Journal of Universal Computer Science, vol. 19, no. 4 (2013), 539-562
submitted: 3/8/12, accepted: 20/2/13, appeared: 28/2/13 © J.UCS

Evolutionary Fuzzy System Ensemble Approach to Model Real Estate Market based on Data Stream Exploration

Bogdan Trawiński
(Wrocław University of Technology, Wrocław, Poland
bogdan.trawinski@pwr.wroc.pl)

Abstract: An approach to predict from a data stream of real estate sales transactions based on ensembles of genetic fuzzy systems was presented. The proposed method relies on incremental expanding an ensemble by models built over successive chunks of a data stream. The output of aged component models produced for current data is updated according to a trend function reflecting the changes of premises prices since the moment of individual model generation or the beginning of the data stream. The impact of different trend functions on the accuracy of single and ensemble fuzzy models was investigated in the paper. Intensive experiments were conducted to evaluate the proposed method using real-world data taken from a dynamically changing real estate market. The statistical analysis of experimental output was made employing the nonparametric methodology designed especially for multiple comparisons including Friedman tests followed by Nemenyi's, Holm's, Shaffer's, and Bergmann-Hommel's post-hoc procedures. The results proved the usefulness of ensemble approach incorporating the correction of individual component model output.

Keywords: Genetic fuzzy systems, Data stream, Sliding windows, Ensembles, Predictive models, Trend functions, Property valuation
Category: H.2.8, I.2.6, I.5.2

1 Introduction

Data stream mining has attracted the attention of many researchers during the last fifteen years. Processing data streams represents a novel challenge because it requires taking into account memory limitations, short processing times, and single scans of arriving data. Many strategies and techniques for mining data streams have been devised. Gaber in his recent overview paper categorizes them into four main groups: two-phase techniques, Hoeffding bound-based, symbolic approximation-based, and granularity-based ones [Gaber, 12]. Much effort is devoted to the issue of concept drift which occurs when data distributions and definitions of target classes change over time [Elwell, 11], [Gama, 05], [Jiang, 09], [Jung, 05], [Maloof, 04], [Sobolewski, 12], [Widmer, 96], [Zliobaite, 09]. Among the instantly growing methods of handling concept drift in data streams Tsymbal distinguishes three basic approaches, namely instance selection, instance weighting, and ensemble learning [Tsymbal, 04], the latter has been systematically overviewed in [Kuncheva, 04], [Minku, 10]. In adaptive ensembles, component models are generated from sequential blocks of training instances. When a new block arrives, models are examined and then discarded or modified based on the results of the evaluation. Several methods have been proposed for that, e.g. accuracy weighted ensembles [Wang, 03] and accuracy updated ensembles [Brzeziński, 11]. In [Bifet, 09a], [Bifet, 09b] Bifet et al. proposed two bagging methods to process concept drift in a data stream: ASHT Bagging using trees of different sizes, and ADWIN Bagging employing a change detector to decide when to discard underperforming ensemble members.

One of the most developed recently learning technologies devoted to dynamic environments have been evolving fuzzy systems [Lughofer, 11a]. Data-driven fuzzy rule based systems (FRBS) are characterized by three important features. Firstly, they are able of approximating any real continuous function on a compact set with an arbitrary accuracy [Castro, 96], [Kosko, 94]. Secondly, they have the capability of knowledge extraction and representation when modelling complex systems in a way that they could be understood by humans [Alonso, 09]. Thirdly, they can be permanently updated on demand based on new incoming samples as is the case for on-line measurements or data streams. The technologies provide such updates with high performance, both in computational times and predictive accuracy. The major representatives of evolving fuzzy approaches are FLEXFIS [Lughofer, 08] and eTS [Angelov, 04] methods. The former incrementally evolves clusters (associated with rules) and performs a recursive adaptation of consequent parameters by using local learning approach. The latter is also an incremental evolving approach based on recursive potential estimation in order to extract the most dense regions in the feature space as cluster centers (rule representatives).

This work is the continuation of our research into application of ensembles of genetic fuzzy systems (*GFSs*) to predict from a data stream of real estate sales transactions reported in two papers during the SUM 2012 and ICCCI 2012 conferences [Trawiński, 12a], [Trawiński, 12b]. The goal of the study presented in this paper was to apply a non-incremental genetic fuzzy systems to build reliable predictive models from a data stream. The approach was inspired by the observation of a real estate market of in one big Polish city in recent years when it experienced a violent growth of residential premises prices. Our method consists in the utilization of aged models to compose ensembles and correction of the output provided by component models by means of trend functions reflecting the changes of prices in the market over time. In this paper we present an analysis of the impact of different trend functions on the accuracy of single and ensemble fuzzy models for residential premises valuation as well as comparison of the accuracy of ensembles created in 20 time points followed by nonparametric tests of statistical significance.

The paper is organized as follows. Section 2 shows the proposed method to predict from a data stream of real estate transactions well as the methods of correcting the output of aged models. Section 3 describes the plan of experiments we conducted. Section 4 presents the results we obtained as well as the thorough analysis of statistical significance of the output. Finally, the paper is concluded in section 5.

2 GFS Ensemble Method to Predict from a Data Stream

2.1 Rationale Behind the Method

The approach based on fuzzy logic is especially suitable for property valuation because professional appraisers are forced to use many, very often inconsistent and imprecise sources of information, and their familiarity with a real estate market and the land where properties are located is frequently incomplete. Moreover, they have to consider various rice drivers and complex interrelation among them. An appraiser should make on-site inspection to estimate qualitative attributes of a given property as well as its neighbourhood. They have also to assess such subjective factors as location attractiveness and current trend and vogue. So, their estimations are to a great extent subjective and are on uncertain knowledge, experience, and intuition rather than on objective data.

So, the appraisers should be supported by automated valuation systems which incorporate data driven models for premises valuation developed employing sales comparison method. The data driven models, considered in the paper, were generated using real-world data on sales transactions taken from a cadastral system and a public registry of real estate transactions. The architecture of the proposed system is shown in Figure 1. 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 as a suggested value of the property is sent back to the appraiser. We explore data-driven fuzzy rule-based systems (*FRBSs*) as a specific data-driven model architecture used in the framework shown in Figure 1, which were recognized to be able of approximating any real continuous function on a compact set with an arbitrary accuracy. Moreover, *FRBSs* allow for knowledge extraction and representation by modelling complex systems in a way understandable by humans. So, the interpretability of fuzzy systems is a characteristic that favours this type of models because it is often required to explain the behaviour of a given real appraisal model.

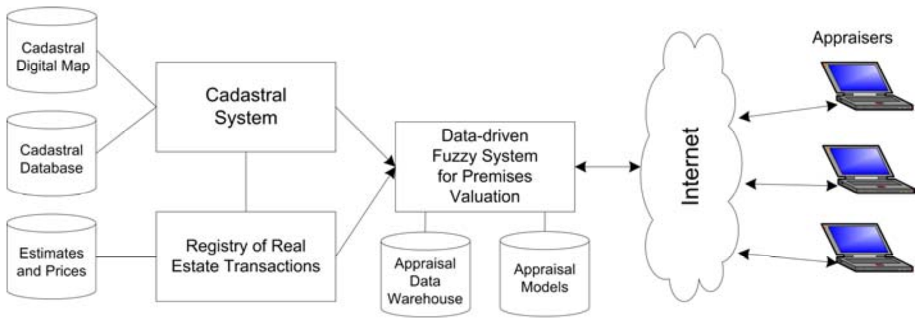

Figure 1: Information systems to assist with real estate appraisals

So far, we have investigated several methods to construct regression models to assist with real estate appraisal based on fuzzy approach: i.e. genetic fuzzy systems as both single models [Król, 08] and ensembles built using various resampling techniques [Kempa, 11], [Lasota, 11b], but in this case the whole datasets had to be available before the process of training models started. All property prices were updated to be uniform in a given point of time. An especially good performance revealed evolving fuzzy models applied to cadastral data [Lasota, 11a], [Lughofer, 11b]. Evolving fuzzy systems are appropriate for modelling the dynamics of real estate market because they can be systematically updated on demand based on new incoming samples and the data of property sales ordered by the transaction date can be treated as a data stream. In this paper we present our attempt to employ evolutionary fuzzy approach to explore data streams to model dynamic real estate market. The problem is not trivial because on the one hand a genetic fuzzy system needs a number of samples to be trained and on the other hand the time window to determine a chunk of training data should be as small as possible to retain the model accuracy at an acceptable level. The processing time in this case is not a decisive issue because property valuation models need not to be updated and/or generated from scratch in an on-line mode.

Our approach is grounded on the observation of a real estate market in one big Polish city with the population of 640,000. The residential premises prices in Poland depend on the form of the ownership of the land on which buildings were erected. For historical reasons the majority of the land in Poland is council-owned or state-owned. The owners of flats lease the land on terms of the so-called perpetual usufruct, and consequently, most flat sales transactions refer to the perpetual usufruct of the land. The prices of flats with the land ownership differ from the ones of flats with the land lease. Moreover, the apartments built after 1996 attain higher prices due to new construction technologies, quality standards, and amenities provided by the developers. Furthermore, apartments constructed in this period were intended mainly for investments and trades which also led to the higher prices. To our study we selected sales transaction data of apartments built before 1997 and where the land was leased on terms of perpetual usufruct. Therefore the dynamics of real estate market concerns more the prices of residential premises rather than other basic attributes of properties such as usable area, number of rooms, floor, number of storeys, etc. Having a real-world dataset referring to residential premises transactions accomplished in the city, which after cleansing counted 5,213 samples, we were able to determine the trend of price changes within 11 years from 1998 to 2008. It was modelled by the polynomial of degree three. The chart illustrating the change trend of average transactional prices per square metre is given in Figure 2 where a time point t_0 and the value of trend function in this point $T(t_0)$ were marked.

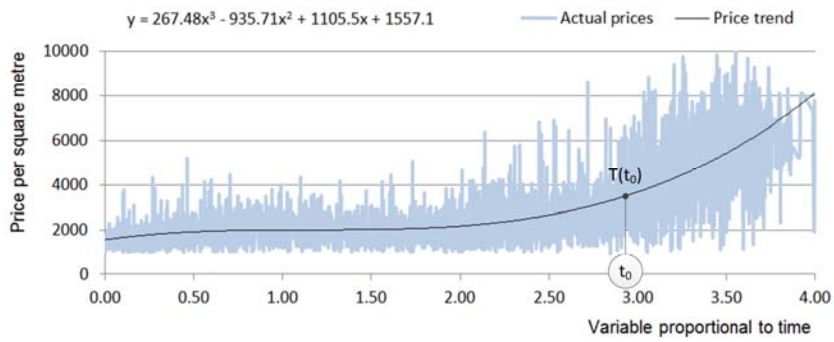

Figure 2: Change trend of average transactional prices per square metre over time

2.2 Description of the proposed method

The outline of the *GFS* ensemble approach to predict from a data stream is depicted in Figure 3. The data stream is partitioned into data chunks according to the periods of a constant length t_c. Each time interval determines the shift of a sliding time window which comprises training data to create *GFS* models. The window is shifted step by step of a period t_s in the course of time. The length of the sliding window t_w is equal to the multiple of t_c so that $t_w=jt_c$, where j=1,2,3,... . The window determines the scope of training data to generate from scratch a property valuation model, in our case GFS_i. It is assumed that the models generated over a given training dataset is valid for the next interval which specifies the scope for a test dataset. Similarly, the interval t_t which delineates a test dataset is equal to the multiple of t_c so that $t_t=kt_c$, where k=1,2,3,.. . The sliding window is shifted step by step of a period t_s in the course of time, and likewise, the interval t_s is equal to the multiple of t_c so that $t_s=lt_c$, where l=1,2,3,… .

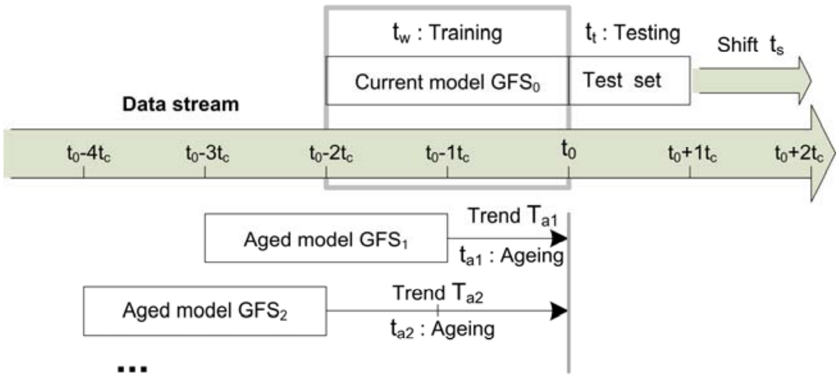

Figure 3: Outline of the proposed ensemble approach to predict from a data stream

We consider in Figure 3 a point of time t_0 at which the current model GFS_0 was built over data that came in between time t_0–$2t_c$ and t_0. The models created earlier, i.e. GFS_1, GFS_2, etc. have aged gradually and in consequence their accuracy has deteriorated. However, they are neither discarded nor restructured but utilized to compose an ensemble so that the current test dataset is applied to each component GFS_i. However, in order to compensate ageing, their output produced for the current test dataset is updated using trend functions $T(t)$. As the functions to model the trends of price changes the polynomials of the degree from one to five were employed, denoted in the rest of the paper by *T1*, *T2*,..,*T5*, respectively. The trends were determined over two time periods: shorter and longer ones. The shorter periods encompassed the length of a sliding window plus model ageing intervals, i.e. t_w plus, t_{ai} for a given aged model GFS_i. The shorter periods are denoted by *Age* in the symbols of methods used in figures and tables presenting the experimental results further on in the paper. In turn, the longer periods took into account all data since the beginning of the stream. Hence, the longer periods are denoted by *Beg* in the symbols of methods used in figures and tables presenting the experimental results further on in the paper.

2.3 Correcting the output of aged models

We proposed two different methods of updating the prices of premises according to the trends of the value changes over time. The first one based on the difference between a price and a trend value in a given time point and we called it the *Delta* method. In turn, the second technique utilized the ratio of the price to the trend value and it was named the *Ratio* method of price correction.

The idea of correcting the results produced by aged models using the *Delta* method is depicted in Figure 4. For the time point $t_{gi}=t_0-t_{ai}$, when a given aged model GFS_i was generated, the value of a trend function $T(t_{gi})$, i.e. average price per square metre, is computed. The price of a given premises, i.e. an instance of a current test dataset, characterised by a feature vector $\boldsymbol{x}$, is predicted by the model GFS_i. Next, the total price is divided by the premises usable area to obtain its price per square metre $P_i(\mathbf{x})$. Then, the deviation of the price from the trend value $\Delta P_i(\mathbf{x})=P_i(\mathbf{x})-T(t_{gi})$ is calculated. The corrected price per square metre of the premises $P_i'(\mathbf{x})$ is worked out by adding this deviation to the trend value in the time point t_0 using the formula $P_i'(\mathbf{x})=\Delta P_i(\mathbf{x})+T(t_0)$, where $T(t_0)$ is the value of a trend function in t_0. Finally, the corrected price per square metre $P'(\mathbf{x})$ is converted into the corrected total price of the premises by multiplying it by the premises usable area. The delta method is denoted by *D* in the symbols of methods used in figures and tables presenting the experimental results further on in the paper.

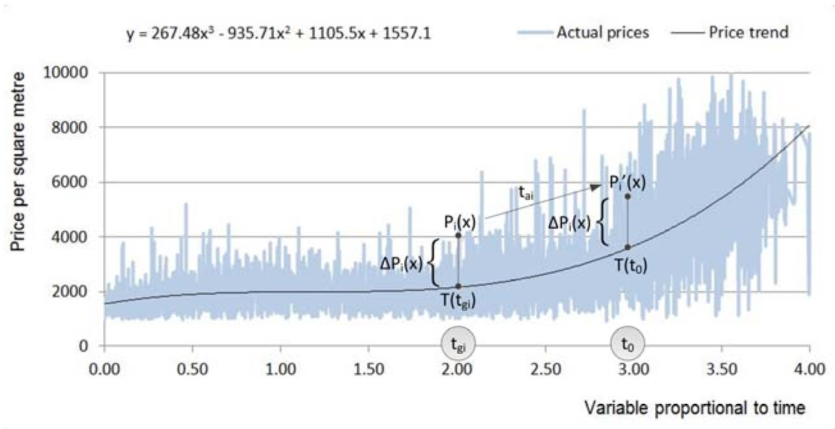

Figure 4: Correcting the output of aged models using Delta method

The idea of correcting the results produced by aged models using the *Ratio* method is depicted in Figure 5. For the time point $t_{gi}=t_0-t_{ai}$, when a given aged model GFS_i was generated, the value of a trend function $T(t_{gi})$, i.e. average price per square metre, is computed. The price of a given premises, i.e. an instance of a current test dataset, characterised by a feature vector $\boldsymbol{x}$, is predicted by the model GFS_i. Next, the total price is divided by the premises usable area to obtain its price per square metre $P_i(\mathbf{x})$. Then, the ratio of the price to the trend value $R_i(\mathbf{x})=P_i(\mathbf{x})/T(t_{gi})$ is calculated. The corrected price per square metre of the premises $P_i'(\mathbf{x})$ is worked out by multiplying this ratio by the trend value in the time point t_0 using the formula

$P_i'(\mathbf{x})=R_i(\mathbf{x})T(t_0)$, where $T(t_0)$ is the value of a trend function in t_0. Finally, the corrected price per square metre $P'(\mathbf{x})$ is converted into the corrected total price of the premises by multiplying it by the premises usable area. The ratio method is denoted by *R* in the symbols of methods used in figures and tables presenting the experimental results further on in the paper.

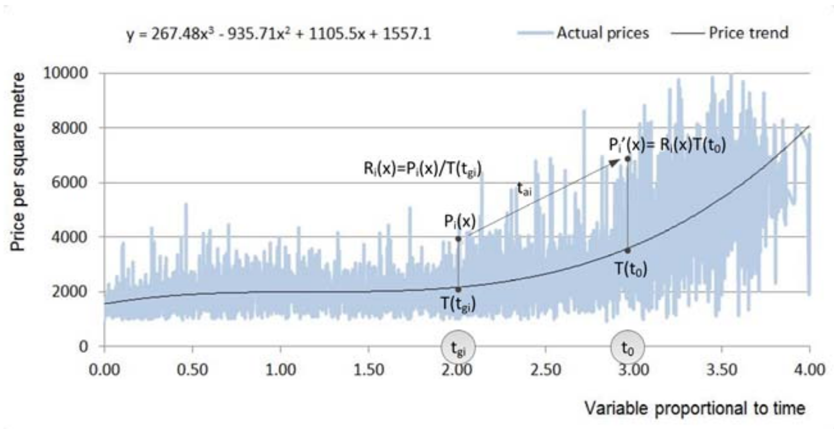

Figure 5: Correcting the output of aged models using Ratio method

Both methods of correcting the prices predicted by aged models are utilised by professional appraisers and the delta method even more frequently. However, the *Ratio* technique seems to be more appropriate than *Delta* method because it allows for the consideration of changes in the variance of prices.

The resulting output of the ensemble for a given instance of the test dataset is computed as the arithmetic mean of the results produced by the component models and corrected by corresponding trend functions.

3 Experimental setup

The investigation was conducted with our experimental system implemented in Matlab environment using Fuzzy Logic, Global Optimization, Neural Network, and Statistics toolboxes. The system was designed to carry out research into machine learning algorithms using various resampling methods and constructing and evaluating ensemble models for regression problems.

Real-world dataset used in experiments was drawn from an unrefined dataset containing above 50,000 records referring to residential premises transactions accomplished in the Polish big city, mentioned in the previous section, within 11 years from 1998 to 2008. In this period most transactions were made with non-market prices when the council was selling flats to their current tenants on preferential terms. First of all, transactional records referring to residential premises sold at market prices were selected. Then the dataset was confined to sales transaction data of apartments built before 1997 and where the land was leased on terms of perpetual usufruct. During the preparation process records with missing attributes and outliers were discarded. The final dataset counted the 5,213 samples. Five following attributes were pointed out as main price drivers by professional appraisers: usable area of a flat (*Area*), age of a building construction (*Age*), number of storeys in a building (*Storeys*), number of rooms in a flat including a kitchen (Rooms), the distance of a building from the city centre (*Centre*), in turn, price of premises (*Price*) was the output variable.

The property valuation models were built from scratch by genetic fuzzy systems over chunks of data stream determined by a sliding window which was 12 months long. The parameters of the architecture of fuzzy systems as well as genetic algorithms are listed in Table 1. Similar designs are described in [Cordón, 99], [Cordón, 04], [Król, 08]. As test datasets the chunks of data stream specified by the time intervals of three months were used. As a performance function the mean absolute error (*MAE*) was used, and as aggregation functions of ensembles arithmetic averages were employed.

Fuzzy system	**Genetic Algorithm**
Type of fuzzy system: Mamdani	Chromosome: rule base and mf, real-coded
No. of input variables: 5	Population size: 100
Membership functions (mf): triangular	Fitness function: MSE
No. of input mf: 3	Selection function: tournament
No. of output mf: 5	Tournament size: 4
No. of rules: 15	Elite count: 2
AND operator: prod	Crossover fraction: 0.8
Implication operator: prod	Crossover function: two point
Aggregation operator: probor	Mutation function: custom
Defuzzyfication method: centroid	No. of generations: 100

Table 1: Parameters of GFS used in experiments

We conducted three series of experiments. In the first one the performance of ageing single models was investigated, in the second the impact of the degree of polynomial trend functions on the accuracy of ensemble fuzzy models was explored, and finally and in the third series the accuracy of *GFS* ensembles created in 20 time points were compared and nonparametric tests of statistical significance were carried out. The trends were modelled using Matlab function *polyfit(x,y,n)*, which finds the coefficients of a polynomial *p(x)* of degree *n* that fits the *y* data by minimizing the sum of the squares of the deviations of the data from the model (least-squares fit).

The analysis of the results was performed using statistical methodology including nonparametric tests followed by post-hoc procedures designed especially for multiple $N \times N$ comparisons [Demšar, 06], [García, 08], [Trawiński, 12c]. The idea behind statistical methods applied to analyse the results of experiments was as follows. The commonly used paired tests i.e. parametric t-test and its nonparametric alternative Wilcoxon signed rank tests are not appropriate for multiple comparisons due to the so called family-wise error. The proposed routine starts with the nonparametric Friedman test, which detect the presence of differences among all algorithms compared. After the null-hypotheses have been rejected the post-hoc procedures should be applied in order to point out the particular pairs of algorithms which produce differences. For $N \times N$ comparisons nonparametric Nemenyi's, Holm's, Shaffer's, and Bergamnn-Hommel's procedures are recommended.

4 Results of experiment

We conducted three series of experiments. In the first one the performance of ageing single models was investigated, in the second the impact of the degree of polynomial trend functions on the accuracy of ensemble fuzzy models was explored, and finally and in the third series the accuracy of GFS ensembles created in 20 time points were compared and nonparametric tests of statistical significance were carried out.

4.1 Performance of ageing single models

The first series of experiments consisted in the investigation of ageing single models created by genetic fuzzy systems (*GFSs*). It was carried out for t_0 equal to January 1, 2006, which corresponded the value of 2.92 on the *x* axis in the chart in Figure 6. In this time the prices of residential premises were growing quickly due to the run on real estate. We compared the performance of single models with and without correcting the results produced by aged models. In order to compensate ageing the output produced by aged models linear trend functions *T1* were employed. They were determined over shorter time periods t_w plus t_{ai} (*Age*) and the *Delta* correction method (*D*) was applied. Such an approach is commonly applied by Polish professional appraisers who having frequently a small number of comparable transactional data tend to use linear polynomial functions to avoid overfitting.

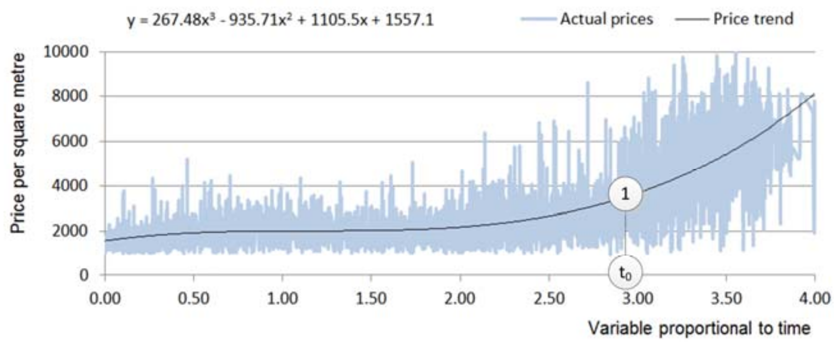

Figure 6: Change trend of prices with marked t_0 equal to January 1, 2006

The performance of ageing single models in terms of *MAE* is illustrated graphically in Figure 7. The *x* axis shows the age of models, i.e. the how many months passed from time when respective models were created to the time point t_0. The models were generated with training datasets determined by sliding windows of t_w=12 months. The windows were shifted of t_s=3 months. The same three test datasets, current for t_0, determined by the interval of 1, 2, and 3 months were employed to all current and aged models. Following denotation was used in the legend of Figure 7: *noT* means that output produced by respective models for the current test dataset was not updated, respectively. The updated results were in turn marked by *Age-DT1(t_i)*. In the graph *MAE* is expressed in thousand PLN, where PLN stands for Polish zlotys.

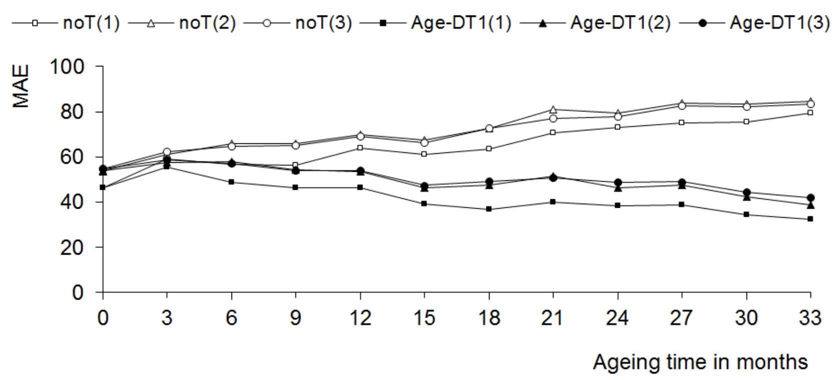

Figure 7: Performance of ageing single models, where (1), (2), (3) stand for test sets built over intervals 1, 2, and 3 months long, respectively

In the chart it is clearly seen that the *MAE* values for models whose output was not corrected by trend functions grow as ageing time increases. The reverse relation can be noticed when the results produced by models were updated with trend functions. Furthermore, the older models with trend correction reveal better performance than the less aged ones. This can be explained by smaller variance of premises prices in earlier time intervals (see Figure 6). Moreover, the shorter time span (starting from t_0) of test data the lower *MAE* value may indicate that data included in test sets also undergo ageing.

4.2 Impact of trend functions with different degree

The second series of experiments aimed at exploring the impact of different trend functions on the accuracy of ensemble fuzzy models for residential premises valuation. The evaluating experiments were conducted for three points of time equal to 1) t_0=2004-01-01, 2) t_0=2006-01.01, and 3) t_0=2008-01-01 denoted by circled numbers in Figure 8, which correspond to the values of 2.19, 2.92, and 3.65 on the *x* axis in the graph, respectively. Thus, 1) the first moment refers to a modest rise of real estate prices just before Poland entered the European Union (EU), 2) the second one corresponds to a stronger growth two years after joining EU by Poland, 3) the last point was selected in the period when the prices of residential premises were increasing rapidly during the worldwide real estate bubble. As the functions to model the trends of price changes the polynomials of degree from one to five were employed, denoted by *T1*, *T2*,..,*T5*, respectively. They were determined over both shorter (*Age*) and longer time periods (*Beg*) and to correct the output of aged models the *Delta* correction method (*D*) was employed in each case. For comparison the results provided by the ensembles with no trend update were also taken into account.

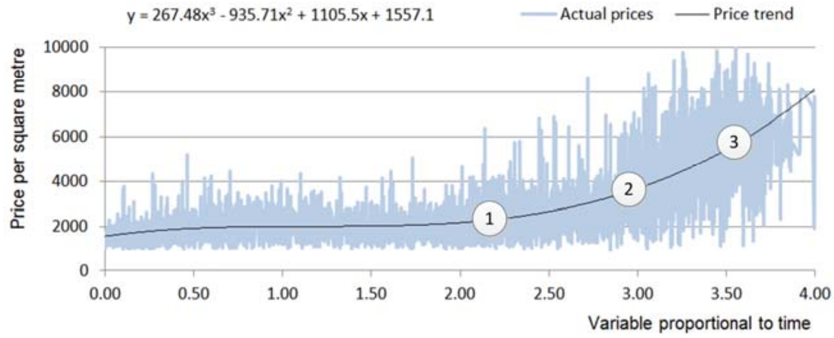

Figure 8: Change trend of average transactional prices with marked three time points equal to 1) January 1, 2004, 2) January 1, 2006, and 3) January 1, 2008

The performance of GFS ensembles is depicted in six following Figures 9-14 for two per each time point so that the results for *Age* and *Beg* periods are illustrated. The ensembles were composed of stepwise growing number of genetic fuzzy systems (*GFS*). To a single model, current for a given t_0, more and more aged models were added, which were built over training data of the previous sliding windows of t_w equal to 12 months shifted by t_s equal to three months. The same test datasets, current for a given t_0, determined by the interval of 3 months were applied to each ensemble. Following denotation was used in the legend of Figures 9-14: *noT* stands for the ensembles which the results produced by component models were not updated with any trend function, *Age-DT*$_i$ and *Beg-DT*$_i$ indicate whether the outcome provided by component models was updated using trend functions determined over shorter (*Age*) or longer (*Beg*) time periods as described in the previous section. In order to be concise, in remaining text of the paper we will call the former *Age Trends* and the latter *Beg Trends*. Moreover, *D* stands for the *Delta* method and i=1,2,..,5 indicates the degree of polynomial function. In each graph *MAE* is expressed in thousand PLN.

For the time point 1), i.e. by the moderate price growth, it can be noticed for *Age Trends* in Figure 9 that the bigger number of component models the better accuracy of the ensemble except for linear and quadratic functions where it was the case only for ensembles composed of more than eight *GFSs*. Better results were achieved using polynomial of degree from three to five than with linear and quadratic functions. The best performance revealed polynomials of degree four and five. The lowest values of *MAE* were obtained for the biggest sizes of ensembles and the ensembles with corrected outcome significantly outperformed the single model built over the current data set. Moreover the former noticeably outperformed the ones which the results produced by component models were not updated with any trend function (*noT*).

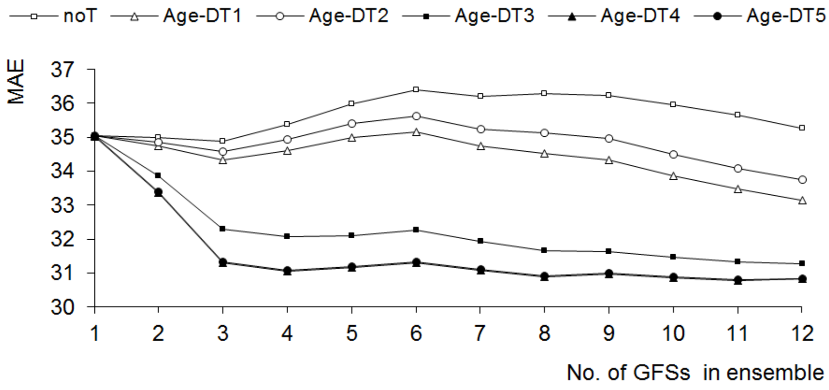

Figure 9: Performance of GFS ensembles for Age Trends for t_0=2004-01-01

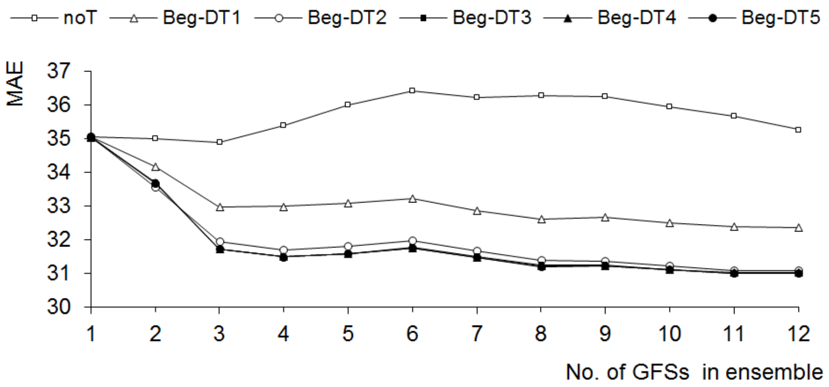

Figure 10: Performance of GFS ensembles for Beg Trends for t_0=2004-01-01

In turn, for *Beg Trends* in Figure 10 it is clearly seen that that the bigger number of component models the better accuracy of the ensemble for all trend functions. The ensembles with corrected output significantly outperformed the *noT* ones. The best performance revealed polynomials of degree three, four, and five and the differences in accuracy among them were unnoticeable.

For the time point 2), i.e. by the stronger price rise, it can be also seen in Figure 11 and 12 that the bigger number of component models in ensembles the greater accuracy except for linear function for *Beg Trends*. Better results were achieved using polynomials of higher degree. For *Beg Trends* the best performance revealed cubic polynomials. The biggest ensembles with corrected output significantly outperformed the single model built over the current data set except for linear functions for *Beg Trends*. For the *noT* ensembles the bigger number of component models the higher value of *MAE*.

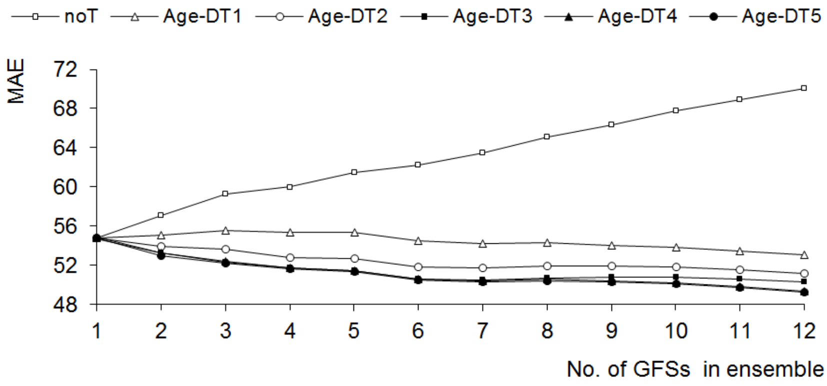

Figure 11: Performance of GFS ensembles for Age Trends for t_0=2006-01-01

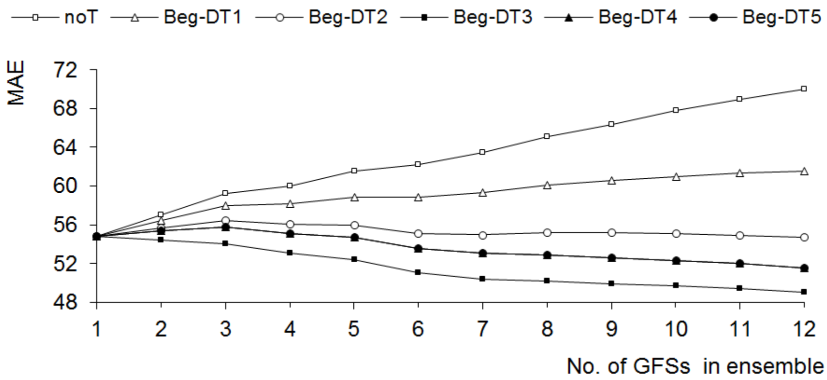

Figure 12: Performance of GFS ensembles for Beg Trends for t_0=2006-01-01

For the time point 3), i.e. by the most dramatic increase of premises prices, it can be observed in Figure 13 that among *Age Trends* the best accuracy provided linear functions. The ensembles with the output corrected with polynomials of degree three, four, and five did not reveal even better performance than a single model does. The reverse relation can be noticed for *Beg Trends* (see Figure 14) where cubic, quartic, and quantic polynomials surpass significantly linear and quadratic functions. For *Beg-DT3*, *Beg-DT4*, and *Beg-DT5* the lowest values of *MAE* were obtained for the ensembles of the biggest sizes and the ensembles with corrected output significantly outperformed the single model built over the current data set. For the *noT* ensembles the bigger number of component models the more dramatic value of MAE. Likewise, the linear polynomials work poorly for *Beg Trends*.

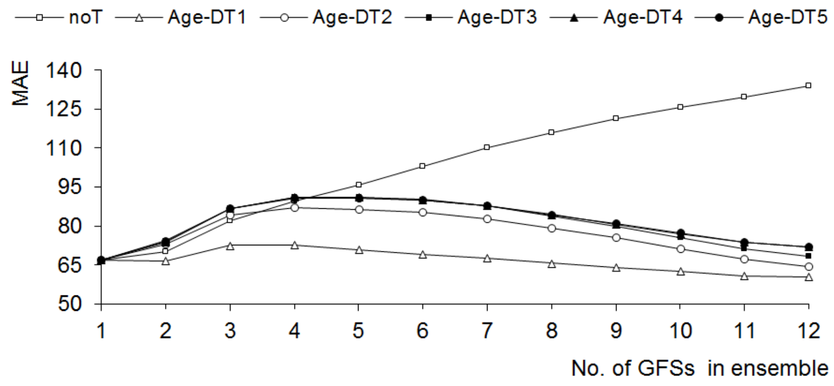

Figure 13: Performance of GFS ensembles for Age Trends for t_0=2008-01-01

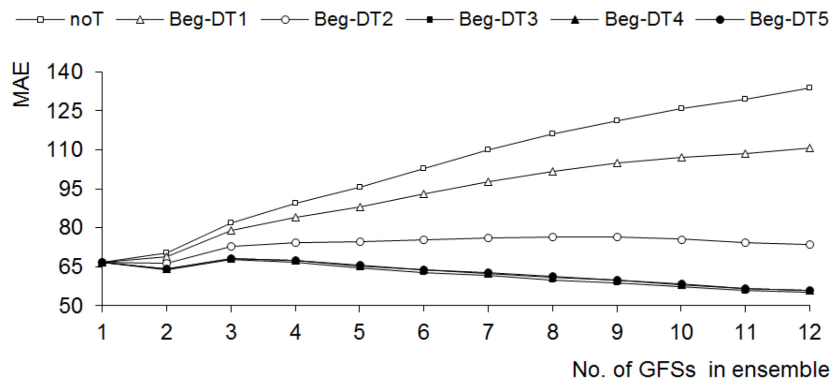

Figure 14: Performance of GFS ensembles for Beg Trends for t_0=2008-01-01

	t_0	**2004-01-01**		**2006-01-01**		**2008-01-01**	
#models	**Trend**	**Age**	**Beg**	**Age**	**Beg**	**Age**	**Beg**
1	noT	35.1	35.1	54.8	54.8	66.8	66.8
12	noT	35.3	35.3	70.0	70.0	133.9	133.9
12	T1	33.1	32.4	53.1	61.6	**60.4**	110.8
12	T2	33.7	31.1	51.1	54.7	64.4	73.6
12	T3	31.3	**31.0**	50.3	**49.1**	68.5	**55.0**
12	T4	**30.8**	31.0	49.4	51.5	71.8	55.8
12	T5	30.8	31.0	**49.3**	51.5	71.9	55.9

Table 2: Accuracy comparison of single models and ensembles composed of 12 GFSs. Performance in terms of MAE is expressed in thousand PLN

Table 2 summarizes the results of this series of experiments. The best results provided the ensembles composed of 12 aged models. The best results for each column in Table 2 are stressed in boldface. For any length of interval to determine the trend function you could find such trend function degree that ensured the better accuracy of ensembles than single models. The ensembles with no output correction by trend functions in each case revealed worse performance. Moreover, the trend functions determined over longer intervals (*Beg*) led to better results than the ones built over shorter time periods (*Age*). Taking into account all three time points with different rate of price changes the best selection seems to the polynomial of degree three build over data which came in since the beginning of the stream.

4.3 Comparison of GFS ensembles

The goal of the third series of experiments was to compare the accuracy of ensembles of *GFSs* created using the aforementioned techniques and to conduct nonparametric tests of statistical significance. The comparison experiments were conducted for 20 points of time from 1) t_0=2003-01-01 to 20) t_0=2007-10-01, denoted by circled numbers in Figure 15. Thus, the time points covered different rates of price growth from modest to massive rise. Taking into account the results of previous experiments in each time point an ensemble of constant size, i.e. composed of 12 aged *GFSs*, was built. Component models were built over training data delineated by the sliding windows of t_w equal to 12 months shifted by t_s equal to three months. The same test datasets, current for a given t_0, determined by the interval of 3 months were applied to each ensemble. The performance of the ensembles was computed for a test set current for individual t_0, determined by the interval of 3 months. As in previous experiments to model the trends of price changes the polynomials of degree from one to five were employed. They were determined over both shorter (*Age*) and longer time periods (*Beg*) and to correct the output of aged models the *Delta* (*D*) and *Ratio* (*R*) correction methods were employed in each case. For comparison the results provided by single models current for individual time points were also utilized.

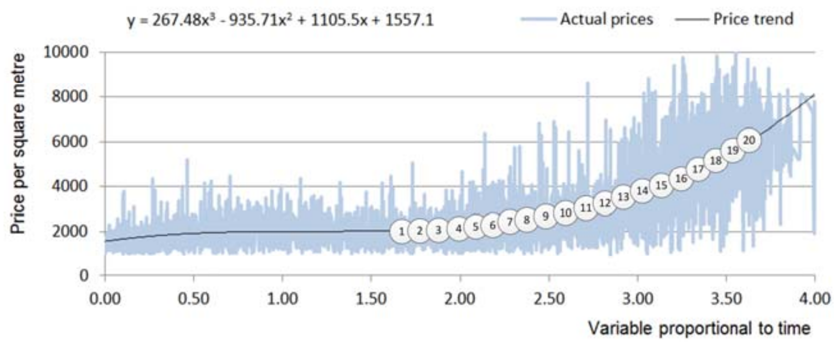

Figure 15: Change trend of average transactional prices with marked 20 time points

The performance of six models, i.e. single ones and ensembles with the correction of component model output by polynomials of degree from one to five using *Age Trends* and *Delta* and *Ratio* methods is illustrated in Figures 16 and 17. It can be easily seen that the single models and *T1* ensembles provide the highest values of

MAE. In turn, the differences among the rest of the ensembles are not apparent therefore we should refer to statistical tests of significance.

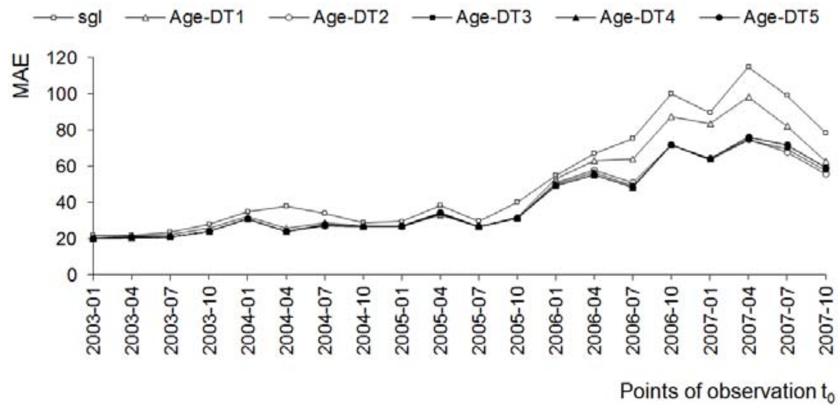

Figure 16: Performance of GFS ensembles for Age Trends with Delta method

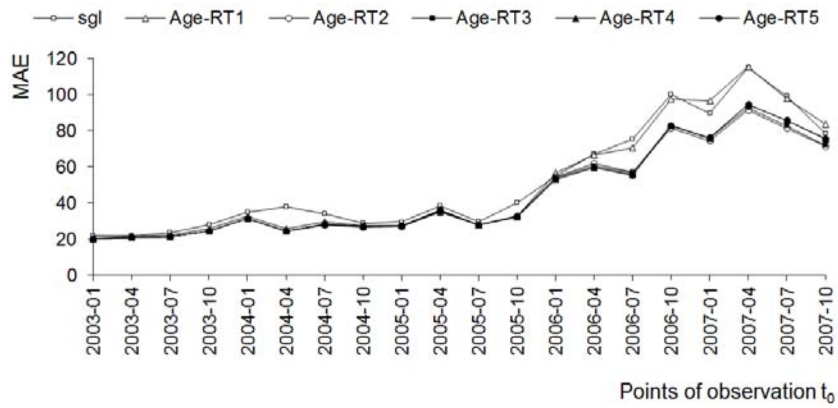

Figure 17: Performance of GFS ensembles for Age Trends with Ratio method

The Friedman test performed in respect of *MAE* values of the models built in 20 time points showed that there were significant differences between some models. Average ranks of individual models are shown in Table 3, where the lower rank value the better model. On both lists the *T3* and *T4* ensembles took the first two positions whereas single models and *T1* ensembles were in the last two places. Adjusted *p-values* for Nemenyi's, Holm's, Shaffer's, and Bergmann-Hommel's post-hoc procedures for *N×N* comparisons for all possible pairs of algorithms are shown in Tables 4 and 5 for *Delta* and *Ratio* methods, respectively. For comparison, the results of paired Wilcoxon tests are placed in both tables. The *p-values* indicating the statistically significant differences between given pairs of algorithms are marked with italics. The significance level considered for the null hypothesis rejection was 0.05.

Following main observations could be done based on the most powerful Shaffer's, and Bergmann-Hommel's post-hoc procedures: *T2*, *T3*, *T4*, and *T5* ensembles revealed significantly better performance than single and *T1* models. There are not significant differences among the *T2*, *T3*, *T4*, and *T5* ensembles as well as among single and *T1* models. The paired Wilcoxon test can lead to over-optimistic decisions because it allows for rejection of a greater number of null hypotheses.

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**
Age Trends, Delta	T3 (1.95)	T4 (2.50)	T2 (2.85)	T5 (3.05)	T1 (4.65)	sgl (6.00)
Age Trends, Ratio	T3 (2.10)	T2 (2.35)	T4 (2.85)	T5 (2.95)	T1 (4.95)	sgl (5.80)

Table 3: Rank positions of models for Age Trends determined by Friedman test

Model vs Model	**pWilcox**	**pNeme**	**pHolm**	**pShaf**	**pBerg**
sgl vs Age-DT3	*8.90E-05*	*1.14E-10*	*1.14E-10*	*1.14E-10*	*1.14E-10*
sgl vs Age-DT4	*8.90E-05*	*4.95E-08*	*4.62E-08*	*3.30E-08*	*3.30E-08*
sgl vs Age-DT2	*8.90E-05*	*1.52E-06*	*1.32E-06*	*1.01E-06*	*7.09E-07*
sgl vs Age-DT5	*8.90E-05*	*9.23E-06*	*7.38E-06*	*6.15E-06*	*4.31E-06*
Age-DT1 vs Age-DT3	*0.000254*	*7.53E-05*	*5.53E-05*	*5.02E-05*	*5.02E-05*
Age-DT1 vs Age-DT4	*0.000339*	*0.004183*	*0.002789*	*0.002789*	*0.001673*
Age-DT1 vs Age-DT2	*0.000189*	*0.035187*	*0.021112*	*0.016421*	*0.009383*
Age-DT1 vs Age-DT5	*0.000593*	0.102613	0.054727	*0.047886*	*0.027363*
sgl vs Age-DT1	*8.90E-05*	0.337414	0.157460	0.157460	0.157460
Age-DT4 vs Age-DT5	*0.008035*	1.000000	1.000000	1.000000	1.000000
Age-DT3 vs Age-DT5	0.100459	0.944686	0.377874	0.377874	0.377874
Age-DT3 vs Age-DT4	0.135358	1.000000	1.000000	1.000000	1.000000
Age-DT2 vs Age-DT3	0.278966	1.000000	0.640951	0.512761	0.384571
Age-DT2 vs Age-DT5	0.525654	1.000000	1.000000	1.000000	1.000000
Age-DT2 vs Age-DT4	0.627446	1.000000	1.000000	1.000000	1.000000

Table 4: p-values for N×N comparisons of models for Age Trends with Delta method

Model vs Model	**pWilcox**	**pNeme**	**pHolm**	**pShaf**	**pBerg**
sgl vs Age-RT3	*8.90E-05*	*6.00E-09*	*6.00E-09*	*6.00E-09*	*6.00E-09*
sgl vs Age-RT2	*8.90E-05*	*8.24E-08*	*7.69E-08*	*5.49E-08*	*5.49E-08*
sgl vs Age-RT4	*8.90E-05*	*9.23E-06*	*8.00E-06*	*6.15E-06*	*4.31E-06*
sgl vs Age-RT5	*8.90E-05*	*2.18E-05*	*1.75E-05*	*1.45E-05*	*1.02E-05*
Age-RT1 vs Age-RT3	*0.000189*	*2.18E-05*	*1.75E-05*	*1.45E-05*	*1.45E-05*
Age-RT1 vs Age-RT2	*0.000163*	*0.000166*	*0.000111*	*0.000111*	*6.65E-05*
Age-RT1 vs Age-RT4	*0.000293*	*0.005786*	*0.003472*	*0.002700*	*0.001543*
Age-RT1 vs Age-RT5	*0.000293*	*0.010848*	*0.005786*	*0.005063*	*0.002893*
sgl vs Age-RT1	*0.027622*	1.000000	1.000000	1.000000	1.000000
Age-RT3 vs Age-RT5	0.262723	1.000000	1.000000	1.000000	1.000000
Age-RT3 vs Age-RT4	0.313464	1.000000	1.000000	1.000000	1.000000
Age-RT2 vs Age-RT5	0.135358	1.000000	1.000000	1.000000	1.000000
Age-RT2 vs Age-RT4	0.167185	1.000000	1.000000	1.000000	1.000000
Age-RT2 vs Age-RT3	0.525654	1.000000	1.000000	1.000000	1.000000
Age-RT4 vs Age-RT5	0.627446	1.000000	1.000000	1.000000	1.000000

Table 5: p-values for N×N comparisons of models for Age Trends with Ratio method

Similar analysis was made for *Beg Trends*. The performance of six models using *Delta* and *Ratio* methods is illustrated in Figures 18 and 19. It is noticeable that the single models and *T1* and *T2* ensembles reveal the highest values of *MAE*. In turn, the differences among the rest of the ensembles are not visible. The Friedman test indicated that there were significant differences between some models. Average ranks of individual models are shown in Table 6. On both lists the *T4*, *T5*, and *T3* ensembles took the first three positions whereas single models and *T1* and *T2* ensembles were in the last three places. Adjusted *p-values* for Nemenyi's, Holm's, Shaffer's, and Bergmann-Hommel's post-hoc procedures for *N×N* comparisons for all possible pairs of algorithms are shown in Tables 7 and 8 for *Delta* and *Ratio* methods, respectively. The *p-values* indicating the statistically significant differences between given pairs of algorithms are also marked with italics. The significance level considered for the null hypothesis rejection was 0.05, too. Following main observations could be done based on the Shaffer's, and Bergmann-Hommel's post-hoc procedures: *T3*, *T4*, and *T5* ensembles revealed significantly better performance than single and *T1* and *T2* models. There are not significant differences among the *T3*, *T4*, and *T5* ensembles as well as among single and *T1* and *T2* models. The paired Wilcoxon test also allowed for the rejection of a bigger number of null hypotheses.

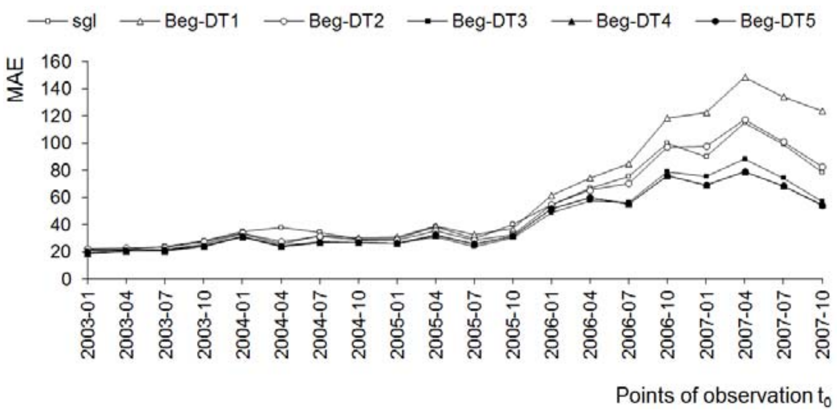

Figure 18: Performance of GFS ensembles for Beg Trends with Delta method

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**
Beg Trends, Delta	T4 (1.75)	T5 (2.05)	T3 (2.40)	T2 (4.65)	sgl (5.05)	T1 (5.10)
Beg Trends, Ratio	T4 (1.85)	T5 (1.95)	T3 (2.55)	T2 (4.80)	sgl (4.80)	T1 (5.05)

Table 6: Rank positions of models for Beg Trends determined by Friedman test

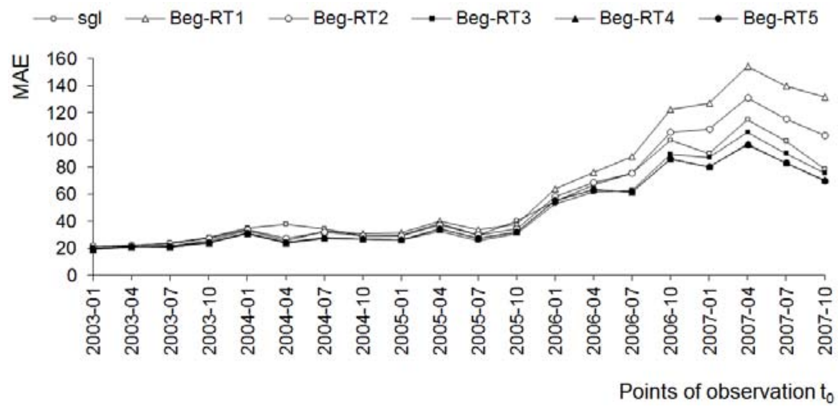

Figure 19: Performance of GFS ensembles for Beg Trends with Ratio method

Model vs Model	pWilcox	pNeme	pHolm	pShaf	pBerg
Beg-DT1 vs Beg-DT4	*0.000120*	*2.24E-07*	*2.24E-07*	*2.24E-07*	*2.24E-07*
sgl vs Beg-DT4	*0.000089*	*3.65E-07*	*3.41E-07*	*2.43E-07*	*2.43E-07*
Beg-DT1 vs Beg-DT5	*0.000140*	*3.80E-06*	*3.29E-06*	*2.53E-06*	*2.53E-06*
sgl vs Beg-DT5	*8.90E-05*	*5.94E-06*	*4.75E-06*	*3.96E-06*	*2.53E-06*
Beg-DT2 vs Beg-DT4	*8.90E-05*	*1.42E-05*	*1.04E-05*	*9.49E-06*	*6.64E-06*
Beg-DT1 vs Beg-DT3	*0.000293*	*7.53E-05*	*5.02E-05*	*5.02E-05*	*3.52E-05*
sgl vs Beg-DT3	*8.90E-05*	*0.000112*	*6.74E-05*	*5.24E-05*	*3.52E-05*
Beg-DT2 vs Beg-DT5	*8.90E-05*	*0.000166*	*8.87E-05*	*7.76E-05*	*4.43E-05*
Beg-DT2 vs Beg-DT3	*8.90E-05*	*0.002143*	*0.001000*	*0.001000*	*0.000429*
Beg-DT1 vs Beg-DT2	*0.008968*	1.000000	1.000000	1.000000	1.000000
Beg-DT3 vs Beg-DT4	0.135358	1.000000	1.000000	1.000000	1.000000
sgl vs Beg-DT2	0.178957	1.000000	1.000000	1.000000	1.000000
Beg-DT3 vs Beg-DT5	0.145401	1.000000	1.000000	1.000000	1.000000
Beg-DT4 vs Beg-DT5	0.370262	1.000000	1.000000	1.000000	1.000000
sgl vs Beg-DT1	0.125860	1.000000	1.000000	1.000000	1.000000

Table 7: p-values for N×N comparisons of models for Beg Trends with Delta method

Model vs Model	pWilcox	pNeme	pHolm	pShaf	pBerg
Beg-RT1 vs Beg-RT4	*0.000140*	*9.51E-07*	*9.51E-07*	*9.51E-07*	*9.51E-07*
Beg-RT1 vs Beg-RT5	*0.000140*	*2.41E-06*	*2.25E-06*	*1.61E-06*	*1.61E-06*
sgl vs Beg-RT4	*0.000103*	*9.23E-06*	*8.00E-06*	*6.15E-06*	*6.15E-06*
Beg-RT2 vs Beg-RT4	*8.90E-05*	*9.23E-06*	*8.00E-06*	*6.15E-06*	*6.15E-06*
sgl vs Beg-RT5	*0.000103*	*2.18E-05*	*1.60E-05*	*1.45E-05*	*8.73E-06*
Beg-RT2 vs Beg-RT5	*8.90E-05*	*2.18E-05*	*1.60E-05*	*1.45E-05*	*8.73E-06*
Beg-RT1 vs Beg-RT3	*0.000293*	*0.000357*	*0.000214*	*0.000167*	*0.000167*
sgl vs Beg-RT3	*8.90E-05*	*0.002143*	*0.001143*	*0.001000*	*0.000571*
Beg-RT2 vs Beg-RT3	*8.90E-05*	*0.002143*	*0.001143*	*0.001000*	*0.000571*
Beg-RT1 vs Beg-RT2	*0.005740*	1.000000	1.000000	1.000000	1.000000
sgl vs Beg-RT1	0.056915	1.000000	1.000000	1.000000	1.000000
Beg-RT3 vs Beg-RT5	0.135358	1.000000	1.000000	1.000000	1.000000
Beg-RT3 vs Beg-RT4	0.145401	1.000000	1.000000	1.000000	1.000000
sgl vs Beg-RT2	0.370262	1.000000	1.000000	1.000000	1.000000
Beg-RT4 vs Beg-RT5	0.654159	1.000000	1.000000	1.000000	1.000000

Table 8: p-values for N×N comparisons of models for Beg Trends with Ratio method

T3 and *T4* ensembles, which revealed the best performance in the above-mentioned experiments, were selected for the final comparison. The Friedman test indicated that there were significant differences between some models. Average ranks determined by Friedman test are placed in Table 9 and adjusted *p-values* for post-hoc procedures for *N×N* comparisons for all possible pairs of models are shown in Table 10. Having performed the analogous analysis following main conclusions can be drawn. The *Delta* correction method results in significantly better performance than the *Ratio* one. Among *Delta* corrected ensembles as well as among *Ratio* corrected ensembles there are no significant differences between *T3* and *T4* models and *Age* and *Beg Trends* methods.

1st	2nd	3rd	4th	5th	6th	7th	8th
Beg-DT4 (2.40)	Age-DT3 (2.85)	Age-DT4 (3.40)	Beg-DT3 (3.95)	Beg-RT4 (4.90)	Age-RT3 (5.90)	Beg-RT3 (6.20)	Age-RT4 (6.40)

Table 9: Rank positions of T3 and T4 models determined during Friedman test

Model vs Model	pWilcox	pNeme	pHolm	pShaf	pBerg
Beg-DT4 vs Age-RT4	*0.000120*	*6.77E-06*	*6.77E-06*	*6.77E-06*	*6.77E-06*
Beg-DT4 vs Beg-RT3	*0.000390*	*2.61E-05*	*2.51E-05*	*1.95E-05*	*1.95E-05*
Age-DT3 vs Age-RT4	*0.000089*	*0.000128*	*0.000119*	*9.62E-05*	*9.62E-05*
Beg-DT4 vs Age-RT3	*0.000103*	*0.000174*	*0.000156*	*0.000131*	*9.97E-05*
Age-DT3 vs Beg-RT3	*0.016882*	*0.000427*	*0.000366*	*0.000321*	*0.000229*
Age-DT3 vs Age-RT3	*0.000103*	*0.002305*	*0.001893*	*0.001729*	*0.000906*
Age-DT4 vs Age-RT4	*0.000103*	*0.003010*	*0.002365*	*0.002258*	*0.001720*
Age-DT4 vs Beg-RT3	*0.015241*	*0.008417*	*0.006313*	*0.006313*	*0.003307*
Beg-DT4 vs Beg-RT4	*0.000163*	*0.034967*	*0.024977*	*0.019981*	*0.016235*
Age-DT4 vs Age-RT3	*0.000163*	*0.034967*	*0.024977*	*0.019981*	*0.016235*
Beg-DT3 vs Age-RT4	*0.003185*	*0.043732*	*0.028114*	*0.024990*	*0.020304*
Beg-DT3 vs Beg-RT3	*0.000293*	0.102917	0.062485	0.058810	*0.033081*
Age-DT3 vs Beg-RT4	*0.003592*	0.227697	0.130112	0.130112	0.073188
Beg-DT3 vs Age-RT3	*0.005734*	0.330992	0.177317	0.177317	0.082748
Beg-DT3 vs Beg-RT4	*0.005734*	1.000000	1.000000	1.000000	1.000000
Beg-DT4 vs Beg-DT3	0.135358	1.000000	0.635431	0.590043	0.544655
Beg-RT4 vs Age-RT4	0.793839	1.000000	0.686498	0.686498	0.544655
Age-DT4 vs Beg-RT4	0.052223	1.000000	0.686498	0.686498	0.544655
Beg-RT4 vs Beg-RT3	0.145401	1.000000	1.000000	1.000000	0.559740
Age-DT3 vs Beg-DT3	0.232226	1.000000	1.000000	1.000000	0.933482
Beg-DT4 vs Age-DT4	0.910825	1.000000	1.000000	1.000000	1.000000
Beg-RT4 vs Age-RT3	0.793839	1.000000	1.000000	1.000000	1.000000
Age-DT3 vs Age-DT4	0.135358	1.000000	1.000000	1.000000	1.000000
Age-DT4 vs Beg-DT3	0.350657	1.000000	1.000000	1.000000	1.000000
Age-RT3 vs Age-RT4	0.313464	1.000000	1.000000	1.000000	1.000000
Beg-DT4 vs Age-DT3	0.940481	1.000000	1.000000	1.000000	1.000000
Age-RT3 vs Beg-RT3	0.156005	1.000000	1.000000	1.000000	1.000000
Beg-RT3 vs Age-RT4	0.262723	1.000000	1.000000	1.000000	1.000000

Table 10: p-values for N×N comparisons of all possible pairs of T3 and T4 models

5 Conclusions and Future Work

A method to predict from a data stream of real estate sales transactions based on ensembles of genetic fuzzy systems was proposed. The approach consists in incremental expanding an ensemble by models built from scratch over successive chunks of a data stream determined by a sliding window. The predicted prices of residential premises computed by aged component models for current data are updated according to a trend function reflecting the changes of the market. The impact of different trend functions on the accuracy of single and ensemble fuzzy models was investigated in the paper.

Moreover, the intensive evaluating experiments were conducted. They consisted in generating ensembles for 20 points of time and comparing their accuracy using nonparametric tests of statistical significance adequate for multiple comparisons. The time points covered different rates of real estate price growth from modest to massive rise. As the functions to model the trends of price changes the polynomials of degree from one to five were employed. The trends were determined over two time periods: shorter and longer ones. The shorter periods encompassed the length of a sliding window plus model ageing time whereas the longer ones took into account all data since the beginning of the stream. Two methods of correcting the output of component models were proposed: *Delta* and *Ratio* ones. The former was based on the difference between a predicted price and a trend value in a given time point and the latter on the ratio of the price to the trend value.

The results proved the usefulness of ensemble approach incorporating the correction of individual component model output. The investigation of different trend functions indicated that the polynomials of degree three build over data which came in since the beginning of the stream provided the best accuracy of single and ensemble fuzzy models. However, the statistical tests based on the results of experiments over 20 points of time covering different rates of price changes showed that significantly better performance revealed ensembles using the *Delta* correction method than the *Ratio* one. Among *Delta* corrected ensembles as well as among *Ratio* corrected ensembles there were no statistically significant differences between *T3* and *T4* models and *Age Trends* and *Beg Trends* methods.

In this paper we presented our attempt to employ evolutionary fuzzy approach to explore data streams to model dynamic real estate market. The problem is not trivial because on the one hand a genetic fuzzy system needs a number of samples to be trained without overfitting and on the other hand the time window to determine a chunk of training data should be as small as possible to diminish the ageing impact and retain the model accuracy at an acceptable level. The processing time in this case is not a decisive factor because property valuation models need not to be updated and/or generated in an on-line mode. Thus, our approach, outlined above, raises a challenge to find the trade-off between the length of a sliding window delimiting the training dataset and the deteriorating impact of ageing models, overfitting, and computational efficiency.

So far in our study an ensemble has been treated as a black box. Further investigation is planned to explore the intrinsic structure of component models, i.e. their knowledge and rule bases, as well as their generation efficiency, interpretability, the problems of overfitting and outliers. Moreover, weighting component models according to their estimated accuracy and the time of ageing will be examined. The balance between the length of a sliding window and a model performance will be explored. When models are built from scratch over relatively small amount of data it may happen that data coming within the next period will not comply with existing models and the accuracy of an ensemble may lower. In order to prevent this undesirable situation clustering, random oracle, or stratification of data could be employed.

It is also planned to conduct experiments using as base learning algorithms other methods capable of learning from concept drifts such as: decision trees, recurrent neural networks, support vector regression, etc. Moreover, we intend to compare the outcome produced by proposed genetic fuzzy models with human based predictions.

Acknowledgements

This paper was partially supported by the Polish National Science Centre under grant no. N N516 483840.

References

[Alonso, 09] Alonso, J.M., Magdalena, L., González-Rodríguez, G.: Looking for a good fuzzy system interpretability index: An experimental approach. International Journal of Approximate Reasoning 51, 115–134 (2009)

[Angelov, 04] Angelov, P.P., Filev, D.: An approach to online identification of Takagi-Sugeno fuzzy models. IEEE Transactions on Systems, Man and Cybernetics, part B 34:1, 484–498 (2004)

[Bifet, 09a] Bifet, A., Holmes, G., Pfahringer, B., Gavalda, R.:Improving Adaptive Bagging Methods for Evolving Data Streams. In Z.-H. Zhou and T. Washio (Eds.): ACML 2009, LNAI 5828, pp. 23–37, Springer, Heidelberg (2009)

[Bifet, 09b] Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., Gavalda, R.: New ensemble methods for evolving data streams. In: Elder IV, J.F. et al. (Eds.) KDD-2009, pp. 139-148, ACM Press, New York (2009)

[Brzeziński, 11] Brzeziński, D., Stefanowski, J.: Accuracy Updated Ensemble for Data Streams with Concept Drift In E. Corchado et al. (Eds.): HAIS 2011, LNAI 6679, pp. 155-163, Springer, Heidelberg (2011)

[Castro, 96] Castro, J.L., Delgado, M.: Fuzzy systems with defuzzification are universal approximators. IEEE Transactions on System, Man and Cybernetics 26, 149–152 (1996)

[Cordón, 04] Cordón, O., Gomide, F., Herrera, F., Hoffmann, F., Magdalena, L.: Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets and Systems 141, 5–31 (2004)

[Cordón, 99] Cordón, O., Herrera, F.: A Two-Stage Evolutionary Process for Designing TSK Fuzzy Rule-Based Systems. IEEE Transactions on Systems, Man and Cybernetics, part B 29:6, 703–715 (1999)

[Demšar, 06] Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, pp. 1–30 (2006)

[Elwell, 11] Elwell, R., Polikar, R.: Incremental Learning of Concept Drift in Nonstationary Environments. IEEE Transactions on Neural Networks 22:10, 1517–1531 (2011)

[Gaber, 12] Gaber, M.M.: Advances in data stream mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2:1, pp. 79–85 (2012)

[Gama, 05] Gama, J., Medas, P.: Learning Decision Trees from Dynamic Data Streams. Journal of Universal Computer Science 11:8, 1353–1366 (2005)

[García, 08] García, S., Herrera, F.: An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. Journal of Machine Learning Research 9, pp. 2677–2694 (2008)

[Jiang, 09] Jiang, T., Feng, Y., Zhang, B.: Online Detecting and Predicting Special Patterns over Financial Data Streams. Journal of Universal Computer Science 15:13, 2566–2585 (2009)

[Jung, 05] Jung, J.J.: Semantic Preprocessing of Web Request Streams for Web Usage Mining. Journal of Universal Computer Science 11:8, 1383–1396 (2005)

[Kempa, 11] Kempa, O., Lasota, T., Telec, Z., Trawiński, B.: Investigation of bagging ensembles of genetic neural networks and fuzzy systems for real estate appraisal. N.T. Nguyen et al. (Eds.): ACIIDS 2011, LNAI 6592, pp. 323–332, Springer, Heidelberg (2011)

[Kosko, 94] Kosko, B.: Fuzzy systems as universal approximators. IEEE Transactions on Computers 43:11, 1329–1333 (1994)

[Król, 08] 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)

[Kuncheva, 04] Kuncheva L.I.: Classifier ensembles for changing environments, Proceedings 5th International Workshop on Multiple Classifier Systems. In F. Roli et al. (Eds.) MCS2004, LNCS 3077, pp. 1–15, Springer, Heidelberg (2004)

[Lasota, 11a] Lasota, T., Telec, Z., Trawiński, B., and Trawiński K.: Investigation of the eTS Evolving Fuzzy Systems Applied to Real Estate Appraisal. Journal of Multiple-Valued Logic and Soft Computing, 17:2-3, 229–253 (2011)

[Lasota, 11b] Lasota, T., Telec, Z., Trawiński, G., Trawiński B.: Empirical Comparison of Resampling Methods Using Genetic Fuzzy Systems for a Regression Problem. In H. Yin et al. (Eds.): IDEAL 2011, LNCS 6936, pp. 17–24, Springer, Heidelberg (2011)

[Lughofer, 08] Lughofer, E.: FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models. IEEE Transactions on Fuzzy Systems 16:6, 1393–1410 (2008)

[Lughofer, 11a] Lughofer, E.: Evolving Fuzzy Systems - Methodologies, Advanced Concepts and Applications. Studies in Fuzziness and Soft Computing, vol. 266, Springer, Heidelberg (2011)

[Lughofer, 11b] Lughofer, E., Trawiński, B., Trawiński, K., Kempa, O, Lasota, T.: On Employing Fuzzy Modeling Algorithms for the Valuation of Residential Premises. Information Sciences 181, pp. 5123–5142 (2011)

[Maloof, 04] Maloof, M.A., Michalski, R.S.: Incremental learning with partial instance memory. Artificial Intelligence 154:1-2, pp. 95–126 (2004)

[Minku, 10] Minku, L.L., White, A.P., Yao, X.: The Impact of Diversity on Online Ensemble Learning in the Presence of Concept Drift, IEEE Transactions on Knowledge and Data Engineering 22:5, pp. 730–742 (2010)

[Sobolewski, 12] Sobolewski, P., Woźniak, M.: Data with Shifting Concept Classification Using Simulated Recurrence. J.-S. Pan, S.-M. Chen, N.T. Nguyen (Eds.): ACIIDS 2012, LNAI 7196, pp. 403–412, Springer, Heidelberg (2012)

[Trawiński, 12a] Trawiński, B., Lasota, T., Smętek, M., Trawiński, G.: An Attempt to Employ Genetic Fuzzy Systems to Predict from a Data Stream of Premises Transactions. In E. Hüllermeier et al. (Eds.): SUM 2012, LNAI 7520, pp. 127–140. Springer, Heidelberg (2012)

[Trawiński, 12b] Trawiński, B., Lasota, T., Smętek, M., Trawiński, G.: An Analysis of Change Trends by Predicting from a Data Stream Using Genetic Fuzzy Systems. N.-T. Nguyen et al. (Eds.): ICCCI 2012, Part I, LNAI 7653, pp. 220–229. Springer, Heidelberg (2012)

[Trawiński, 12c] Trawiński, B., Smętek, M., Telec, Z., Lasota, T.: Nonparametric Statistical Analysis for Multiple Comparison of Machine Learning Regression Algorithms. International Journal of Applied Mathematics and Computer Science (AMCS) 22:4, pp. 867–881 (2012)

[Tsymbal, 04] Tsymbal, A.: The problem of concept drift: Definitions and related work. Technical Report. Department of Computer Science, Trinity College, Dublin (2004)

[Wang, 03] Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: Getoor, L., et al. (Eds.) KDD-2003, pp. 226–235, ACM Press, New York (2003)

[Widmer, 96] Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Machine Learning 23, pp. 69–101 (1996)

[Zliobaite, 09] Zliobaite, I., Kuncheva, L. I.: Determining the Training Window for Small Sample Size Classification with Concept Drift. In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops ICDMW '09, IEEE Computer Society, Washington, DC, USA, pp. 447–452 (2009)