axis-Parallel Rectangles," Artificial Intelligence Journal 1997, 89, pp-31-71.

[11] Maron O, Ratan AL, "Multiple-Instance Learning for Natural Scene Classification," Proc. 15th ICML, pp 341-349, 1998.

[12] Zhang Q, Goldman SA, Yu W, Fritts JE, "Content-Based Image Retrieval Using Multiple-Instance Learning," Proc. of the 19th ICML, July 2002, pp 682-689.

[13] George JK, Yuan B, "Fuzzy sets and fuzzy logic," Prentice hall PTR A Simon & Schuster Company, upper saddle river, NJ 07458.

[14] Rantz MJ, Marek KD, Aud MA, Johnson RA, Otto D, Porter R, "Tiger Place: A New Future for Older Adults," J. of Nursing Care Quality, vol. 20, no. 1, pp 1-4, 2005.

[15] Herrera F., Herrera-Viedma E., Chiclana F., A Study of the Origin and Uses of the Ordered Weighted Geometric Operator in Multicriteria Decision Making, Int. J. of Int. Systems, vol. 18, 689–707 (2003)