Analysis of Bagging Ensembles of Fuzzy Models for Premises Valuation

Marek Krzystanek[1], Tadeusz Lasota[2], Zbigniew Telec[1], and Bogdan Trawiński[1]

[1] Wrocław University of Technology, Institute of Informatics,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
[2] Wrocław University of Environmental and Life Sciences, Dept. of Spatial Management
Ul. Norwida 25/27, 50-375 Wrocław, Poland
marek.krzystanek@gmail.com, tadeusz.lasota@wp.pl,
zbigniew.telec@pwr.wroc.pl, bogdan.trawinski@pwr.wroc.pl

Abstract. The investigation of 16 fuzzy algorithms implemented in data mining system KEEL from the point of view of their usefulness to create bagging ensemble models to assist with real estate appraisal were presented in the paper. All the experiments were conducted with a real-world dataset derived from a cadastral system and registry of real estate transactions. The results showed there were significant differences in accuracy between individual algorithms. The analysis of measures of error diversity revealed that only the highest values of an average pairwise correlation of outputs were a profitable criterion for the selection of ensemble members.

Keywords: genetic fuzzy systems, bagging, real estate appraisal, KEEL.

1 Introduction

Ensemble methods have drawn considerable attention of many researchers for the last decade. Ensembles combining diverse machine learning models have been theoretically and empirically proved to ensure significantly better performance than their single original models. Although many multiple model creation techniques have been developed [28], according to [27] five commonly used groups of them can be distinguished, namely bagging [4], boosting [29], AdaBoost [13], stacked generalization [30], and mixture of experts [17].

Bagging, which originates from the term of bootstrap aggregating, belongs to the most intuitive and simplest ensemble algorithms providing a good performance. The diversity of regressors is obtained by using bootstrapped replicas of the training data. That is, different training data subsets are randomly drawn with replacement from the original training set. So obtained training data subsets, called also bags, are used then to train different regression models. Finally, individual regressors are combined through an algebraic expression, such as minimum, maximum, sum, mean, product, median, etc. [27]. Bagging has been employed to regression trees [3], Gaussian process [9], and neural networks [16]. There are also a few works on applying ensemble fuzzy systems to solve classification [6] and prediction problems [18].

N.T. Nguyen, M.T. Le, and J. Świątek (Eds.): ACIIDS 2010, Part II, LNAI 5991, pp. 330–339, 2010.
© Springer-Verlag Berlin Heidelberg 2010

Many articles consider the bias-variance-covariance decomposition of error functions and the problem of diversity-accuracy trade off as the basis of the construction of hybrid ensembles [5], [7], [21].

Very often the whole generated ensemble turns out to be not an optimal solution, e.g. due to bias in the learners. Numerous approaches to select classifiers/regressors to compose ensembles were proposed i.e. genetic algorithms [10], [31], ensemble pruning [15], [26], instance selection [14], overproduce-and-choose strategy [12], negative correlation learning [25], and many others.

In our previous works [19], [20], [22] we tested different machine learning algorithms, among others genetic fuzzy systems trying to select the most appropriate ones to build data driven models for real estate appraisal using MATLAB and KEEL. We began also our study on the bagging approach, to investigate whether it could lead to the improvement of the accuracy machine learning regression models devoted to assist with real estate appraisals [23], [24]. In this paper we present the extension of our previous experiments with bagging ensembles to 16 fuzzy algorithms implemented in KEEL. We aimed to consider if this class of algorithms fits to design hybrid ensembles.

Actual data used to generate and learn appraisal models came from the cadastral system and the registry of real estate transactions referring to residential premises sold in one of big Polish cities at market prices within two years 2001 and 2002. They constituted an original dataset of 1098 instances of sales/purchase transactions. Four attributes were pointed out as price drivers: usable area of premises, floor on which premises were located, year of building construction, number of storeys in the building, in turn, price of premises was the output variable.

The concept of a data driven models for premises valuation, presented in the paper, was developed based on the sales comparison method. The appraiser accesses the system through the internet and input the values of the attributes of the premises being evaluated into the system, which calculates the output using a given model. The final result, that is a suggested value of the property, is sent back to the appraiser.

2 Plan of Experiments

The main goal of the investigations was to carry out the comparative analysis of 16 fuzzy algorithms employed to create ensemble models for premises property valuation. For this purpose KEEL, a non-commercial Java software tool designed to assess evolutionary algorithms for data mining problems, was used [2]. KEEL algorithms used in study are listed in Table 1, and details of the algorithms and references to source articles can be found on KEEL web site: www.keel.es. Most algorithms were evolutionary fuzzy ones. Two families of algorithms were distinguished during the analysis of results achieved: MOGUL - a methodology to obtain genetic fuzzy rule-based systems under the Iterative Rule Learning approach and the Wang-Mendel algorithm tuned by means of evolutionary post-processing algorithms.

Table 1. Fuzzy algorithms used in study

Alg.	KEEL name	Description
COR	Regr-COR_GA	Genetic fuzzy rule learning, COR algorithm inducing cooperation among rules
FRS	Regr-FRSBM	Fuzzy and random sets based modeling
FGP	Regr-Fuzzy-GAP	Fuzzy rule learning, grammar-based GP algorithm
PFC	Regr-Fuzzy-P_FCS1	Pittsburgh fuzzy classifier system #1
FSA	Regr-Fuzzy-SAP	Fuzzy rule learning, grammar GP based operators and simulated annealing based algorithm
SEF	Regr-Fuzzy-SEFC	Symbiotic evolution based fuzzy controller design method
THR	Regr-Thrift	Genetic fuzzy rule learning, Thrift algorithm
IRL	Regr-Fuzzy-MOGUL-IRL	Iterative rule learning of descriptive Mamdani rules
IHC	Regr-Fuzzy-MOGUL-IRLHC	Iterative rule learning of Mamdani rules - high constrained approach
ISC	Regr-Fuzzy-MOGUL-IRLSC	Iterative rule learning of Mamdani rules - small constrained approach
ITS	Regr-Fuzzy-MOGUL-TSK	Local evolutionary learning of TSK fuzzy rule based system
W-M	Regr-Fuzzy-WM	Fuzzy rule learning, Wang-Mendel algorithm
WAG	Regr-Fuzzy-WM & Post-A-G-Tuning-FRBSs	Wang-Mendel algorithm tuned using approxi-mative genetic tuning of FRBSs
WGG	Regr-Fuzzy-WM & Post-G-G-Tuning-FRBSs	Wang-Mendel algorithm tuned using global genetic tuning of the fuzzy partition of linguistic FRBSs
WGS	Regr-Fuzzy-WM & Post-G-S-Weight-RRBS	Wang-Mendel algorithm tuned using genetic selection of rules and rule weight tuning
WGT	Regr-Fuzzy-WM & Post-G-T-Weights-FRBSs	Wang-Mendel algorithm tuned using genetic tuning of FRBS weights

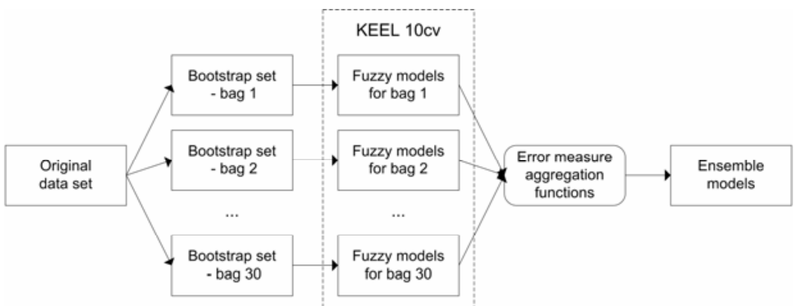

Fig. 1. Schema of bagging ensemble model development

Schema of the experiments is depicted in Fig. 1. On the basis of the original data set 30 bootstrap replicates (bags) of the cardinality equal to the original dataset were created. The bags were then used to generate models employing each of 16 above mentioned algorithms, and as the result 480 models for individual bags were obtained. For the reference 16 base models for individual algorithms using the original dataset were also produced. During the pre-processing phase normalization of data was performed using the min-max approach. As the accuracy measure the mean square error (MSE) was applied. All models were generated using 10-fold cross validation (10cv). As aggregation functions for combining the ensemble members' individual outputs simple averages were used.

3 Results of Experiments

At the first step the performance of single models built over the original data set was compared (see Fig. 2). It can be observed that MSE differentiate the algorithms substantially, ranging from 0.0025 for SEF and FRS to 0.060 for W-M and 0.0078 for PFC. The MSE values of individual base models were then used as reference levels when analyzing the bagging ensembles.

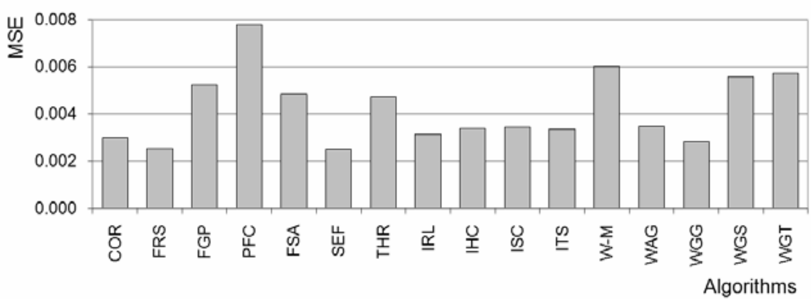

Fig. 2. Comparison of MSE for models built using the original data set

Having the values of MSE for individual algorithms over 30 bags, the Friedman and Iman-Davenport non-parametric statistical tests were performed in respect of average ranks, which use χ^2 and F statistics, respectively [11]. The calculated values of these statistics were 391.85 and 195.42, respectively, whereas the critical values at α=0.05 are $\chi^2(15)$=27.49 and F(15,435)=1.69, so the null-hypothesis were rejected. Average ranks of individual algorithms are shown in Table 2, where the lower rank value the better algorithm.

Table 2. Average rank positions of individual algorithms over all 30 bags

Rank	Alg.	Rank	Alg.	Rank	Alg.	Rank	Alg.
2.30	ITS	4.40	SEF	7.93	WAG	12.80	FSA
4.03	IRL	5.37	IHC	10.10	THR	13.63	W-M
4.13	WGG	5.57	ISC	11.50	WGS	14.67	PFC
4.17	FRS	7.23	COR	12.37	WGT	15.80	FGA

Figures 3 a)-p) show the differences in accuracy of bagging ensembles built by means of individual evolutionary fuzzy algorithms over from 2 to 30 successive bags. The greater distance between the horizontal lines reflecting the MSE of original models and the tops of bars, which represent the MSE of a given ensemble, the bigger error reduction provided by that ensemble. Moreover, due to the same scale of all charts it is possible to assess visually the differences between respective models. So you can notice that PFC, THR, and all algorithms belonging to the Iterative Rule Learning and Wang-Mendel families ensure the best error reduction. Of course, the relative values of MSE should be taken into account, because in fact for PFC, THR, W-M, WGS, and WGT ensembles lead to the error reduction, but their accuracy remains still at a rather low level. Maximal percentage MSE reduction of bagged ensembles compared to MSE provided by the original models is shown in Table 3.

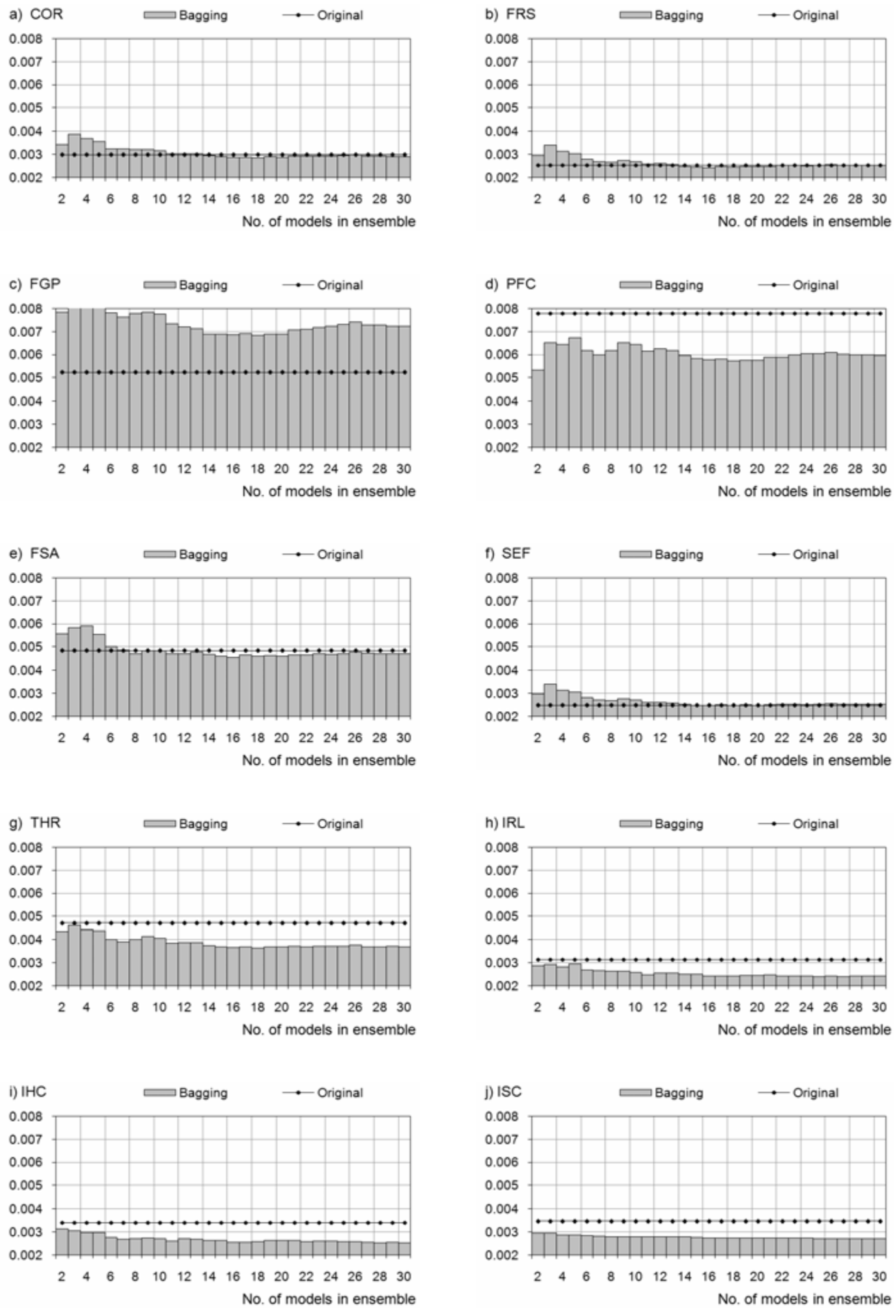

Fig. 3 a)-j). Performance of bagged ensembles compared with original model for individual algorithms in terms of MSE

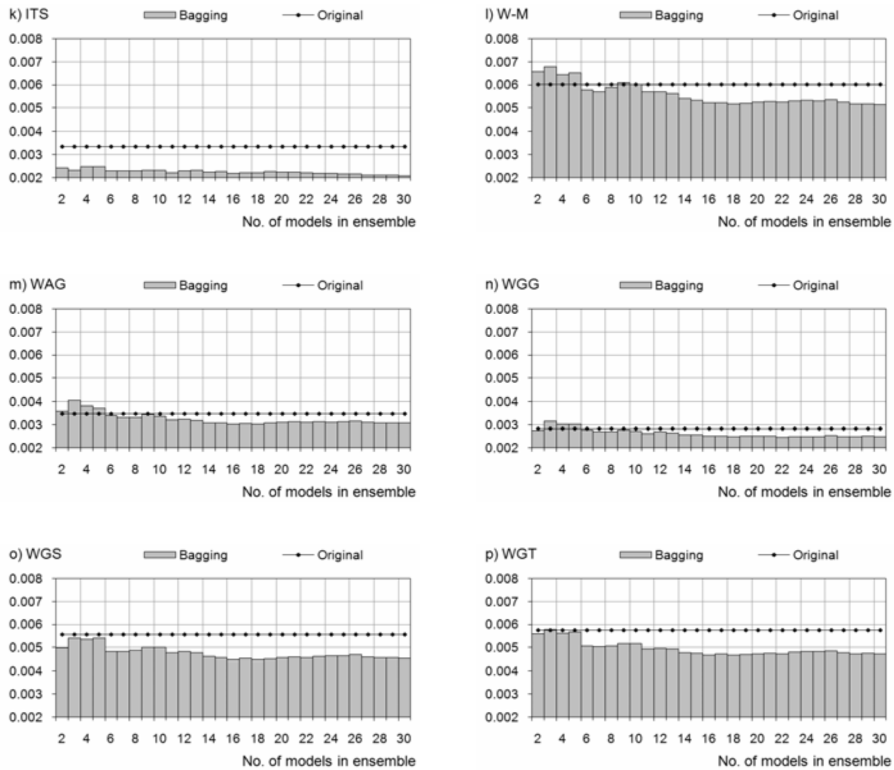

Fig. 3 k)-p). Performance of bagged ensembles compared with original model for individual algorithms in terms of MSE

Table 3. Maximal error reduction of bagged ensembles compared to single models for individual algorithms

Alg.	No. of bags	MSE	% red	Alg.	No. of bags	MSE	% red
COR	18	0.00286	4.39%	IRL	27	0.00240	23.56%
FRS	16	0.00243	3.89%	IHC	30	0.00252	25.75%
FGP	18	0.00685	-30.81%	ISC	25	0.00272	21.70%
PFC	2	0.00534	31.45%	ITS	30	0.00209	37.66%
FSA	16	0.00454	6.30%	W-M	30	0.00514	14.66%
SEF	16	0.00247	0.62%	WAG	18	0.00302	13.21%
THR	18	0.00364	22.63%	WGG	22	0.00245	13.18%

These observations encourage us to seek the optimal subset of bags composing an ensemble with the best performance. The simplest method is to take a number of models which reveal the lowest accuracy error. Figures 4, 5, 6 and 7 depict the ensembles comprising from 10 to 25 best models generated with the best representatives of FRS, SEF, ITS, and WGG. The gain in accuracy can be easily noticed. In Table 4 three best algorithms which provide the lowest MSE for each bag, and two worst ones were placed in five leftmost columns. So, still better results might provide hybrid ensembles comprising the best models created by different algorithms.

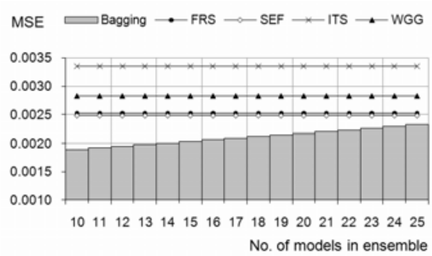

Fig. 4. MSE of ensembles comprising best of FRS models

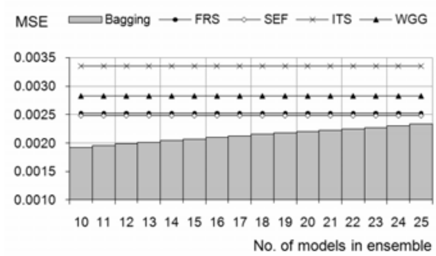

Fig. 5. MSE of ensembles comprising best of SEF models

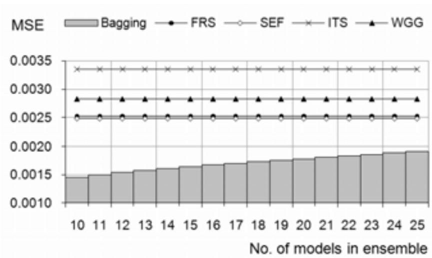

Fig. 6. MSE of ensembles comprising best of ITS models

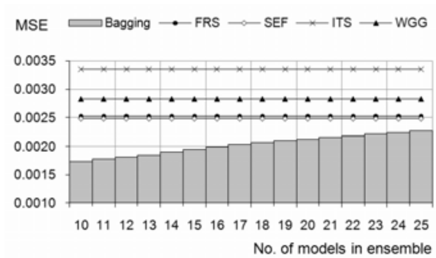

Fig. 7. MSE of ensembles comprising best of WGG models

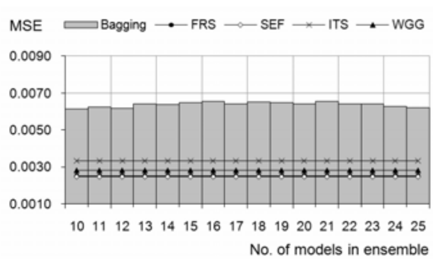

Fig. 8. MSE of ensembles comprising models with the lowest CV

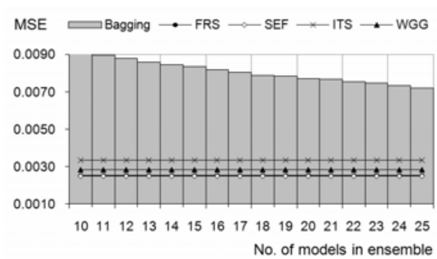

Fig. 9. MSE of ensembles comprising models with the highest CV

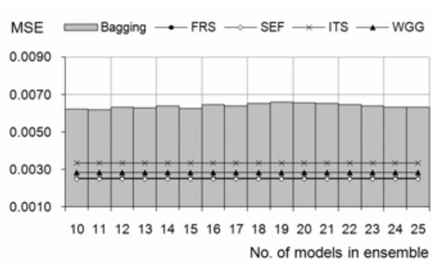

Fig. 10. MSE of ensembles comprising models with the lowest APC

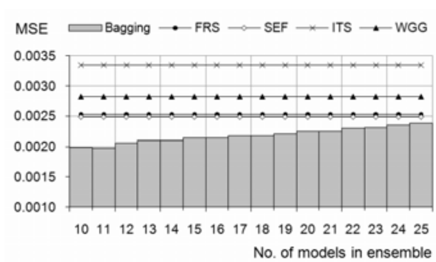

Fig. 11. MSE of ensembles comprising models with the highest APC

Table 4. Fuzzy models holding highest and lowest rank positions according to accuracy and diversity criteria for individual bags

Bag	Best MSE			Worst MSE		CV		APC	
	1st	2nd	3rd	15th	16th	Lwst	Hgst	Lwst	Hgst
1	ITS	IHC	FRS	PFC	FGP	PFC	W-M	PFC	SEF
2	ITS	IRL	FRS	FWM	FGP	PFC	FGA	PFC	WGG
3	ITS	ISC	IHC	PFC	FGP	PFC	FGA	PFC	IRL
4	FRS	SEF	IRL	PFC	FGP	PFC	W-M	PFC	WGG
5	ITS	FRS	SEF	PFC	FGP	PFC	FGA	PFC	WGG
6	ITS	IRL	WGG	FSA	PFC	PFC	ITS	PFC	WGG
7	FRS	WGG	SEF	FWM	FGP	PFC	W-M	PFC	**WGG**
8	FRS	ITS	SEF	PFC	FGP	PFC	FGA	PFC	WGG
9	ITS	IRL	IHC	FGP	PFC	PFC	W-M	PFC	IRL
10	ITS	IRL	FRS	PFC	FGP	PFC	W-M	PFC	WGG
11	ITS	WGG	IRL	FGP	FSA	PFC	WGT	PFC	WGG
12	SEF	FRS	COR	FWM	PFC	PFC	FGA	PFC	WGG
13	FRS	SEF	WGG	FSA	FGP	PFC	W-M	PFC	WGG
14	ITS	IRL	WGG	FSA	FGP	PFC	WGT	FGA	WGG
15	FRS	SEF	COR	WGT	FGP	PFC	FGA	PFC	WGG
16	ITS	IRL	IHC	PFC	FGP	PFC	FGA	PFC	WGG
17	IHC	IRL	ITS	PFC	FGP	PFC	FGA	PFC	IRL
18	SEF	FRS	WGG	PFC	FGP	PFC	W-M	PFC	WGG
19	SEF	FRS	WGG	PFC	FGP	PFC	FGA	PFC	WGG
20	ITS	WGG	FRS	FWM	FGP	PFC	W-M	ITS	WGG
21	ITS	WGG	SEF	PFC	FGP	PFC	FGA	PFC	WGG
22	WGG	IRL	ITS	PFC	FGP	PFC	FGA	PFC	WGG
23	ITS	IRL	ISC	PFC	FGP	PFC	W-M	PFC	WGG
24	SEF	FRS	ITS	PFC	FGP	PFC	W-M	PFC	IRL
25	ITS	IRL	ISC	PFC	FGP	PFC	FGA	PFC	WGG
26	ITS	IHC	ISC	PFC	FGP	PFC	FGA	PFC	WGG
27	ITS	WGG	IRL	PFC	FGP	PFC	WGT	PFC	WGG
28	IHC	ITS	WAG	PFC	FGP	PFC	WGS	PFC	WGS
29	ITS	FRS	SEF	PFC	FGP	PFC	W-M	PFC	ITS
30	ITS	ISC	IHC	PFC	FGP	PFC	FGA	PFC	WGG

Inspired by some works on the application of diversity of errors to construct ensembles [1], we tried to check the usability of two methods which measure the diversity of regressors' outputs regardless of whether they are correct or incorrect. These were the coefficient of variability (CV) measured for each instance for individual algorithms, hoping the higher CV the bigger error reduction could be obtained. But it was not the case. In Table 4 the algorithms with the lowest and the highest CV for each bag are enumerated, but Figures 8 and 9 indicate there is no gain in performance in both cases. Another diversity measure we tested was the average pairwise correlation (APC) of the results produced by algorithms over individual bags. In Table 4 the algorithms with the lowest and the highest APC for each bag are listed, and the performance was depicted in Figures 10 and 11. Ensembles comprising models of the lowest APC also did not lead to the improvement of accuracy. Only the highest APC turned to be beneficial and ensembles composed of those models provided noticeable error reduction.

4 Conclusions and Future Work

Sixteen fuzzy algorithms implemented in data mining system KEEL were tested from the point of view of their usefulness to create bagging ensemble models to assist with real estate appraisal. All the experiments were conducted with a real-world data set derived from cadastral system and the registry of real estate transactions. 30 bootstrap replicates (bags) were created by random drawing with replacement the elements contained in the original dataset of the cardinality equal to 1098. The bags were then used to generate models employing each of 16 above mentioned algorithms, and as the result 480 models for individual bags were obtained.

The results in terms of MSE as the accuracy measure showed there were significant differences between individual algorithms. The best performance in terms of both accuracy and error reduction rate revealed algorithms FRS, WGG, and whole family of Iterative Rule Learning. Further considerations led to the conclusion it was possible to construct hybrid bagging ensembles which would provide still better output. The analysis of two measures of error diversity showed that only the highest values of an average pairwise correlation of outputs were a valuable criterion for the selection of ensemble members.

Further research is planned to construct optimal hybrid ensembles of real estate appraisal models with the use new sets of actual data and with the consideration of time impact on the prices of land and premises. Moreover, an open problem remains the stability of such ensembles, therefore further investigations are needed.

Acknowledgments. This paper was supported by Ministry of Science and Higher Education of Poland under grant no. N N519 407437.

References

1. Aksela, M., Laaksonen, J.: Using diversity of errors for selecting members of a committee classifier. Pattern Recognition 39, 608–623 (2006)
2. Alcalá-Fdez, J., et al.: KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems. Soft Computing 13(3), 307–318 (2009)
3. Banfield, R.E., et al.: A Comparison of Decision Tree Ensemble Creation Techniques. IEEE Trans. on Pattern Analysis and Machine Intelligence 29(1), 173–180 (2007)
4. Breiman, L.: Bagging Predictors. Machine Learning 24(2), 123–140 (1996)
5. Brown, G., Wyatt, J., Harris, R., Yao, X.: Diversity Creation Methods: A Survey and Categorisation. Journal of Information Fusion 6(1), 5–20 (2005)
6. Canul-Reich, J., Shoemaker, L., Hall, L.O.: Ensembles of Fuzzy Classifiers. In: Proc. IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2007, pp. 1–6 (2007)
7. Chandra, A., Yao, X.: Evolving hybrid ensembles of learning machines for better generalisation. Neurocomputing 69, 686–700 (2006)
8. Chawla, N.V., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P.: Learning Ensembles From Bites: A Scalable and Accurate Approach. Journal of Machine Learning Research 5, 421–451 (2004)
9. Chen, T., Ren, J.: Bagging for Gaussian Process Regression. Neurocomputing 72(7-9), 1605–1610 (2009)
10. Cordón, O., Quirin, A.: Comparing Two Genetic Overproduce-and-choose Strategies for Fuzzy Rule-based Multiclassification Systems Generated by Bagging and Mutual Information-based Feature Selection. Int. J. Hybrid Intelligent Systems (2009) (in press)
11. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)

12. Dos Santos, E.M., Sabourin, R., Maupin, P.: A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. Pattern Recognition 41(10), 2993–3009 (2008)
13. Freund, Y., Schapire, R.E.: Decision-theoretic generalization of on-line learning and an application to boosting. J. Computer and System Sciences 55(1), 119–139 (1997)
14. García-Pedrajas, N.: Constructing Ensembles of Classifiers by Means of Weighted Instance Selection. IEEE Transactions on Neural Networks 20(2), 258–277 (2009)
15. Hernandez-Lobato, D., Martinez-Munoz, G., Suarez, A.: Pruning in ordered regression bagging ensembles. In: Yen, G.G. (ed.) Proceedings of the IEEE World Congress on Computational Intelligence, pp. 1266–1273 (2006)
16. Islam, M.M., et al.: Bagging and Boosting Negatively Correlated Neural Networks. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cyb. 38(3), 771–784 (2008)
17. Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E.: Adaptive mixtures of local experts. Neural Computation 3, 79–87 (1991)
18. Kim, D.: Improving the Fuzzy System Performance by Fuzzy System Ensemble. Fuzzy Sets and Systems 98(1), 43–56 (1998)
19. Król, D., Lasota, T., Trawiński, B., Trawiński, K.: Investigation of Evolutionary Optimization Methods of TSK Fuzzy Model for Real Estate Appraisal. International Journal of Hybrid Intelligent Systems 5(3), 111–128 (2008)
20. Krzystanek, M., Lasota, T., Trawiński, B.: Comparative Analysis of Evolutionary Fuzzy Models for Premises Valuation Using KEEL. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCCI 2009. LNCS, vol. 5796, pp. 838–849. Springer, Heidelberg (2009)
21. Kuncheva, L.I., Whitaker, C.J.: Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. Machine Learning 51, 181–207 (2003)
22. Lasota, T., Mazurkiewicz, J., Trawiński, B., Trawiński, K.: Comparison of Data Driven Models for the Validation of Residential Premises Using KEEL. International Journal of Hybrid Intelligent Systems (2009) (in press)
23. Lasota, T., Telec, Z., Trawiński, B., Trawiński, K.: A Multi-agent System to Assist with Real Estate Appraisals using Bagging Ensembles. In: Nguyen, N.T., Kowalczyk, R., Chen, S.-M. (eds.) ICCCI 2009. LNCS, vol. 5796, pp. 813–824. Springer, Heidelberg (2009)
24. Lasota, T., Telec, Z., Trawiński, B., Trawiński, K.: Exploration of Bagging Ensembles Comprising Genetic Fuzzy Models to Assist with Real Estate Appraisals. In: Corchado, E., Yin, H. (eds.) IDEAL 2009. LNCS, vol. 5788, pp. 554–561. Springer, Heidelberg (2009)
25. Liu, Y., Yao, X.: Ensemble learning via negative correlation. Neural Networks 12(10), 1399–1404 (1999)
26. Margineantu, D.D., Dietterich, T.G.: Pruning Adaptive Boosting. In: Proc. 14th Int. Conf. Machine Learning, pp. 211–218 (1997)
27. Polikar, R.: Ensemble Learning. Scholarpedia 4(1), 2776 (2009)
28. Rokach, L.: Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. Comp. Stat. and Data Analysis 53, 4046–4072 (2009)
29. Schapire, R.E.: The Strength of Weak Learnability. Machine Learning 5(2), 197–227 (1990)
30. Wolpert, D.H.: Stacked Generalization. Neural Networks 5(2), 241–259 (1992)
31. Zhou, Z.H., Wu, J., Tang, W.: Ensembling Neural Networks: Many Could Be Better Than All. Artificial Intelligence 137, 239–263 (2002)