

A Survey on the Classifiers in On-line Handwritten Uyghur Character Recognition System

Wujiahemaiti Simayi¹, Mayire Ibrayim¹, Dilmurat Tursun¹ and Askar Hamdulla²

¹*Institute of Information Science and Engineering, Xinjiang University, China
hojaahmad2014@163.com*

²*School of Software, Xinjiang University, Urumqi, China 830046,
Corresponding Author: askarhamdulla@gmail.com*

Abstract

With the fast development of information technology made people eager to get access the convenient implementations of modern technology in every walk of life. Online handwritten recognition technology for Uyghur is also receiving great need, too. Precious work form researchers for this technology has been gifted many gains. This paper observe the classifiers used in previous work on this field in order to see their adaptabilities for Uyghur online handwritten recognition, and acquire clues for classifier implementation in future work.

Keywords: *Uyghur, on-line handwritten recognition, classification, classifier*

1. Handwritten Recognition Technology and its Significance

The emerging of first printing technology made the human thoughts and history capable for saving and transmitting generation after generation, and sometimes with world-wide effect. After the passing through the industrial revolution, Electrical revolution, Atomic age, human-kind entered into our Information era. The surprising speed of electronic processing and popularization of Internet made the information already globalized for many areas. This fast developing age is putting big threat for many ethnic groups' language, culture and existence, meanwhile providing great opportunities to express and strengthen themselves via sharing the most updated information and technologies. So, millions of books in dark and dusty corners of old libraries are becoming digitized in order to get access to read by computers and transmit by Internet. Everyone can see the huge workload and time cost that this task requires and automatic word recognition tool is a must. Optical Character recognition System, abbreviated as OCR, handled this problem out in a great extent, and we see unending materials in electronic library with quite easy access. After the recognition of printed letters, the next step should be the recognition of handwritten scripts. However, this is a obstacle technology from the very beginning [1]. Everyone has unique handwriting style, and handwriting style of a same one is also influenced by many factors during the actual writing. In addition to that, people always break the standard writing rules. Urgent need for modern communication made the online handwritten technology a necessary. After the great success of Chinese online handwritten technology, minority groups in China also started their endeavor for their own scriptures including Uyghur. As a main language in north-western region of China and central Asia, also holding tight linguistic bonds with many languages in Turkish family of Altaic language system, the study on Uyghur online handwritten recognition technology has significant research value.

2. Uyghur Online Handwritten Recognition System

The research and experiments conducted on Uyghur online handwritten recognition can be summarized as the framework shown in Figure 1

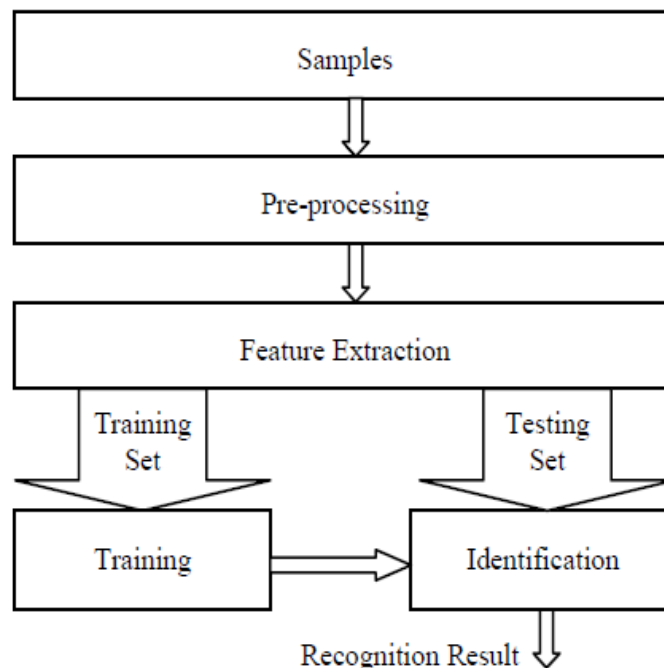


Figure 1. Block Diagram of Uyghur on-line Handwritten Character Recognition System

As shown in Figure 1, collecting handwritten samples always is the priority task, for any idea or method can't be tested without experiment and any experiment can't be run without samples. In online handwritten recognition, the content of sample data used to be the two-dimensional coordinate of the position which touched on the notepad screen or the sensitive surface of intelligent device. Also, the time order of coordinates of touched positions are recorded during handwriting. Usually, the first hand handwritten samples are written in various ways in size, angle and screen's blocks, which means some people write into a corner instead of the center of screen. Therefore, a pre-processing operation should be done before these samples go into the later stages of recognition. Normalization and noise elimination are the basic components in pre-processing stage [2]. The all samples receive the feature extraction process which is vital for recognition accuracy and system efficiency. Some distinctive characteristics of samples are chosen to represent a kind of character or a word. The distinctive characteristics of a category or class are called feature. Features can be observed usually in three ways: structural properties of the object and the statistical characteristics or a hybrid application of those two methods [3]. Some part of collected samples are participated in training process which aims to build model for characters or in more generally speaking, to build models for recognition categories. Sometimes, a few alternative models are built for one single character or category considering the reluctance and randomness of handwriting. In recognition experiment, the untrained samples are expected to go through the identification test. The classification techniques in training and testing units are same in method and parameters. The recognition criteria such as accuracy, recall rate, precision etc are obtained from the identification process.

3. The Classifiers used for Uyghur Handwritten Recognition

This paper observed the classifier applied or proposed in Uyghur online handwritten recognition technology until the recent days. Some scenarios from off-line printed or handwritten recognition are taken into observation for reference, too. Table 1 gives some information of the classifiers or classification techniques used in Uyghur handwritten recognition.

Table 1. Classifiers and Data Sets used for Uyghur Handwritten Recognition

Ref No.	Classifier / Training algorithm	Distance Measure	Data-set Size	Accuracy
4	Minimum Distance Classifier	Minkowsky Distance	NG (Not Given)	NG
5	(MQDF) Modified Quadratic Discriminat Function		48800 samples for test	99.48%
6	Template Matching		NG	>98%
7	Piecewise Linear Sorter	Euclidean Distance	NG	>90%
8	Artificial Neural Networks		10 different users	91%
9	Artificial Neural Networks (BP with Gradient Descent)		230 samples for training	98.21%
10	Nearest Neighbor Classifier		10 different handwriters	NG
11	Minimum Distance Classifier	Euclidean Distance	465 samples	>86%
12	Minimum Distance Classifier	Weighted Euclidean & Chi-square Distance combined	handwritten samples from Chinese, English, Uyghur, Tibetan	>92%
13	BP and HMM (Freeman code) combined method		26617samples for mathematical formula, 400 samples for test	91.75%
14	BP		480	96.46%
15	SVMwith Gradient Descent		400×128=51200 samples	>90%
16	Template Matching with XOR operation		400 samples for test	97%
17	Minimum Distance Classifier	Weighted Euclidean Distance.	6×100=600 samples 400 for training , 200 for testing	89%
18	Minimum Distance Classifier	Euclidean Distance,	70 groups for training, 45 groups for testing,	75%
19	Minimum Distance Classifier	Euclidean Distance	400 × 128=51200 samples, 50% trained, 50% tested	
20	HMM		2000 samples , 90% for training, tested on five different fonts of Uyghur typed letters,	
21	SVM with RBF kernel Surpassed MQDF, DLQDF, LVQ (modified quadratic discriminant function , the discriminative learning quadratic discriminant function , the learning vector quantization classifier)		400 × 128=51200 samples, 70% trained, 30% tested	89.08%
22	SVM		32* 50 for training, 32*100 for testing,	93%
23	Minimum Distance Classifier	Euclidean Distance & Manhattan Distance	120 users	NG
24	Combined classifiers using KNN	DTW with Euclidean Distance	51200 samples	
25	Weighted Naïve Bayes		102×128=13056samples, 60 groups for training, 42 groups for testing	93.15%
26	KNN	Chi-square distance	1000 signatures	98.5%.
27	KNN	Euclidean distance,	115 group of samples, 75groups for training. 40 groups for testing	95%
28	Classifier Combination Using Nearest Neighbours, MST and DTW clustering	Euclidean Distance	51200 samples	89%
29	Minimum Distance Classifier	Euclidean Distance	12800 samples, 280 groups for training, 120 groups for testing	78%

Ref No.	Classifier / Training algorithm	Distance Measure	Data-set Size	Accuracy
30	Semi-supervised K-means clustering		Samples from IFN/INET2.0 Arabic Storehouse and self-collected 160 ones	
31	Minimum Distance Classifier	Euclidean distance	12800 and 51200 samples, 80% for training, 20% for testing	94.50%
32	Minimum Distance Classifier	Euclidean distance	12800 and 51200 samples, 80% for training, 20% for testing	98.77%
33	Combination of Gaussian Mixture Model and Hidden Markov Model			99%
34	HMM and Statistical Language Model		24 000 Uyghur words, 759 Arabic words	97.45%, 85.80%

With a quick glance to Table 1, it is very clear to see that the sample data library for many of the experiments are not from one source and some of them are even not addressed. In addition to that, the samples are different in volume. This situation makes us unable to compare the recognition results and inconvenient to analyze the performance of different classifiers.

From Table 1, it is not difficult to find that see that minimum distance classifier has been a hot attractive point for its simple nature in theory and application. Some classic or popular classifiers including HMM, SVM, NB have shown themselves in seldom appearance. The new developing technologies such as artificial neural networks also have been tested with the application of BP, mostly. Ensemble application of different classifiers appeared in most recent work and produced the highest classification or recognition rate in experiment.

4. Popular Classifiers

From Table 1, we can easily see that many of the popular classical classifiers have not been applied yet [35, 36], except SVM, HMM, BP, NB etc. perhaps this is due to the lacking of research staff on Uyghur handwritten recognition and for some other reasons. However, applied classification methods and the obtained recognition results using them do not necessarily mean that they are the ones which most fit for Uyghur handwritten recognition. So, there are lots of work has to be done for classification study and its application on Uyghur handwritten recognition. In this section, some popular classifiers with their advantages as well as shortcomings are introduced [37~40].

- **RF--Random Forests**

Random Forests are showing themselves as very successful ensemble learning method for classification regression and other tasks, which are integrated form multiple decision trees. The easily overfitting problem is beaten in Random Forests. Random forests found themselves as good solver for lots of classification problems. For their distinct advantages of being fast and scalable and convenient for tuning multiple parameters simultaneously, Random Forests are earning quite popularity in these days.

- **SVM—Support Vector Machines**

SVMs are the most popular and most successful learning algorithm for classification tasks and regression analysis, nowadays. High accuracy, nice theoretical guarantees regarding over fitting and with an appropriate kernel they can handle well the linearly inseparable data in base feature space. SVMs are especially popular in text classification problems where very high-dimensional feature spaces have to be worked on.

Disadvantages: SVMs are mostly Memory-intensive, hard to interpret and kind of annoying to run and tune. They are very much relied on adequate choosing of kernel methods; in addition to that, there is no generalized solution for kernel choosing. SVMs sensitive for missed data, quite dependent of the feature set size and take a considerable amount of time to train, this leads to decrease the interactive efficiency.

- **NB--Naive Bayes**

Naive Bayes is a popular algorithm in the family of simple probabilistic classifiers. The Bayes theorem with strong independence assumptions between data features are the main theoretical base of naïve Bayes classifiers. It has a property of simplicity in mathematical foundation and training computation and quick to execution. Stable classification results are the main advantage of Naïve Bayes algorithm. The classification performance is not much influenced by missed data and small size of parameters can meet the requirements. However, Naïve Bayes classifiers lose performance on big data, and most notably, they can't learn interactions between data features. Naive bayes is not powerful enough to create accurate models on large data.

- **KNN—K Nearest Neighbours**

k-NN is one of the most popular and simplest, non-parametric learning algorithm used for classification and regression problems. It is a typical instance-based, lazy learning algorithm, where the function is only approximated locally and all computation is deferred until classification. KNN has advantages of being pretty easy and convenient for real world applications. When training data size is becoming larger, the KNN shows its power and effectiveness, but with the high cost of computational time consuming. As a Lazy Learning algorithm, KNN almost does not learn anything and classification is influenced by un-equally prepared volume of sample data.

- **Logistic Regression**

Logistic regression can be seen as a special case of generalized linear model. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. Logistic Regression has nice probabilistic interpretation and is not influenced by the correlation relationship of data features. The trained model can be easily updated to take in new data. It is a very reasonable alternative for a probabilistic dynamic framework that the new received data can be quickly incorporated to update the model in time.

- **DT--Decision Trees**

One of the most easily applied predictive modeling approaches, Decision Trees have Advantages of being easy to interpret and explain, can easily handle the interactions between the features of data. Decision Tree methods are non-parametric, so they are not influence by whether data is linearly separable. However, DTs easily over fit and neglect some feature interactions very often. They are sensitive for missed data and features generated from insufficient data volume. They don't support online learning, so new tree has to be built when new examples come on.

- **ANN--Artificial Neural Networks**

Artificial neural networks are generally presented as systems of interconnected neurons which exchange messages between each other. The tunable weighted connections between neurons make artificial neural networks' adaptive to inputs and capable of self-learning. Distributed processing and noise protective mechanism, High classification accuracy, powerful online memorizing and learning ability make artificial neural

networks quite applicable for complex nonlinear problems and tasks that are hard to solve using ordinary rule-based algorithms. Sometimes, it is hard to adapt neural networks. Because they require large volume of parameters, can't observe the learning process, results are difficult to explain and take as granted. Most considerably, the learning time is too long that neural networks sometimes even can't learn.

● Genetic Algorithm

Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution. Genetic algorithms have strong stochastic searching ability and advisable for ensemble application. Beginners perhaps find Genetic algorithms inconvenient for programming. They often show low accuracy when they are applied alone.

● Adaboost

Adaboost is a kind of classifier with high precision. It provides the framework for classification and makes the establishing of sub-classifiers very easy. It is simple in foundation and does not need feature selection. You don't worry about over fitting problem.

● HMM

A HMM can be presented as the simplest dynamic Bayesian network. HMM is applicable for more complex data structures and the modeling of non-stationary data.

In fact, Hundreds of classifiers are being used in different fields. We would like to recommend a few more classifiers from different categories of classification algorithms [41].

1. Elm_kernel_m from Artificial Neural Networks
2. C5.0_t from Boosting algorithms
3. fda.t from Discriminant Analysis
4. RandomCommittee_w from ensemble classifiers
5. Multinom_t from Logistic and Multinomial Regression algorithms
6. gcvEarth_t Multivariate adaptive regression splines
7. glmnet_R from Generalized Linear Models
8. treebag_t From Bagging Algorithms
9. RandomSubSpace_w from Decision Tree algorithms
10. C5.0Rules_t from Rule based methods

5. Recommendations

We believe the following recommendations will be helpful for the advancement of Uyghur handwritten recognition.

First of all, it is strongly recommended that an authorized handwritten sample library with enough volume has to be established, so that each one can use it freely and compare the performance of different classifiers and new methods. An open data source does provide a great forum of learning each other and prevent the redoing of some work.

Second, different classifiers should be comparatively tested, instead of putting just one kind of classifier on experiment. A good classifier in common does not necessarily will be the best one for all scenarios. Perhaps, some data structures or features require the quite sophisticated classification methods with time- costing, while some data structures need to be processed with handy and quick classifiers. Some classifiers may give better accuracy, but produce low efficiency of whole system due to lazy learning and slow solving nature. In contrast, although some classifiers can't make competent classification accuracy, the speed from their application makes the user more satisfactory.

Thirdly, proposing new kinds of classification methods or the adapted modification of classic classifiers are very much hoped. The natural languages are very complicated that they are uttered obeying to natural rules and language holder's subjective influences such as thought, ambition, temper and emotion. So, the classifier that simply used on the natural or mechanical stochastic data may will not perform as good as usual. The different characteristics of natural language data should be studied and appropriate classification methods are applied. It is hopeful that classic classifiers can be adopted with special modifications, too.

Then, the ensemble application of classifiers should be studied for a successful application of integrated classifiers can meet acceptable accuracy as well as the satisfied speed [42]. It is possible that data structure from natural languages like handwritten samples can be observed in different angles to find their static or dynamic, simple or complicated, structural or statistical, surface layer or deep-in characteristics. When appropriate classifiers are applied for each part or feature of data, the classification accuracy and overall system efficiency can be improved greatly. However, this is dependent on careful studying both of data structure the classifier attributes.

6. Conclusion

The great need for new communication technologies and intelligent applications require handwritten recognition technology to have high accuracy and acceptable speed. Studies on Uyghur handwritten recognition have been gifted with many achievements, too. Still, lots of work is open for researchers for better improve the accuracy and efficiency standards of both online and offline implementations of Uyghur handwritten recognition. This paper examined the classification methods applied in Uyghur handwritten recognition and the sample data sources the classifiers run on. Although, the sample data sources with different size and standards make the comparing of classifier performances impossible, many advances in classifier application have been obtained. A brief introduction of popular classifiers and a few recommendations for applying them in more advanced ways are given, and this can be the main content of future work on Uyghur handwritten recognition.

Acknowledgement

This work is supported by Natural Science Foundation of China (No. 61263038 and No. 6146208) and Natural Science Foundation of Xinjiang (No.2015211C285) and PHD Start Foundation of Xinjiang University (No.BS130117).

References

- [1] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Fourth ed, Academic press, USA, (2009).
- [2] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Third edition, Publishing House of Electronics Industry (with permission to publish), China, (2010).
- [3] A. T. Al-Taani, "Recognition of On-line Arabic Handwritten Characters Using Structural Features", *Journal of Pattern Recognition Research*, (2010), pp. 23-37.
- [4] A. Halmurat, "Research and Development of a Multifont Printed Uyghur Character Recognition System", *Chinese Journal of Computers*, vol. 27, no. 11, (2004), pp. 1480~1484.
- [5] W. Hua and D. Xiao_Qing, "HALMURAT, Multi-font multi-size printed Uyghur character recognition", *Journal of Tsinghua University (Sci& Tech)*, vol. 4, no. 7, (2004), pp. 946~49.
- [6] Y. Abaydul and M. Mamutjan, "The Research Entering Electronic Text and Written Style Editor of Uyghur Language", *The Proceedings of 2nd Chinese Conference on Human-Computer Interaction, (CHCI'06)*, (2006), pp. 66-71.
- [7] W. Fang, "Research and Implementation of the On-line Uyghur Handwriting Recognition Technology", Ms thesis, Xinjiang University, Urumqi, (2007).
- [8] D. Abilhayer and G. Altenbek, "Handwritten Kazakh character recognition system using artificial network", *Computer Engineering and Applications*, vol. 44, no. 1, (2008), pp. 225- 228.
- [9] J. Jian_Zhong, "The Research of feature selection and classifier design for Printed Offline Uyghur

- character recognition”, Ms thesis, Xinjiang university, Urumqi, (2008).
- [10] L. Yuan, “Uyghur Handwriting Distinction Key Method Research”, Ms thesis, Xinjiang university, Urumqi, (2008).
- [11] C. Zhi-hong and L. Yuan, “Research and Realization of Feature Extraction on On-line Uyghur Handwritten Characters”, Technological Trend, vol. 14, (2009).
- [12] L. Xin, D. Xiao-Qing and P. Liang-Rui, “A Microstructure Feature Based Text-independent Method of Writer Identification for Multilingual Handwritings”, Acta Automatica SINICA, vol. 35, no. 9, (2009) September.
- [13] P. Wei-Min, Q. Xiang-Wei and Y. Abaidula, “Terminal Embedded Searching Algorithm of Handwriting Recognition for Connection Strokes in Mathematical Formula”, Journal of Jiangnan University (Natural Sciences), vol. 37, no. 2, (2009), pp. 40~43.
- [14] R. Hong-Yu, Y. Bao-She and T. Yuan, “On-line & Handwritten Uyghur Character Recognition Based on BP Neural Network”, Microelectronics & Computer, vol. 27, no. 8, (2004) August, pp. 239-241.
- [15] S. Mutallip, “Research on the Technology of Trainable Online Uyghur Handwritten Character Recognition”, M. S. thesis, Dalian Maritime University, Dalian, (2010).
- [16] Y. Bao-She and W. Slam, “Handwriting Recognition Algorithm of Uyghur Letter”, Computer Engineering, vol. 36, no. 2, (2010).
- [17] G. Naby, K. Ubul and K. Moydin, “Uyghur handwriting identification technology based on combining of multi-directional features”, Computer Engineering and Applications, vol. 49, no. 3, (2013), pp. 139-142.
- [18] T. Fuxiu, “On-line Uyghur Handwritten Character Recognition Based on Mobile Platform”, Ms thesis, Xidian University, (2011).
- [19] Ranagul, “Research on the Key technologies of Online Handwritten Uyghur Word Recognition”, M.S. thesis, Xinjiang University, Urumqi, (2011).
- [20] E. Jume, A. Halidan and H. Hao, “Recognition of extracting Uyghur texts from videos images”, Computer Engineering and Applications, vol. 47, no. 36, (2011), pp. 190-192.
- [21] M. Ibrayim, Z. Heng, L. Cheng-Lin and A. Hamdulla, “A Method for Online Handwritten Uyghur Character Recognition”, Pattern Recognition and Artificial Intelligence, vol. 25, no. 6, (2012) December, pp. 979-986.
- [22] H. Lin-Feng and Z. Hui, “SVM Based Recognition Algorithm of Online Uyghur Handwritten Characters”, Computer Applications and Software, vol. 29, no. 3, (2012) March, pp. 179-182.
- [23] S. Jie, K. Moydi and Z. Wei-Yu, “Uighur Handwriting Identification Based on Microstructure Feature”, Computer Engineering, vol. 38, no. 24, (2012) November, pp. 175-178.
- [24] M. Ibrayim, K. Moydin and A. Hamdulla, “Multi-classifier combination scheme for recognition of online handwritten Uyghur characters”, Computer Engineering and Applications, vol. 48, no. 31, (2012), pp. 140-145.
- [25] X. Ya-Mei, L. Zhao-Yang and L. Jiang, “Handwritten Uyghur Character Recognition based on Radical Dictionary and time division direction feature”, Journal of Jilin University(Engineering and Technology Edition), vol. 43, no. 3, (2013) May, pp. 740-746.
- [26] G. Abliz, K. Ubul and K. Moyidin, “Research on off-line Uyghur signature recognition technology based on multiresolution geometric features”, Computer Engineering and Applications, vol. 49, no. 16, (2013), pp. 168-171.
- [27] J. Wen, L. Zhao-Yang and L. Jing, “Recognition of Uyghur Handwritten Character Recognition Based on Directional Element Feature”, Micro-Electronics & Computer, vol. 30, no. 10, (2013) October, pp. 97-100.
- [28] M. Ibrayim, “Key Technologies for Recognition of On-line Handwritten Uyghur Characters and Words”, PhD dissertation, Wuhan University, (2013) May.
- [29] K. Zulpiya, “Research on Online Uyghur Handwritten Character Recognition Based on Feature Combination”, M. S. thesis, Xinjiang University, Urumqi, (2013).
- [30] J. Jianzhong, G. Shengrong and Y. A. Ahudoulikemu, “Recognising Offline Handwritten Uyghur Text Based on Mixed Features of Elastic Mesh”, Computer Applications and Software, vol. 31, no. 9, (2013) September, pp. 172-176.
- [31] W. Simayi, M. Ibrayim, D. Tursun and A. Hamdulla, “Research on On-line Uyghur Character Recognition Technology Based on Center Distance Feature”, Proceedings of 13th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Athens, Greece, SP-6046, (2013) December 12-15, pp. 000293 – 000298.
- [32] A. Hamdulla, W. Simayi, M. Ibrayim and D. Tursun, “Research on On-line Uyghur Handwritten Character Recognition Technology Based on Modified Center Distance Feature International Journal of Signal Processing”, Image Processing and Pattern Recognition, vol. 7, no. 5, (2014) October, pp. 409-424.
- [33] R. Tursun and W. Silamu, “A Uyghur Online-Handwritten Word Recognition System”, Journal of Chinese Information Processing, vol. 28, no. 3, (2014) May, pp. 112-115, 122.
- [34] N. Kadier and P. Liang-Rui, “Halimulati, Uyghur and Arabic Recognition Methods Based on HMM and Statistical Language Model”, Computer Applications and Software, vol. 32, no. 1, (2015) January, pp. 171-174.

- [35] J. Han and M. Kamber, "Data Mining: Concepts and Techniques, Second Edition", Machinery Industry Press (with permission), China, (2007).
- [36] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining", Knowl Inf Syst, vol. 14, (2008), pp. 1-37.
- [37] Y. Xiaofan and C. Tinghuai, "Inherent advantages and disadvantages of artificial neural networks", Computer Science, vol. 21, (1994), pp. 23-26.
- [38] M. Dave, S. B. Singh and S. Manchanda, "An Empirical Comparison Of Supervised Learning Processes", International Journal of Engineering, vol. 1, no. 1, (2006), pp. 21-38.
- [39] T. V. Spina, J. A. Montoya-Zegarra, F. Andrijauskas, F. A.Faria, C. E. A. Zampieri, S. M. Pinto-Cáceres, T. J. de Carvalho and A. X. Falcao, "A Comparative Study among Pattern Classifiers in Interactive Image Segmentation", Sibgrapi Conference on Graphics, (2009), pp. 268-275.
- [40] J. Demšar, "Statistical Comparisons of Classifiers over Multiple Data Sets", Journal of Machine Learning Research, vol. 7, no. 1, (2006), pp. 1-30.
- [41] M. Fernández-Delgado, E. Cernadas, S. Barro and D. Amorim, "Do we Need Hundreds of Classifiers to Solve Real-World Classification Problems", Journal of Machine Learning Research, vol. 15, no. 1, (2014), pp. 3133-3181.
- [42] S. Tulyakov, S. Jaeger, V. Govindaraju and D. Doermann, "Review of classifier combination methods", In Machine Learning In Document Analysis & Recognition, vol. 90, (2008), pp. 361-386.

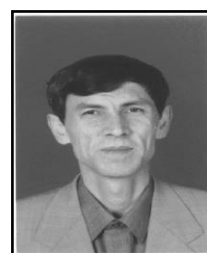
Authors



Wujiahemaiti Simayi, He has received his B.E. and M.S. degree in Electronics, Signal & Information processing from Xinjiang University, China, in 2009 and 2014, respectively. Currently, he is a PhD candidate in Computer Applications in Xinjiang University, and working as research assistant at the Key Laboratory of Intelligent Information Processing, Xinjiang University, China. His research interests include feature extraction and classification techniques, handwritten character recognition, computer simulation for atmospheric pollution dispersion modeling.



Mayire Ibrayim, He received her M.S. and Ph.D. degree in communication engineering, communication and information system from Xinjiang University and Wuhan University, China, in 2009 and 2013, respectively. Since 2013, he has been working as a teacher in School of Information Science and Engineering, Xinjiang University, and now she is doing her postdoctoral research work in a workstation of Computer science and Technology of Xinjiang University. Her research interests include image processing and pattern recognition.



Dilmurat Tursun, He received his B.E. in 1983 in Electrical Engineering, from Xinjiang University of China. Currently, he is a professor in the Institute of Information Science and Engineering of Xinjiang University. He has published more than 50 technical papers on experimental phonetics and image processing.



Askar Hamdulla, He received B.E. in 1996, M.E. in 1999, and Ph.D. in 2003, all in Information Science and Engineering, from University of Electronic Science and Technology of China. In 2010, he was a visiting scholar at Center for Signal and Image Processing, Georgia Institute of Technology, GA, USA. Currently, he is a professor in the School of Software Engineering, Xinjiang University. He has published more than 150 technical papers on speech synthesis, natural language processing and image processing. He is a senior member of CCF and an affiliate member of IEEE.