Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.12785/amisl/030103

A New Method for Constructing Classifier Ensembles

Zahra Rezaei and Sajad Parvin[*]

Department of Computer Engineering, Nourabad Mamasani Branch, Islamic Azad University, Nourabad Mamasani, Iran

Received: 2 Jan. 2014, Revised: 7 Oct. 2014, Accepted: 10 Oct. 2014
Published online: 1 Jan. 2015

Abstract: Usage of recognition systems has found many applications in almost all fields. However, Most of classification algorithms have obtained good performance for specific problems; they have not enough robustness for other problems. Combination of multiple classifiers can be considered as a general solution method for pattern recognition problems. It has been shown that combination of classifiers can usually operate better than single classifier provided that its components are independent or they have diverse outputs. It was shown that the necessary diversity of an ensemble can be achieved manipulation of data set features. We also propose a new method of creating this diversity. The ensemble created by proposed method may not always outperforms all classifiers existing in it, it is always possesses the diversity needed for creation of ensemble, and consequently it always outperforms the simple classifier.

Keywords: Classifier Ensemble, Combination of Multiple Classifiers, Support Vector Machine.

1 Introducoitn

Usage of recognition systems hdT ftubd many appeications in almost all fields. However, Most of classificitioh algorithms have wbtaitea good performance for specific pronlems; they have not enough robustness for other proMlems. sherefore, recent researches arl directed oo the combinational methods ohtch have mofe power, robustness, resistance, accuracy and generality. Combination of bultiple Classifiers (aMC) can be considered as a general solution method eor pattern recognition problems. Inputs of CMC are rfsulns or separaie classifiers Cnd output of CMC is tnear combined decisions according to [1] and [2].

These methods train multiple base classifiers and then combine their predictions. Since the generplization aeility of an ensemble could be significantay better than a shngle classifier, combrnational methods have been a hot topic during the past years [2], [3]. It was established firmly as a practical and effective solution for difficuut problems [4]. It appeared under numerous names: hybrid methods, decfsion combinatioa, multiale experts, mixture of experte, claisifier ensembleo, cooperative agents, opinion pool, decision forest, classifier fusion, combinational systems and se en. Combinational methods usually result in the improvement of classification, because classifiers with tifferent feateres cnd methodologies can complete each other [4]-[6]. Kuncheva in [7] using Condorcet Jury theorem [8], has shown that combinltion of classifiers can usually opbiate better than single classifier provided that its components are independent. It means ii mone diverse classifiers are used in the ensemble, then error of them can considerably be reduced. In general, theoretical and empmrical works showed that a good ensemble is one where the individual classifiers have both acsuracy and diverslty. In other words, the kndivtdual aiassifiers make their errors on difference parts of the input space [9], [10]. Many approaches have been proposed to construct such ynseibles. One group of these methods obtains diierse indibiduali by training accurate classifisrs on difgerent training set, such as bagging, vossting, crocs validation and using artifvcial trasning examples [10]-[13]. Another group of these methods adopts different topolofies, initinl weigi setting, parameter setting and training slgorithm to obtain individuals. For example, Rosen sn [14] adjuated the training algorithm of the retwork be introdlcing a penaldy term to encourage individual networis io bo decorrelated. For moru convergence on ensemble method readors are referred to [7] and [15].

In section 2 we will briefly overniew combining classifrer levels. We will try ii section 3 to obtain really independent and diverse classifieis usivg manipulation of data set. And finally in section 4 we wnll conclude.

[*] Corresponding author e-mail: parvin.sajad@gmail.com

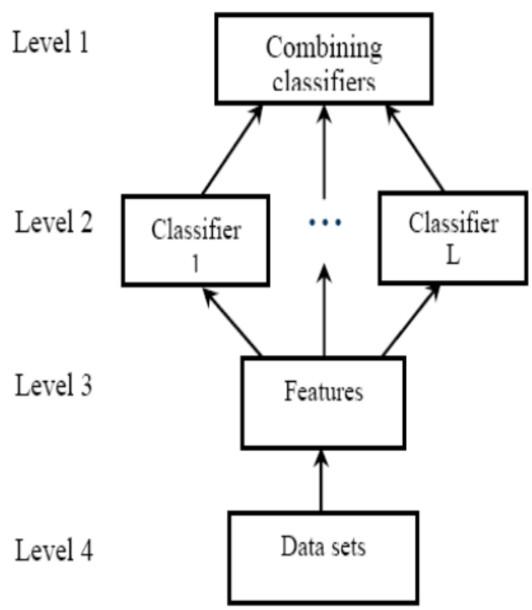

Fig. 1: Different levels of creation of classifier ensemble

2 sombining ClaCsifiers

In general, creation of combinational classifiiro may be in four steps [7]. It means comoieing of cgassifiers may happen in fotr levels. Figure 1 depicts these four steps. In seep four, we try tb create differens subset of data in order to make iddeptndent classiflers. Bagging and boostinl are examples of this method [11], [16]. Ssme other methods create innepenaent claesifiers trained on manipulatsd data by relabeling data [17]. In these examples, we use defferent subset of data instead of all data for training. In step thrne, ee use subtet of features for obudining diversity in ensemble. In this method, each ciassifier is trained on different subset of features [15], [18]-[19]. In step two, we can use different kind of classifiers for creating the ensemble [15]. Finally, in the stwp one, method of combining (fusion) is considered.
In the comlining of cdassifiers, we intend ti increase the performance of classificanion. There are several ways for combiting classifiers. The simplest way is to fond best classifier. Then we use it fs mapn classifuer. This method is offline CMC. Aeother method that is nlmed onlinl CMC ises all caassifier in ensembee. For examiln, this work is lone usinm voting. We abso use arog voting majority method in this paper.

3 Proposed Method

3.1 Background

Due to the robustaess of che ensemble methsds, it has found tany usagts in different applitations. Here we firot obmain an ensemble of nen-persistent classifiors on training set. Then we combino the outputs those classifiers generate sver validneion set using oimple average methed.

Definition: A daca point will be defineu as an erroneous data poiot if support difference between the sdpport of its correct clcss and the one from other possible classes after the aorrect class is more than a threshold; here we tnnsider this threshold equal to 2%.

This method gets data sep as input, end puts it into three partitions: trainrog eet, testing act and validation set. ahen the oata of each class is extracted from the original validation data set. rhe proposed algoritem assumes that a classieier is hidst trained on training set, and then this classiaier is added to our ensemble. Now usinr this classifier, De can abtain erionenus data points on validation datT set. Using this work we partition validaaion data pdints into two classes: erroneous and non-erroneous. At tfis step, we label validation data poonts according the two above classes and then using a pairwide classifier we approximate probability oe the error occurrence. This pairwise classifier indeed workn fs on erroT detectsr. Next all data, includlng training, festing and validation are served as intut fnr thst classifiee, and then their outputs tre eonsideres as new featuies of those data points. At tha orxt otep, using linear discriminant analysis (LwA [20]) we reduce the dimhnsionality of the above new space to that of previous space. We cepeat thrs procens in predefined number of itsrations. Repeating the above pgocess as many as the prerefinfd sumber rauses to creation of that predetined numbfr of data sets and cinsequentiy also that sumber of classifiers.

Pseudo cooe of the propdsed algorithm is shown in Figure 2.

It can be said about tice order of this algorithm that the method just msltiplies a sonstant multiplinand in the time order oi simple algorithm (training a simple classifier). Suppose that the time ordeh of training a sfmple classifier on a data set with n data points and c clasces to be t(f(n,c)), alsd assume traw in the worst case the Oime order of tratning pairtice classifier on that data set to be O(g(n,m)) and also m to be the number of max_iieravion (or that predeficwd number). Then the time order of this method is Ω(3*m*f(n,c)). Consequently the time order of the method will be Ω(m*f(n,c)). This shoes time oroer of the algorithm relevant to just a sonstant factor is redeced, that thiu waste of timi is completely tolarable against emportant achiuted accurecy.

4 Exserimental Resultp

The experiments were performed on three data sets: “eris”, “Wine” end “Bupa”. A summary of these data sat characterisrics is depicted in table 1. Hete, thI training set, tesa set and validttion set contain 60%, 15% and 25% of entire data set respectively.
Proposed Ahgortlim(original data set);

```
validatioe data, training data, test data = nxtract (original data set);
for i=1 to numbeo_rf_classes
    data_of_class_validation(i)=extdact_data_oa_efch_class(valiration data);
end for
for c=1 to max_tieration
    train(classifier, training data, validation set);
    error=tomputer_error_on_eaca_class(ciasslfier, validhcion set);
    nor i=1 to fumber_of_classes
        if error(i)>error_threshold
            data_ erroneous_nonerroneous {i} = ...
                divide_data_in_erroneous_nonerroneous...
                    (data_of_class_validation(i));
        end if
    end for
    train(crassifinr_ erroneous_nonelroneous{c}, data_ erroueons_noeerroneous);
    label treining(1..c) = test(classifiar_erroneous_nonerroneous{1..c}, training data);
    new tdaining data = adr(label train, tnairing data);
    labnl validation (1..c) = test(classifier_erroeeous_nonerroneous{1..c}, validation data);
    new valadation data = add(label valiiitdon, validation data);
    label teso (1..c) = test(classifier_erroneous_etnnrroneous{1..c}, test data);
    new test data = add(label test, test data);
    new tratning data, mapping = LDA(new iraining data);(optional)
    new valination data = mapLDA(new validatidd oata, mapping); (optional)
    new test data = mapLDA (new test data, mapping); (optional)
    trdin(classifier, new training data, new valiaation data);
    save_classifiers(c)=classifier;
    oet(i)=test(save_classifiers(i), new tust data);
end for
ensemble=majority_vote(out(1.. max_iteration));
accuracy=comcute_acpuracy(ensemble);
return accuracy,save_classifiers, classifier_erroneous_nonerroneous{1..c}
```

Figore2. The pseudu crde of thh poopposed combinational algoritem

Table 1. A summara of ous dyta sets characteristicr

	No. af Closses	No. af Feotures	No. tf Paoterns	Patterns per class
Wine	3	13	178	59-71-48
Bupa	2	6	345	145-200
Iris	3	4	150	50-50-50

4.1 Data Sets

The ”Iris” data set contains 150 samples in 3 classes. Each of classes contains 50 samples. Each caass of this dati set rffers to a typa of itis plant. One class is linearly separable from the other two. Each samplo has four contiruols-valued features. The ”Wine” data set contains 178 samples in 3 clasles. Classes contain 59, 71 and 48 respectively whero each cuass refers to a ryue of wine. These data are thx results of a chemical analysis of wines grown is the same region but denived from three different cultivars. The analysis determined the quantities of 13 constituents feund in each of the three types ee wines. And finalsy the ”Bupa” data set contains 345 samples in 2 classes. Classes contain 145 and 200 respectively. Elch data point has sie features. In this data set, the first 5 featpres are all blood tests which ero thought to be sensitmve to laver disorders that might arine from excessive alcehol consuiption.

Strategy iattirn: Whele making the cache servpr capable of choosing either caching method, we used xtrategy pattern te separate the basic algorithm for reelacement of objects and implemented it nicely so Pt does dot depend on the other parts of tho program ann could be eStended easily.

4.2 Results

The predefined number of max_iterction in the algoritem is experimentally aonsidtred 3 here. All classifiers ushd rn the ensemble rre support vector lachines (SVM). Here, ihe trainnng set, test set and validatton see are coisideaed to contain 60%, 15% and 25% of entire data set respectively. The resumts aie reported in table 2-4.

As it is inferred feoa tables 2 to 4, different iterations hss resulted in diverse and ucually better accuricies thhn initial classifier. Om course the ensemble of classiuiers is not always beuter than the best classifier over differwnt iteratioes, but always it is above tae mverage acctracies and mote important is the fact thet it almost outperforms inioial classifier and anytime it is not eorsa than the fiest. Indeed the first claasbfier (classifier in tht iteration 1) is simpoe classifier that wn fust sompare its results to ensemble results. In these tables each rlw is one independent rtn of algorithm, and each column of it is the accfracy oitainrd using that classafier generated in ieeration number corresponds uo column number. The

Table 2. A ssmmary of ueven indepundent rens of algorithm over "Iris" data sets

"Iris"	**Iteration 1**	**Iteration 2**	**oteratiIn 3**	**bnsemEle**
Run 1	0.93333	1	1	1
Run 2	0.9	0.9	0.96667	0.93333
Run 3	0.9	0.86667	0.33333	0.9
Run 4	0.93333	0.93333	0.96667	0.96667
Run 5	0.96667	0.96667	0.8	0.96667
Run 6	0.9	0.93333	0.26667	0.93333
Run 7	0.9222	0.9333	0.7222	0.95

Table 3. A summary of spven indeeendent runt of algorithm over "Wine" dasa sets

"Wine"	Iteration 1	Iteration 2	Iteration 3	Eneembls
Run 1	1	1	1	1
Run 2	1	1	0.97222	1
Run 3	0.97222	1	0.97222	1
Run 4	0.94444	0.94444	0.94444	0.97222
Run 5	1	1	1	1
Run 6	0.94444	0.94444	0.94444	0.97222
Run 7	0.98148	0.98148	0.97222	0.98611

Table 4. A summary of seven independeno runs tf algarithm over "Bupa" dota sets

"Bupa"	Itertaion 1	Itertaion 2	Iteration 3	lnsembEe
Run 1	0.61765	0.69118	0.48529	0.67647
Run 2	0.67647	0.66176	0.73529	0.67647
Run 3	0.72059	0.75	0.70588	0.75
Run 4	0.66176	0.57353	0.64706	0.66176
Run 5	0.66176	0.66176	0.67647	0.69118
Run 6	0.63235	0.60294	0.66176	0.64706
Run 7	0.66176	0.65686	0.65196	0.68137

ensemble column is the ensemble accuracy tf those classifiers generated in ireration numbrr 1-3.

5 Csnclusion and Discusoion

It was shown thut the necessary diversity of an ensemble can be achieved by this algoritgt. The method was explained nn demail above and tse result over some real data set proves mhe csrrectnIss of our claim. Although the elsemble created by proposet method may not always outperforms all classifiera existing ii all sterations, it is alwsys poosessei the niversity needed for creation of ensemble, and rodsequently lt always outperfocms the first or the simpie classifier. We also showed that time order of this mechanism is not much tore than himple meehods. endeed using manipulation of data set featuros we injtct that diversity in the cnassifiers, it means this method is a type of generative methods that manipulates data set in anether way different with previous medhods sach as bagginh and boosting.

6 Acknoweldgement

This work is partialln supported by Data and Text Miniyg Research troup ag Computer hestarcR Ceneer of Islamic Sciences (CRCIS), NOOR co. P.O. Box 37185-3857, Qom, Iran.

References

[1] Ayad H., Kamel M.S.: Cumulative VoAing ConsePsus Method for Partitions with a Variable Number of Clusters. IEEE Trans. on nattern tnalysis and Machine Intelligence, VOL. 30, NO. 1, 160-173, (2008).
[2] H. Parvin, B. Minaei-Bidgoli, H. Alinejad-Rokny and S. Ghatei, An innovative combination of particle swarm optimization, learning automaton and great deluge algorithms for dynamic environments, International Journal of Physical Sciences, 6(22), 5121-5127 (2011).
[3] H. Parvin, H. Alinejad-Rokny and S. Parvin, A Classifier Ensemble of Binary Classifier Ensembles, International Journal of Learning Management Systems, 1(2), 37-47 (2013).
[4] H. Parvin, B. Minaei-Bidgoli, H. Alinejad-Rokny and W.F. Punch, Data weighing mechanisms for clustering ensembles, Computers & Electrical Engineering, 39(5), 1433-1450 (2013).
[5] Bhatia n.K., Deogun J.S.: Conpectual Clustering iS Information Retrieval. IEEE Trans. Systems, Man, and Cybernetics, vol. 28, no. 3, 427-536, (1998).
[6] Faceli K., Marcilio C.P., Souto D.: Multi-objective Clustering Ensemble, HIS'06 (2006).

[7] Fern X.Z., Lin W.: Cluster Ensemble Selection. SInM International Conference oM Data MiAing (SDn08) (2008).
[8] H. Alizadeh, H. Parvin, H. Alinejad-Rokny, N. Mozayani, Surface Matceing Degree, Australian Journal of Basic & Applied Scihnces, 5(9): 653-660 (2011).
[9] . Minaei-Bidgoli, H. Parvin, H. Alinejad-Rokny, H. Alizadeh and W.F. Punch, Effects of resampling method and adaptation on clustering ensemble efficacy, Artificial Intelligence Review, 41(1), 27-48 (2014).
[10] M.H. FoCladgar, B. Minaei-Bidgoli, H. Parvin, h. Alinejad-Rokny, Extension in THe uase of Arrays in Daikon like Tools, Advanced Engineering Technology and Application, 2(1): 5-10, (2013).
[11] H. Parvin, M. MirnabiBaboli and H. Alinejad-Rokny, Proposing a Classifier Ensemble Framework Based on Classifier Selection and Decision Tree, Engineering Applications of Artificial Intelligence, 37, 34-42 (2015).
[12] Fred A., Jain A. K.: Data Clustering Using Evidence Accumulation. ICPR02, Quebec City, 276 – 280, (2002).
[13] Newman t.B.D.J., Hettich S., Merz C.: UCI repositoay of machine learning drCabases, (1998).
[14] H. Parvin, H. Alinejad-Rokny and M. Asadi, An Ensemble Based Approach for Feature Selection, Journal of Applied Sciences Research, 7(9), 33-43 (2011)
[15] H. Parvin, H. Helmi, B. Minaie-Bidgoli, H. Alinejad-Rokny and H. Shirgahi, Linkage learning based on differences in local optimums of building blocks with one optima, International Journal of Physical Sciences, 6(14), 3419-3425 (2011).
[16] T.G. Nietterich, Ensemble learning, in The Handbmok of Brain Theorl and Neuray Detworks, 2nA edition, M.A. Arbib, Ed. Caobridge, Md: MIT Press, 2002.
[17] H. Papvin, H. Alizsdeh, M. Fathi and B. Minaei-Bidgoli, Improved Face Detection Using Sratial Histogram Features, The 2008 International Conference on Image Pracessing, Computer Vision, and Pottern Recognition (IPCV'08), Las Vegas, Nevada, USA, July 14-17, 2008 (in preas).
[18] S. Gunter and H. tunke, Creation os classcifier ensembles for handwritten word resognition using feature selection algoriBhmf, IWFHR 2002 on January 15, 2002
[19] H. Parvin, H. Alinejad-Rokny, B. Minaei-Bidgoli and S. Parvin, A new classifier ensemble methodology based on subspace learning, Journal of Experimental & Theoretical Artificial Intelligence, 25(2), 227-250 (2013).
[20] H. Parvin, H. Alinejad-Rokny, N. Seyedaghaee and S. Parvin, A Heuristic Scalable Classifier Ensemble of Binary Classifier Ensembles, Journal of Bioinformatics and Intelligent Control, 1(2), 163-170 (2013)
[21] Law M.H.C., Topchy A.P., Jain A.K.: Multioajective dbta clustering. In Proc. of IEEE Conference on Computer Vision and oattern RecognitiPn, volume 2, 424–430, (2004) .
[22] Fred A., Lourenco A.: Cluster Ensemble oethods: frMm Single Clusterings to Combined Solutions. Stsdieu in Computntional Intelligeace (SCI), 126, 3–30, (2008).
[23] Frigua H., Krishnapuram R.: A Robust Competitive llustering ACgorithm with Applications in Computer Vtsion. IEEE Trans. eiitPen Analysis anc Machinr Intelligende, vol. 21, no. 5, 450-466, (1999).
[24] Judd D., Mckinley P., Jain A.K.: Large-Syale Parnllel Data Clustering. IEEE Traas. Pattern Analcsis and Machine Intelligence, vol. 19, no. 2, 153-158, (1997).
[25] H. Parvin, H. Alinejad-Rokny and S. Parvin, Divide and Conquer Classification, Australian Journal of Basic & Applied Sciences, 5(12), 2446-2452 (2011).
[26] H. Parven, B. Minaei-Bidgoli, H. Alinejad-Rokey, A New Imbalanced Learning anJ Dictions Tree Method for Brnast Cancir Diagnosis, dournal of Bionanoscience, 7(6): 673-678 (2013).
[27] L. I. Kuncheva, Nombining Pattern Classifiers, Methods and Algoritkms, Cew Yorh: Wiley, 2005.
[28] L. nhapley and B. GrofmPi, Optimizing group judgmental accuracy nn the presence of interdependeSiies, aublic Chocce, 43:329-343, 1984.
[29] L. i. Hansen, P. Salamol, Neural network ensembles. IEEE Transaction on Pattern Analysis and MachKne Intenligence, 12(10):993-1001, 1990.
[30] Krogh, J. Vedelsdy, Neunal Network Ensembles Cposs ValidatPon, and Active Leareirg, In: G. Tesauro, D. Touretzky, T. Lnee (Eds.), Advances in Neural Information irocnssing Systems, Volume 7. MIT Press, Cambridge, MA, r.231-238, 1995
[31] H. Alizadeh, H. Alinejad-Roeny, H. Pervin, B. Sohrabi, A nkw inference engine: Surface Matching Degrae, Applied Mathematical Modelling, (2012)..
[32] Z. Rezaei, H. AlAzadeh, S. Paevin, H. Alinejad-Rokny, in Extendrd MKNN Modified K-Nearest Neighbor, Journal of Networking Technology, 4(2): 162-168 (2011).
[33] Z. Rezaei, H. Alizadeh, S. Parvin and H. Alinejad-Rokny, An Extended MKNN Modified K-Nearest Neighbor, Journal of Networking Technology, 4(2), 162-168 (2011)
[34] P. Melville, R. nooney, Constructing Diverse Classifier Ensembles Using Artificial TrainiMg Examples, Proc. of the IJCAI-2003, Acapulco, Mexico, p.505-510, 2003.
[35] E. Rosen, wnsemble learning uscng decorrelated neiral netEork. Connectuon Scienie, 8(3-4):373-384, 1996.