Proceedings of the 2nd World Congress on Electrical Engineering and Computer Systems and Science (EECSS'16)
Budapest, Hungary - August 16 - 17, 2016
Paper No. MVML 108
DOI: 10.11159/mvml16.108

Takagi-Sugeno-Kanga Fuzzy Fusion In Dynamic Multi-Classifier System

Maciej Krysmann[1]
[1]Department of Systems and Computer Networks, Wroclaw University of Technology
27 Wybrzeze Wyspianskiego Street, Wroclaw, Poland
maciej.krysmann@pwr.edu.pl

Abstract - In this paper, the approach to implementation of Takagi-Sugeno-Kanga fuzzy system into the Dynamic Ensemble Selection multi-classifier. Paper presents DES system with its working idea, provides in-depth information about that system. Dynamic creation of classifier ensemble, which selects classifiers for particular classified object x has proven its advantages. It is shown that described TSK system can improve classification quality even better, even in situation in which base classifiers are not fully trained. Proposed rule set and for TSK system is described. Paper presents complete algorithm with pointing all phases of work. Experimental study presents positive results and prove proposed system advantages basing on well known UCI Machine Learning benchmark databases. Paper also is discussing real life situation in which system can be used, however also points out classification time increase.

Keywords: Dynamic Ensemble Selection, Takagi-Sugeno-Kanga, Fuzzy Systems

1. Introduction

Nowadays classification gives manny possibilities, impossible to be done without it. It is well known how advantageous classification was and is for medicine, energetics, data processing and manny more. However in manny cases single or separate classifier cannot deal with task, due to its own limitations. That situation can be overcome by building multi-classifier systems. That approach can make possible creation of EEG based control systems for bioprosthetics, allow to analyse big amounts of data. Thanks to multi-classifiers there is possible to easily integrate data which come from different sources, or have different nature. Moreover due to its improvement possibility in fusion part of multi-classifier system and thanks to repealing single classifiers mistakes an ensemble can provide significantly better results. It is very important nowadays, because in manny cases classifier is trained basing only on part of data, which does not give full view for particular problem, in the other words, classifier is trained on set which does not fully represent feature space.

2. Multi-classifier Systems

It is well known that in classification task, to increase classification quality, multi-classifier systems can be build [3]. Thanks to that, few weak classifiers can work together in order to obtain classification quality higher than each of them separately. Those systems often work on paradigm of having base classifiers trained before building multi-classifier. Then their final decision can be made by directly combining knowledge from each of them - it is called Majority Voting or by adjustable fusion algorithm. Those second one could be divided in two main approaches:

- Static - multi-classifier is build once, fusion algorithm is adjusted to the data during learning process
- Dynamic - multi-classifier is build separately, from base classifiers, for each classified object, during classification (classifier work) phase.

Dynamic Ensemble Selection main assumption is that there is trained ensemble of base classifiers $\Psi = \{\psi_1, \psi_2, ..., \psi_l\}$, from which each can be part of final decision making ensemble. Those classifiers have been trained earlier and they are not modified during DES system building or work. However there is available part of training set (with correct labels) called the validation set, which allow to perform dynamic classifier selection before classification process for classified object. Comprehensive description of dynamic multi-classifier systems could be found in [1]. It is important to mention that performing dynamic selection of classifiers allow us to create ensemble which should be specialised for classification of particular object x. That action prevents form several issues such as local minima problem or prevent from voting made by classifier which is unable to match decision boundary in particular part of feature space. Dynamic approach can be divided into two groups:

- Individual based - each classifier is evaluated separately, then if it meet assumed condition classifier gets into ensemble. Grading of classifier can be made on several basis:
  - Ranking,

- Accuracy,
- Probabilistic,
- Behaviour,
- Oracle.

- Group based - classifiers are selected in subgroups, evaluated how they perform as such subgroup on base of:
  - Diversity,
  - Ambiguity,
  - Data Handling.

In previous author's works [6, 5] selection was made individually according to only probabilistic approach developed and described in [2]. In this approach there is as masure of classifier work the competence presented. Classifier after being evaluated to acquire competence can be used in ensemble if competence value is high enough. Dynamic Ensemble Selection based on Randomized Reference Classifier [2], have proven its advantages, therefore author is developing system based on RRC by implementing Fuzzy Systems in DES schema [4] and also by replacing RRC by custom measures evaluated through various fuzzy systems.

3. Dynamic Ensemble Selection System

Dynamic Ensemble Selection is utilising RRC in order to acquire competence set, which is then generalised through learning procedure (Figure 1). It is needed because during classification process there is need of having competence value for classified object, what is impossible to get directly. RRC DES system requires validation set (Equation 1), where x_N is feature

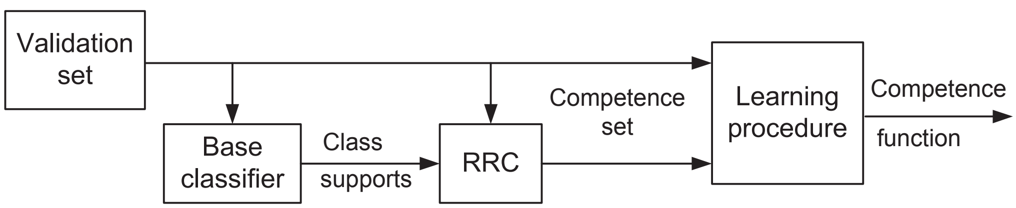

Fig. 1: Use of validation set to perform dynamic ensemble selection

vector of object form validation set and j_N is correct class of x_N. That set is needed in order to be able to evaluate each of classifier ψ_l. That requirement implies necessity of having enough big training set. In testing environment validation set is separate part of training set, which is invisible to classifiers during their learning process.

$$\mathscr{V} = \{(x_1, j_1), (x_2, j_2), \ldots, (x_N, j_N)\}; \; x_k \in \mathscr{X}, \; j_k \in \mathscr{M} \tag{1}$$

Using validation set classifier is evaluated by RRC in order to get competence values for objects from $\mathscr{V}$. This new set, which contains objects (points in feature space) and calculated competence values is called competence set (Equation 2), and it is created separately for each classifier ψ_l from ensemble Ψ.

$$\mathscr{C}_l = \{(x_1, C(\psi_l|x_1)), (x_2, C(\psi_l|x_2)), \ldots, (x_N, C(\psi_l|x_N))\}. \tag{2}$$

3.1. Randomized Reference Classifier

The RRC is stochastic classifier - build as equivalent of evaluated classifier ψ_l. It is defined through probability distribution over the product of class supports $[0,1]^M$. It can be said that the RRC uses the maximum rule and a vector of class supports $[\delta_1(x), \delta_2(x), \ldots, \delta_M(x)]$ for the classification of the feature vector x, where the j-th support is a realisation of a random variable (rv) $\Delta_j(x)$. The rvs probability distributions are satisfying (index l of the classifier ψ_l and its class supports is dropped for better clarity, while E is expected value operator):
(1) $\Delta_j(x) \in [0,1]$;
(2) $E[\Delta_j(x)] = d_j(x), \; j = 1, 2, \ldots, M$;
(3) $\sum_{j=1,2,\ldots,M} \Delta_j(x) = 1$,
Above equations shows that RRC is equivalent of classifier ψ because it produces for feature vector x the same vector of class supports.

The RRC performs classification in a stochastic manner, so the probability of classification an object x to the i-th class, can be calculated as:

$$P^{(RRC)}(i|x) = Pr[\forall_{k=1,\dots,M,\ k\neq i}\ \Delta_i(x) > \Delta_k(x)]. \tag{3}$$

In particular, if the object x belongs to the i-th class, from (3) we simply get the conditional probability of correct classification $Pc^{(RRC)}(x)$.

Most important in above approach is the choice of probability distributions for the rvs $\Delta_j(x), j \in \mathscr{M}$ to satisfy 1-3 conditions. In this paper beta probability distributions are used with the parameters $\alpha_j(x)$ and $\beta_j(x)$ $(j \in \mathscr{M})$, according to the [2].

To get the probability of correct classification of RRC at a point $x_k \in V$, we apply the RRC to a validation point x_k and assume for (3) $i = j_k$:

$$\begin{aligned} Pc^{(RRC)}(x_k) \quad &= \quad \int_0^1 b(u, \alpha_1(x_k), \beta_1(x_k)) \\ & \quad\quad [\prod_{j=2}^{M} B(u, \alpha_j(x_k), \beta_j(x_k))]du, \end{aligned} \tag{4}$$

where $B(.)$ is a beta cumulative distribution function.

Randomized Reference Classifier is considered as equivalent to the base classifier $\psi_l \in \Psi$, thus we can use the probability (4) as the competence of the classifier ψ_l at the validation point $x_k \in \mathscr{V}$:

$$C(\psi_l|x_k) = Pc^{(RRC)}(x_k). \tag{5}$$

As result there is competence set $\mathscr{C}_l$ (2) created. Howevers values of competence are calculated only for objects (points) form validating set $\mathscr{V}$ (1). In order to perform classifier evaluation for particular object x which will be classified (if classifier is competent enough), there is need of approximation (generalisation) of competence set $\mathscr{V}$ for considered point x using all validation points $x_k \in \mathscr{V}$. Previous works [5, 6] describes research on competence generalisation methods, so called methods of classifier competence learning. Various methods was considered resulting in choice of potential function (6) as both fast and accurate.

$$C(x) = \frac{\sum_{k=1}^{N} C_k * K(x, x_k)}{max(C_k)}$$

where

$$K(x, x_k) = e^{-d(x, x_k)} \tag{6}$$

In previous works fusion (selection) was performed using threshold [2] or then moving threshold [5, 6]. In first approach classifier was regarded as competent when its competence value for object x was greater than $\frac{1}{\mathscr{M}}$, where $\mathscr{M}$ is number of classes in particular classification task.

4. Takagi-Sugeno-Kanga Fuzzy System

Fuzzy systems lead out from fuzzy sets theory presented by L.A.Zadeh [8] in 1965. This theory assumes that object can partially belong to particular set. Fact of partially belonging of object to the set is described by membership function $\mu_A : \mathscr{X} \rightarrow [0, 1]$. Fuzzy set is described as pairs $(x, \mu_A(x))$, that way: $A = \{(x, \mu_A(x))|x \in [0, 1]\}$. Over fuzzy sets are defined multiplication (t-norm) and sum (s-norm) most often as: $min[\mu_A(x), \mu_B(x)]$ and $max[\mu_A(x), \mu_B(x)]$. On that base Fuzzy reasoning systems were created. They allow to utilise linguistic knowledge to perform classification. Knowledge base is stored in IF-THEN rules, which can be obtained from human expert, or through extraction from numeric data. Most popular fuzzy reasoning system is Mamdani System. However in this work will be used Takagi-Sugeno-Kanga system, which in opposite to Mamdani System which uses in conclusion fuzzy sets, TKS system utilises functions. In most cases this functions are linear functions. For TSK system typical rule looks like that: $\mathscr{R}$: IF $(X_1$ is $A_1)$ and $(X_2$ is $A_2)$ and ... and $(X_n$ is $A_n)$ THEN $y = f(X)$. Since in this approach there is need of two input variables final function will have shape: $y = a(competence) + b(diversity) + c$. Rules list are described in compact form of triplets $[a, b, c]$ in Table 1.

Table 1: Rules list for TSK system.

Competence / Diversity	Competent	Neutral	Incompetent
High	[1,0,0]	[0.3,0.6,0.1]	[0.4,0.4,0.2]
Neutral	[0.6,0.2,0.2]	[0.3,0.4,0.3]	[0.3,0.5,0.2]
Low	[0.7,0.3,0]	[0.4,0.6,0]	[0,0,0]

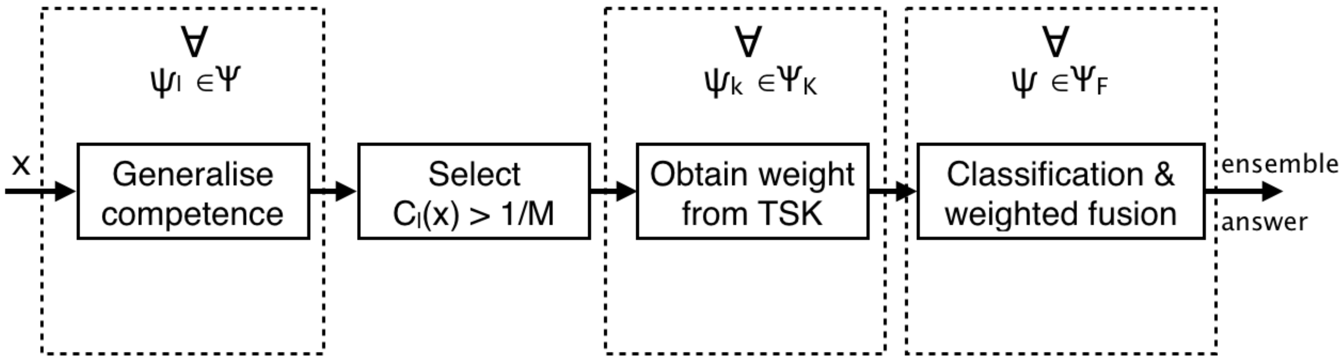

Fig. 2: General system shema

5. TSK Fussion In DES

In paper Takagi-Sugeno-Kanga system is utilised in fusion of base classifiers thus it creates weights for competent classifier outputs. General system schema is presented on Figure 2.

Let denote $C_l(x)$ competence of classifier ψ_l for object we want to classify x (1 on Figure 3) . Then we create subset $\Psi_K(x)$ of base classifiers which are better than random classifier, it means their competence is higher than $\frac{1}{\mathscr{M}}$, where $\mathscr{M}$ is number of classes in classification task(2 on Figure 3):

$$\Psi_K(x) = \{\exists_{k \in l} \psi_k \Rightarrow C_k(x) > \frac{1}{\mathscr{M}}\} \tag{7}$$

Next $\Psi_K(x)$ is sorted in descending order (3 on Figure 3) and there is ψ_n denoted as classifier with highest competence. Subsequently the for $k = 2$ process begins, lets denote Ψ_F as selected for final ensemble classifier set, on the beginning looks like that: $\Psi_F = \psi_n$. Fuzzy system (5 on Figure 3) is implemented with two input variables and one output variable. First input is, for classification of object x, considered classifier ψ_k competence $C_k(x)$. Second input is diversity, calculated between selected classifiers Ψ_F and ψ_k: $Div(\Psi_F, \psi_k)$, using Kohavi-Wolpert [9] method (4 on Figure 3). As TSK system output we get w_k weight which will be used for support vector from ψ_k weighting in final fusion. Then ψ_k is added to Ψ_F, what gives $\Psi_F = \{\psi_n, \psi_k\}$, k is incremented. Process repeats until last classifier from $\Psi_K(x)$.

5.1. Experiments

There had been performed experimental evaluation of proposed TSK Fuzzy Fusion in DES system. Proposed system was compared with base DES RRC system described in [2]. Comparison was performed on well known benchmark datasets (see Table 2) from UCI Machine Learning Repository [10]. For each database there had been performed 2CV test 10 times, then results were statistically tested, for obtaining statistically significant differences. Moreover the same base classifiers were used, in particular repetition, for both compared systems.

Base classifier set was heterogenous and build as not fully trained (to obtain better increase in quality) classifiers as follows:

- 3,5,15-Nearest Neighbour,
- Nearest Mean,

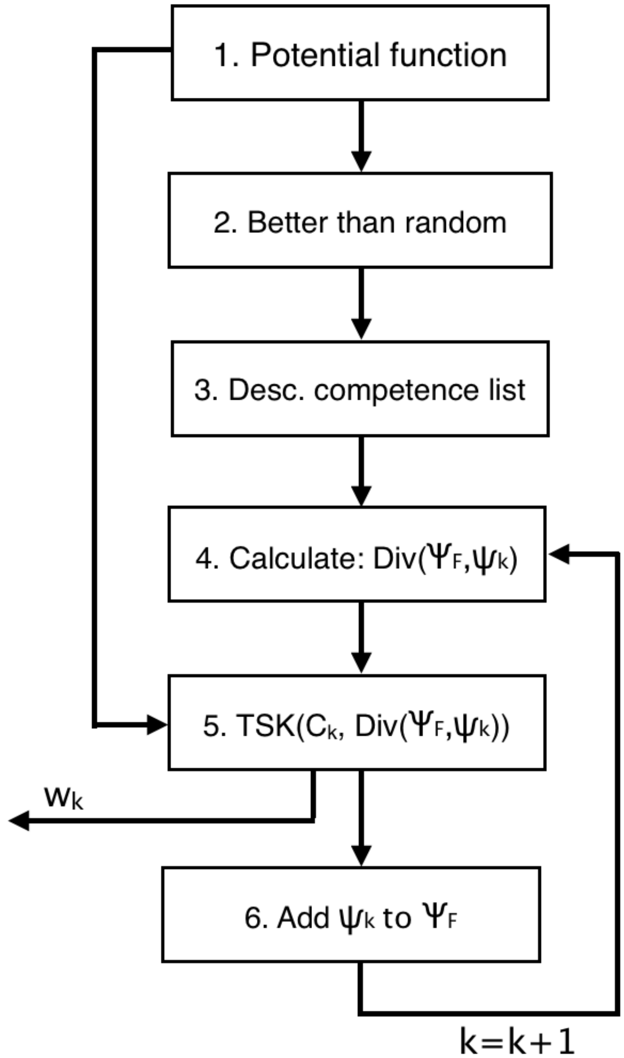

Fig. 3: TSK in DES approach

Table 2: Databases used in evaluation

Database	Features	Instances	Classes	Database	Features	Instances	Classes
EColi	34	336	6	Glass	9	214	6
Ionosphere	34	351	2	Haberman	3	306	2
Iris	4	150	3	Derma	34	366	6
Wine	13	178	3	Yeast	8	1484	10
Pima	8	768	2				

- Parzen Density Classifier: with h_{opt} and with $\frac{h_{opt}}{2}$,
- Classification Tree with Gini splitting parameter,
- Neural Network: 1 hidden layer with 10 neurones, 2 hidden layers with 5 neurones each.

This set of heterogenous classifiers was used in previous authors works, thus keeping set the same makes easier to compare other systems. Experimental evaluation was performed in Matlab environment using own authors code, PRToools toolbox [11] and built-in Matlab toolboxes.

6. Results And Discussion

Results as classification average quality are presented in Table 3. Comparison was made between DES RRC - its advantage over classical multi-classifier is known from papers [2] and proposed Takagi-Sugeno-Kanga Fuzzy Fusion Dynamic Ensemble System - TSK-FDES. Overall quality values are not the best possible to acquire in classification task, because base classifiers were weakly trained, in order to present differences better. Moreover testing with this approach gives very important information - if base classifiers are weak, presented approach can improve results significantly.

Table 3: Evaluation results

Database	DES RRC	TSK-FDES
EColi	85.22	84.76
Ionosphere	86.62	90.07
Iris	95.35	95.30
Wine	86.31	87.23
Pima	61.3	73.45
Glass	56.76	71.64
Haberman	61.75	79.92
Derma	87.63	97.86
Yeast	56.9	60.54

It can be easily seen that proposed system is not always better than DES RRC, especially in "easy" classification tasks (such as for example Iris database). However for the "harder" classification tasks there is great improvement in classification quality. It proves that even having weak base classifiers there is possibility of improvement, by advanced fusion and weighting approach. Moreover dynamic selection of classifiers and their dynamic weighting is performed before classification as itself. It is possible because of dynamic nature of proposed system which can tune-up itself for each classified object, moreover with taking into account current state of system - by comparing candidate classifier with one selected previously (for classified object x). However implementing that approach in DES system increases classification time, so the classifier will be working slower than before. Although it is small price for that significant improvement.

7. Conclusion

Paper presents successful approach to implementation Takagi-Sugeno-Kanga fuzzy system into Dynamic Ensemble Selection multi-classifier. Presented algorithm has proven its advantages, especially in hard classification problems in which previous approach was defiantly worse. That improvement is important, because experimental evaluation was performed in weakly trained base classifiers environment. Despite that proposed algorithm improved quality, what can be approximated on situation in which classifiers during learning phase didn't obtained full feature space data. That situation is common in real life classification tasks nowadays. Some data can be lost, or simply didn't shown in training phase (incoming internet data). However it is important to notice, that implementing TSK system into fusion block of TSK-FDES system is generating increase of work time of whole system. Moreover due to limitations system would not be efficient for Big-Data problem. Nevertheless proven by this paper advantage of implementation of fuzzy systems into the Dynamic Ensemble Selection schema opens up new ways of improvement of DES systems and wider development of that area is already in progress.

Acknowledgments

This work is financed from Grant For Young Scientists and PHd Students Development, under agreement: B50069/K0402.

References

[1] Alceu S. Britto Jr., Robert Sabourin, Luiz E.S. Oliveira, *Dynamic selection of classifiersA comprehensive review*, Pattern Recognition, Volume 47, Issue 11, November 2014, Pages 3665-3680,

[2] Woloszynski T., Kurzynski M. (2011) *A probabilistic model of classiffier competence for dynamic ensemble selection* Pattern Recognition 44 2656-2668.

[3] Kuncheva I. (2004). *Combining Pattern Classiffiers: Methods and Algorithms* Wiley-Interscience

[4] Kurzyski M., Krysmann M. *Fuzzy inference methods applied to the learning competence measure in dynamic classifier selection*, 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2014: Rio de Janeiro, Brazil, 27-30 August 2014 Los Alamitos, Ca. cop. 2014. s. 180-187,

[5] Krysmann M.,Kurzynski M. (2012). *On new methods of dynamic ensemble selection based on randomized reference classifier* Journal of Medical Informatics Technologies, vol. 20, 101107.

[6] Krysmann M.,Kurzynski M. (2013). *Methods of Learning Classifier Competence Applied to the Dynamic Ensemble Selection*, Advances in Intelligent Systems and Computing Volume 226, 151-160.
[7] Lysiak R., Kurzynski M., Woloszynski T.(2011). *Probabilistic approach to the dynamic ensemble selection using measures of competence and diversity of base classiffers* Lecture Notes in Artiffcial Intelligence 6679, 229-236
[8] Zadeh L.A. *Fuzzy sets. Information and Control* 1965; 8: 338353.
[9] Kuncheva, L. I. and Whitaker C.J. *Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy.*,Machine Learning, 51(2):181207, 2003.
[10] UCI Machine Learning Repository *http://archive.ics.uci.edu/ml/*.
[11] PR Tools Machine Learning Matlab Toolbox *http://www.prtools.org/*.