A CLASSIFICATION ALGORITHM WITH LINEAR DISCRIMINANT ANALYSIS AND AXIOMATIC FUZZY SETS

WENJUAN JIA, YINGJIE DENG, CHENYANG XIN AND XIAODONG LIU*

School of Control Science and Engineering
Faculty of Electronic Information and Electrical Engineering
Dalian University of Technology
Dalian, China

WITOLD PEDRYCZ
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, T6G 2G7, Canada

(Communicated by Wan-Quan Liu)

Abstract. In exploratory data mining, most classifiers pay more attention on the accuracy and speed of learned models, but they are lacking of the interpretability. In this paper, an interpretable and comprehensible classifier is proposed based on Linear Discriminant Analysis (LDA) and Axiomatic Fuzzy Sets (AFS). The algorithm utilizes LDA to extract features with the largest inter-class variance. Besides, the proposed approach aims to explore a transformation from the selected feature space to a semantic space where the samples in the same class are made as close as possible to one another, whereas the samples in the different class are as far as possible from one another. Moreover, the descriptions of each class can be obtained by the proposed approach. When compared with well-known classifiers such as LogisticR, C4.5Tree, SVM and KNN, the proposed method not only can achieve better performance in terms of accuracy but also has the capability of interpretability and comprehension.

1. Introduction. In knowledge discovery and machine learning, classification is one of the most important tasks. There are numerous classifiers exploiting various fundamentals, such as Decision Trees, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Linear Regression Analysis (LDA). In these classifiers, LDA is used to identify attributes that account for the most variance between classes. In essence, LDA is a supervised method using known class labels. The ultimate goal of a supervised classification method is to identify to which class a given instance belongs based on a given set of correctly labeled instances. The basic idea of LDA is the high-dimensional pattern sample projection to best identify vector space, to extract the classification information and the compression feature space dimension. Meanwhile, the effect of samples in the new subspace projection

2010 Mathematics Subject Classification. Primary: 03B52, 03E72, 28E10, 94D05.

Key words and phrases. Feature extraction, semantic interpretation, Axiomatic Fuzzy Set theory, machine learning.

The work is supported by National Natural Science Foundation of China under grants 61673082 and 61533005.

* Corresponding author: Xiaodong Liu.

73
guarantee pattern after the biggest class within the class and the minimum distance, the distance between the pattern has the best separability in the space. However, LDA lacks semantic interpretability.

Axiomatic fuzzy sets (AFS) theory was proposed in [15], which can build the relationship between the data and semantic concepts. Unlike the traditional fuzzy set approaches, the difference is that the membership function of AFS was related to data distribution instead of artificially formed with some guidance of human. AFS has been used in various fields [16, 8, 9, 20, 17], such as business intelligence [1], computer vision [10, 18], among others. AFS theory can be applied to machine learning and some other research areas according to the above characteristics. Inspired by the idea of LDA and AFS theory, the new approach is proposed, which not only can help learn and understand a model, but also becomes endowed with sound logic operation and membership function depending upon the data. In sum, the structure of AFS is related to the data and the selected features from LDA. Besides, the global difference and absolute difference is utilized to explore the best separability among patterns in the space.

The sections of this paper are formed as follows: Section 2 is a brief introduction to the AFS theory. In Section 3, we introduce the proposed classifier in detail. Experimental results is covered in Section 4. Finally, a conclusion is given in Section 5.

2. Axiomatic fuzzy set theory. To make this paper self-contained, this section serves as a brief introduction to axiomatic fuzzy set (AFS) [11].

2.1. AFS concepts. Here’s a strict mathematical definition [12] to explain concepts in detail. Let \( \zeta \) be a concept on \( X \), \( R_\zeta \) is called a sub-order relationship on \( \zeta \). And let \( X \) be a collection, \( R \) is a binary relationship on \( X \). For any \( x, y, z \in X \) and \( x \neq y \), \( R \) is called a sub-order relationship on \( X \), if \( R \) satisfies the following relationship:

(1): if \( (x, y) \in R \), then \( (x, x) \in R \);
(2): if \( (x, x) \in R \) and \( (y, y) \notin R \), then \( (x, y) \in R \);
(3): if \( (x, y), (y, z) \in R \), then \( (x, z) \in R \);
(4): if \( (x, x) \in R \) and \( (y, y) \in R \), then \( (x, y) \in R \) or \( (y, x) \in R \)

If \( R_\zeta \) is a sub-order relationship on \( X \), then \( \zeta \) is called a simple concept. Otherwise, \( \zeta \) is called a complex concept.

2.2. AFS algebras. AFS algebras is a way to study the logical relationship between simple concepts and complex concepts. More specifically, it is a family of completely distributive lattices that are made up of \( EI, EII, ..., EI^n, E^#I, E^#II, ..., E^#I^n \). Let \( M \) be a set, \( EM^* \) is defined as follows:

\[
EM^* = \left\{ \sum_{i \in I} \left( \prod_{m \in A_i} m \right) | A_i \subseteq M, i \in I \right\}.
\]

(1)

where \( I \) is an arbitrarily non-empty indexing set. \( \prod_{m \in A} m \) indicates that simple concepts which come from \( A \subseteq M \) by operation “\( \wedge \)”. \( \sum_{i \in I} (\prod_{m \in A_i} m) \) denotes that all \( \prod_{m \in A_i} m \) by operation “\( \vee \)”. AFS logic operation [13] can be written as follows:

\[
\sum_{i \in I} \left( \prod_{m \in A_i} m \right) \vee \sum_{j \in J} \left( \prod_{m \in B_j} m \right) = \sum_{k \in I \cup J} \left( \prod_{m \in C_k} m \right),
\]

(2)
\[
\sum_{i \in I} \left( \prod_{m \in A_i} m \right) \wedge \sum_{j \in J} \left( \prod_{m \in B_j} m \right) = \sum_{i \in I, j \in J} \left( \prod_{m \in A_i \cup B_j} m \right).
\]

2.3. **AFS structure.** AFS structure reflects the key relationship of the data distribution and semantic concepts [15], which is defined as follows:

Let \(X, M\) be sets, \(2^M\) is the power set of \(M\), \(\tau : X \times X \to 2^M\). \((M, \tau, X)\) is treated as AFS structure. Now, here are two axioms:

**Axiom 1:** \(\forall (x_1, x_2) \in X \times X, \tau(x_1, x_2) \supseteq \tau(x_1, x_1)\);

**Axiom 2:** \(\forall (x_1, x_2), (x_2, x_3) \in X \times X, \tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3)\).

Where \(X\) is the domain, \(M\) is simple concept set on \(X\), \(\tau\) can be treated as a kind of structure.

2.4. **AFS membership function.** Let \(X\) be a set, \(M\) is a simple concepts set on \(X\), \((M, \tau, X)\) is an AFS structure based on a data set. If \(x \in X\), \(A \subseteq M\), the definition of \(A^\tau(x)\) reads as follows:

\[
A^\tau(x) = \{ y | y \in X, \tau(x, y) \supseteq A \}.
\]

The definition of membership function is given as follows[14]: for \(\forall m \in M\), \(x \in X\): and concept \(\xi = \sum_{i \in I} \left( \prod_{m \in A_i} m \right) \in EM\),

\[
\mu_\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\sum_{u \in A_i} \rho_\gamma(u)N_u}{\sum_{u \in X} \rho_\gamma(u)N_u}, \forall x \in X,
\]

\[
\mu_\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\int_{A^\tau(x)} \rho_\gamma(t) dP(t)}{\int_{X} \rho_\gamma(t) dP(t)}, \forall x \in \Omega.
\]

where \(N_u\) represents how many times sample \(x \in X\) is observed, \(M\) is a simple concepts set on \(X\), the weight function of simple concept \(\gamma\) is \(\rho_\gamma(x)\). If \(\rho_\gamma(x)\) is continuous on \(\Omega\) for every simple concept \(\gamma \in M\) and \(X\) is a randomly selected data set on \((\Omega, F, P)\). For any \(X \in \Omega\) and \(|X|\) tends to infinity, the conclusion can be obtained that the membership function defined by Eq.5 converges to that defined by Eq.6.

3. **The proposed classifier.** LDA is a supervised learning approach which chooses a projected hyperplane in k-dimensional space such that the samples belonging to the same class on the hyperplane are as close as possible, while the distances between the different classes are as large as possible. Due to these characteristics, LDA is applied to transform the data into a new abstract space. All the steps of whole process is shown in Fig.1. In the new space, we define simple concept \(M\) on \(X\) with abstract features \(f\). Then, we define simple concept set \(M\) and weight function \(\rho_m\), and construct the ensuing AFS structure, so that the membership function can be obtained. To find the descriptions of samples, we need to select the best simple concepts, then complex concepts are formed. According to the label information, the original descriptions of class can be obtained by Algorithm 1.

Moreover, according to the global difference \(G_C\) and absolute difference \(A_C\), to obtain the descriptions, so that some interferences can be excluded to achieve higher
Classifying Based on the Class Characterization

Finding the Semantic Descriptions of Samples

Best Simple Concept

Complex Concept

Finding the Semantic Descriptions of Samples

Best Description of Samples

Candidate Set

Fig. 1. The proposed classifier flow chart

Algorithm 1 Find the Semantic Descriptions of Training Samples

Require: \( X \): a train data set, \( M \): the set of simple concepts on \( X \), \((M, \tau, X)\): an AFS structure, \( 0 < \varepsilon \ll 1, 0 < \theta_2 < \theta_1 < 1, \psi = \sum_{m \in M} m \in EM \).

Ensure: The description \( \zeta \) for all train data.

1: for \( x_i \in X \) do
2: Choose the simple concept \( B_{\varepsilon x_i} \) by a threshold \( \varepsilon \),
\[
B_{\varepsilon x_i} = \{ m \in M | \mu_m(x_i) \geq \mu_{\psi}(x_i) - \varepsilon \}.
\] (7)
3: For \( x_i \in X_{C_i}, \forall y \in X - X_{C_i} \),
\[
A_{\varepsilon x_i} = \{ \gamma = \prod_{m \in H} m | H \in B_{\varepsilon x_i} : \mu_{\gamma}(x_i) \geq \theta_1; \mu_{\gamma}(y) < \theta_2 \},
\] (8)
4: The description \( \zeta_{x_i} \) for \( x_i \),
\[
\zeta_{x_i} = \begin{cases} \arg \max_{\gamma \in A_{\varepsilon x_i}} \{ \mu_{\gamma}(x_i) \}, & A_{\varepsilon x_i} \neq \emptyset, \\ \emptyset, & A_{\varepsilon x_i} = \emptyset. \end{cases}
\] (9)
5: end for

accuracy and more accurate description.

\[
G_{C_i} = \arg \max_{\zeta \in \psi_{C_i}} \left\{ \frac{\sum_{x \in X_{C_i}} \mu_{\zeta}(x)}{X_{C_i}} - \frac{\sum_{x \in X - X_{C_i}} \mu_{\zeta}(y)}{X - X_{C_i}} \right\}
\] (10)
\[
A_{C_i} = \arg \max_{\zeta \in \psi_{C_i}} \left\{ \max_{x \in X_{C_i}} \mu_{\zeta}(x) - \max_{x \in X - X_{C_i}} \mu_{\zeta}(x) \right\}
\] (11)
Finally, we use the descriptions of each class to classify the samples. Similar to Eq. (4), for test sample \( s, A \subseteq M \), one has
\[
A^\tau(s) = \{ x \in X \mid s \succeq m, m \in A \} \subseteq X.
\] (12)
The membership degree \( \mu_{\Xi C_i}(s) \) of the sample \( s \) belonging to \( \Xi C_i = \sum_{i \in I} \prod_{m \in A_i} \mu_{\Xi C_i}(s) \), for \( i = 1, 2, ..., c \) can be obtained by Eq. (5), namely,
\[
\mu_{\Xi C_i}(s) = \sup_{i \in I} \frac{\prod_{\gamma \in A_i} \sum_{x \in A^\tau_i(s) \cup \mathcal{S}_s} \rho_{\gamma}(x)}{\sum_{x \in X \cup \mathcal{S}_s} \rho_{\gamma}(x)}
\] (13)
Then, the class label \( l \) of the sample \( s \) can be achieved by
\[
l = \arg \max_{1 \leq k \leq c} \mu_{\Xi C_k}(s).
\] (14)

4. Experimental results. In order to demonstrate the superiority of the proposed method, several data sets and several methods were used in this experiment. At the same time, by way of eliminating the randomness error, the 10 times 10-fold cross-validation method used in this paper. The data set are obtained from the UCI repository [4], which are heart.statlog(heart), Breast_Cancer_Coimbra (breast.c), seeds, wine, iris, heat, Data_User_Modeling_Dataset_Hamdi_Tolga_KAHRAMAN(USD), vertebral_column(column_2c), caesarian, immunotherapy [6, 7], SomervilleHappinessSurvey2015(SHS2015).

4.1. Case study (iris data set). The selected features obtained by the LDA algorithm are called abstract features. After normalizing these features, then we define the simple concepts: “small”, “medium” and “large”. To better illustrate the approach, iris data set can be used as an example. The iris data has 4 features. By using the LDA, the 4 features can be transformed into 2 abstract features \( f_1 \) and \( f_2 \), respectively. The maximum, average and minimum of each feature is shown in Table 1. Then, the simple concepts can be defined as below: \( m_1 \): \( f_1 \) is small, \( m_2 \): \( f_1 \) is medium, \( m_3 \): \( f_1 \) is large, \( m_4 \): \( f_2 \) is small, \( m_5 \): \( f_2 \) is medium, \( m_6 \): \( f_2 \) is large.

Fig.2 is a scatter plot that shows samples in abstract features space.

| feature | \( f_1 \) | \( f_2 \) |
|---------|---------|---------|
| minimum | -1.73   | -2.36   |
| average | 0.00    | 0.00    |
| maximum | 1.61    | 2.44    |

The membership of every simple concept and all their combinations on all samples are determined with the aid of Eq.5, where \( \rho = 1 \). Now, the membership values of simple concepts and their combinations on all samples can be obtained. The next step is to from the descriptions of samples. The proposed method have three threshold values to filter simple concepts and their combinations. Fig.3 shows the membership degree of the best description \( m_1 \) of the sample \( S_1 \) on all data. Step by step, the descriptions of each sample can be obtained. By using the proposed approach, all class descriptions of iris data are given as shown below:

Class 1 : \( m_1(\text{f}_1 \text{ is small}) \).
Class 2 : \( m_2(\text{f}_1 \text{ is medium}) \).
Class 3 : \( m_3(\text{f}_1 \text{ is large}) \).
Fig. 2. Samples in abstract features space

Fig. 3. The membership degree of $m_1$ on all data

The membership of class descriptions of all simples are shown in Fig.4, Fig.5 and Fig.6. In Fig.4, Fig.5 and Fig.6, it illustrates that the proposed method can distinguish three classes on iris data set. Fig.7 shows the membership function of three classes. The semantic descriptions of three classes can be obtained:

**Class 1:** If abstract $f_1$ is small, then the sample belongs to Class 1.

**Class 2:** If abstract $f_1$ is medium, then the sample belongs to Class 2.

**Class 3:** If abstract $f_1$ is large, then the sample belongs to Class 3.

4.2. **Comparative analysis.** For comparative analysis, some classical classification methods are considered in this paper, such as logistic regression (LogisticR), C4.5Tree, support vector machine (SVM) and k-NearestNeighbor (KNN). These algorithms are implemented by scikit-learn with Python and all the parameters are set by default[19]. For SVM, the regularization coefficients $C$ are chosen from \{0.01, 0.1, 0.5, 1, 2, 10\} for the RBF kernel with $\sigma \in \{0.001, 0.01, 0.1, 0.5, 1, 5, 10\}$ by
Figure 4. The membership degree of description Class 1

Figure 5. The membership degree of description Class 2

Figure 6. The membership degree of description Class 3
cross validation. The number of neighbors are set 3 for KNN. For the proposed approach, we set the parameter $\varepsilon = 0.2$, $\theta_1$, $\theta_2$ are chosen by 10-folds cross-validation.

Table 2. The experiments results-accuracy rates (standard deviation)

| dataset   | LogisticM | C4.5 Tree | SVM  | KNN  | Our method |
|-----------|-----------|-----------|------|------|------------|
| wine      | 0.9556±0.0032 | 0.9364±0.0127 | 0.7900±0.0118 | 0.7074±0.0141 | 0.9866±0.0005 |
| iris      | 0.9593±0.0021 | 0.9520±0.0053 | 0.9826±0.0071 | 0.9647±0.0032 | 0.9867±0.0000 |
| heart     | 0.8399±0.0038 | 0.7544±0.0230 | 0.6918±0.0181 | 0.6604±0.0118 | 0.8407±0.0000 |
| breast_C  | 0.7220±0.0209 | 0.7142±0.0209 | 0.6034±0.0241 | 0.5405±0.0118 | 0.7753±0.0038 |
| seeds     | 0.9228±0.0020 | 0.9286±0.0089 | 0.9271±0.0055 | 0.8828±0.0088 | 0.8590±0.0093 |
| USD       | 0.7309±0.0052 | 0.9321±0.0090 | 0.9510±0.0017 | 0.8247±0.0133 | 0.7912±0.0090 |
| column_2c | 0.8258±0.0030 | 0.8067±0.0205 | 0.8625±0.0031 | 0.8280±0.0064 | 0.7697±0.0040 |
| caesarian | 0.6741±0.0189 | 0.5263±0.0341 | 0.6551±0.0161 | 0.5589±0.0326 | 0.7253±0.0125 |
| immunotherapy | 0.7971±0.0139 | 0.8073±0.0262 | 0.7807±0.0090 | 0.7235±0.0298 | 0.7517±0.0091 |
| SHS2015   | 0.5572±0.0133 | 0.6126±0.0133 | 0.6443±0.0213 | 0.5445±0.0561 | 0.6498±0.0028 |

Experimental results-accuracy rates is shown in Table 2, where the bold number represents the highest accuracy rate. As shown in Table 2, the proposed method can achieve the better performance than other methods. Besides, the proposed method has semantic interpretation realized in abstract feature space, but other methods do not have this characteristic. Thus, the proposed method is a novel classifier with good performance and attractive features.

5. Conclusions. In this paper, a new classifier is proposed, which combines linear discriminant analysis (LDA) and axiomatic fuzzy set (AFS) theory. It provides an innovative approach to transform the data into another feature space. The proposed method transforms features of raw data into abstract features space, then it converts abstract features into the semantic space, so that the proposed method has semantic interpretation which can help the users to understand the classifier easily. In a conclusion, the proposed classifier with semantic interpretation based on axiomatic fuzzy set theory is an excellent classifier. In the future, there are two issues to study: how to automatically obtain the parameters values based on data distribution and how to develop semantic interpretation about features present in raw data.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under grants 61673082 and 61533005.
REFERENCES

[1] W. Bi, M. Cai and M. Liu, et al., A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn, IEEE Transactions on Industrial Informatics, 12 (2016), 1270–1281.

[2] A Das and A. Desarkar, Decision tree-based analytics for reducing air pollution, Journal of Information & Knowledge Management, 17 (2018), 1850015.

[3] K. J. Dsouza and Z. Ansari, Experimental exploration of support vector machine for cancer cell classification, IEEE International Conference on Cloud Computing in Emerging Markets, 2017, 29–34.

[4] D. Dua and E. Karra Taniskidou, UCI Machine Learning Repository, http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science, 2017.

[5] L. I. Jinmeng, Y. Lin and T. Zhu, k-Nearest Neighbor Classification Algorithm Based on Hubness and Class Weighting, Computer Engineering, 44 (2018), 248–252.

[6] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir and A. Khosravi, et al. An expert system for selecting wart treatment method, Computer in Biology and Medicine, 81 (2017), 167–175.

[7] F. Khozeimeh and F. Jabberi Azad, et al., Intralesionional immunotherapy compared to cryotherapy in the treatment of warts, International Journal of Dermatology, 56 (2017), 474–478.

[8] Q. Li, Y. Ren, L. Li and W. Liu, From low-level geometric features to high-level semantics: An axiomatic fuzzy set clustering approach, Journal of Intelligent & Fuzzy Systems, 31 (2016), 775–786.

[9] Q. Li, Y. Ren and L. Li, et al. Fuzzy based affinity learning for spectral clustering. Pattern Recognition, 60 (2016), 531–542.

[10] Z. Li, Q. Zhang and X. Duan, et al., A novel semantic approach for multi-ethnic face recognition, International Journal of Pattern Recognition & Artificial Intelligence, 32 (2017), 15.

[11] X. Liu, W. Jia and Y. Wang, et al., Knowledge discovery and semantic learning in the framework of axiomatic fuzzy set theory, Wires Data Mining And Knowledge Discovery, 8 (2018), e1268.

[12] X. Liu, K. Zhu and H. Z. Huang, The representations of fuzzy concepts based on the fuzzy matrix theory and the AFS theory, IEEE International Symposium on Intelligent Control, 2003.

[13] X. Liu, The fuzzy theory based on afs algebras and afs structure, Journal of Mathematical Analysis and Applications, 217 (1998), 459–478.

[14] X. Liu, W. Wang and T. Chai, et al., Approaches to the representations and logic operations of fuzzy concepts in the framework of axiomatic fuzzy set theory I, Information Sciences, 177 (2007), 1007–1026.

[15] X. Liu, The fuzzy sets and systems based on AFS structure, EI algebra and EII algebra, Fuzzy Sets & Systems, 95 (1998), 179–188.

[16] X. Liu, X. Feng and W. Pedrycz, Extraction of fuzzy rules from fuzzy decision trees: An axiomatic fuzzy sets (AFS) approach. Data & Knowledge Engineering, 84 (2013), 1–25.

[17] Y. Ren, M. Song and X. Liu, New approaches to the fuzzy clustering via AFS theory, International Journal of Information and Systems Sciences, 3 (2007), 307–325.

[18] Y. Ren, Q. Li and W. Liu, et al., Semantic facial descriptor extraction via Axiomatic Fuzzy Set, Neurocomputing, 171 (2016), 1462–1474.

[19] A. Swami and R. Jain, Scikit-learn: Machine learning in python, Journal of Machine Learning Research, 12 (2012), 2825–2830.

[20] W. Wang and X. Liu, Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification, Information Sciences, 294 (2015), 78–94.

[21] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques (Fourth Edition), Elsevier, Singapore, 2017.

[22] X. Zhang, J. M. Chiou and Y. Ma, Functional prediction through averaging estimated functional linear regression models, Biometrika, 105 (2018), 945–962.

E-mail address: wenjuan_jia@yahoo.net
E-mail address: yingjie_deng@163.com
E-mail address: xinchenyang@mail.dlut.edu.cn
E-mail address: xdliuros@dlut.edu.cn
E-mail address: wpedrycz@ualberta.ca