A Fuzzy Set Tool in the Classification and Prediction Software System (CLAPSS)

Extended Abstract*

Krzysztof Pancerz and Jaromir Sarzyński

Department of Computer Science, Faculty of Mathematics and Natural Sciences
University of Rzeszów
Prof. S. Pigonia Str. 1, 35-310 Rzeszów, Poland
kpancerz@ur.edu.pl

Abstract. In the paper, we give the outline of a fuzzy set tool implemented in the Classification and Prediction Software System (CLAPSS). CLAPSS is being developed for solving different classification and prediction problems using, among others, some specialized approaches based mainly on fuzzy sets and rough sets which are not available in other machine learning software systems. Theoretical background as well as the module embedded in CLPASS, for fuzzification of attribute values in information/decision systems, are described. Moreover, possible further steps in the usage of CLAPSS (generation of fuzzy decision trees as well as fuzzy flow graphs) are mentioned.

Key words: Fuzzy sets, Fuzzification, Software system, CLAPSS.

1 Theorethical Background

Most of the methods implemented in the CLAPSS system are applied for information/decision tables representing information/decision systems understood as Pawlak's knowledge representation systems (cf. [12]).

A decision system is a tuple $DS = (U, C, D, \{V_a\}_{a \in C \cup D}, f_{inf}, f_{dec})$, where U is the non-empty, finite set of objects, C is the non-empty, finite set of condition attributes, D is the non-empty, finite set of decision attributes, $\{V_a\}_{a \in C \cup D}$ is the family of non-empty sets of condition and decision attribute values, $f_{inf} : C \times U \to \bigcup_{c \in C} V_c$ is the information function such that $f_{inf}(c, u) \in V_c$ for each $c \in C$ and $u \in U$, $f_{dec} : D \times U \to \bigcup_{d \in D} V_d$ is the decision function such that $f_{dec}(d, u) \in V_d$ for each $d \in D$ and $u \in U$. An information system is a specific case of a decision system. In this case $D = \emptyset$ and the decision function is not defined. Further, only information systems, in the form $IS = (U, A, \{V_a\}_{a \in A}, f_{inf})$, will be considered.

Fuzzification is the process that transforms the real value variables into linguistic variables whose domains contain linguistic values which can be described by fuzzy sets (their membership functions). Let $IS = (U, A, \{V_a\}_{a \in A}, f_{inf})$ be an information system such that $V_a \subseteq \mathbb{R}$ for each $a \in A$. For each attribute $a \in A$, we can define a linguistic variable λ_a. With each linguistic variable λ_a, a set $L^{\lambda_a} = \{l_1^{\lambda_a}, l_2^{\lambda_a}, \ldots, l_{k_a}^{\lambda_a}\}$ of linguistic values is associated. Each linguistic value $l_i^{\lambda_a}$, where $i = 1, 2, \ldots, k_a$, is described by a membership function $\mu_{l_i^{\lambda_a}} : \mathbb{R} \to [0, 1]$. In CLAPSS, the user has a broad set of membership functions which can be used to make a fuzzification process. This set consists of:

- a triangular shaped membership function,
- a trapezoidal shaped membership function,
- a Gaussian shaped membership function,
- a generalized bell shaped membership function,
- an S shaped membership function,
- a π shaped membership function,
- a sigmoidal shaped membership function,
- a fuzzy singleton membership function,
- a sinusoidal shaped membership function,
- a Z shaped membership function,
- a pseudo-exponential shaped membership function,
- an L-R shaped membership function,
- a two Gaussian shaped membership function,
- a D-sigmoidal shaped membership function,
- a P-sigmoidal shaped membership function.

Let:

- $IS = (U, A, \{V_a\}_{a \in A}, f_{inf})$ be an information system, where $card(U) = n$ and $card(A) = m$, such that $V_a \subseteq \mathbb{R}$ for each $a \in A$,
- $\{L^{\lambda_a}\}_{a \in A}$ be the family of sets of linguistic values associated with linguistic variables from the family $\{\lambda_a\}_{a \in A}$ defined for attributes from A, where $L^{\lambda_a} = \{l_1^{\lambda_a}, l_2^{\lambda_a}, \ldots, l_{k_a}^{\lambda_a}\}$ for each $a \in A$.

A fuzzified information system is a tuple $\mathcal{F}(IS) = (U^{\mathcal{F}}, \Phi, \{V_\phi\}_{\phi \in \Phi}, f_{inf}^{\mathcal{F}})$, where $U^{\mathcal{F}}$ is the non-empty, finite set of objects such that each $u^* \in U^{\mathcal{F}}$ corresponds exactly to one $u \in U$, $\Phi = \Phi_{a_1} \cup \Phi_{a_2} \cup \cdots \cup \Phi_{a_m}$ is the non-empty, finite set of fuzzified attributes, $\{V_\phi\}_{\phi \in \Phi}$ is the family of sets of fuzzified attribute values, $f_{inf}^{\mathcal{F}} : \Phi \times U^{\mathcal{F}} \to \bigcup_{\phi \in \Phi} V_\phi$ is the information function such that $f_{inf}^{\mathcal{F}}(\phi_{l_i^{\lambda_{a_j}}}, u^*) \in V_\phi$ for each $\phi_{l_i^{\lambda_{a_j}}} \in \Phi$ and $u^* \in U^{\mathcal{F}}$, $f_{inf}^{\mathcal{F}}(\phi_{l_i^{\lambda_{a_j}}}, u^*) = \mu_{l_i^{\lambda_{a_j}}}(f_{inf}(a_j, u))$, where $\mu_{l_i^{\lambda_{a_j}}}$ is a membership function describing $l_i^{\lambda_{a_j}}$.

2 CLAPSS

CLAPSS is our tool developed for solving different classification and prediction problems using, among others, some specialized approaches based mainly on fuzzy sets and rough sets. CLAPSS is equipped with the graphical user interface (see Figure 1). In general, our main idea is to implement in CLAPSS those specialized approaches which are not available in other machine learning software systems. Selected functionalities of CLAPSS were earlier described in [7], [8], and

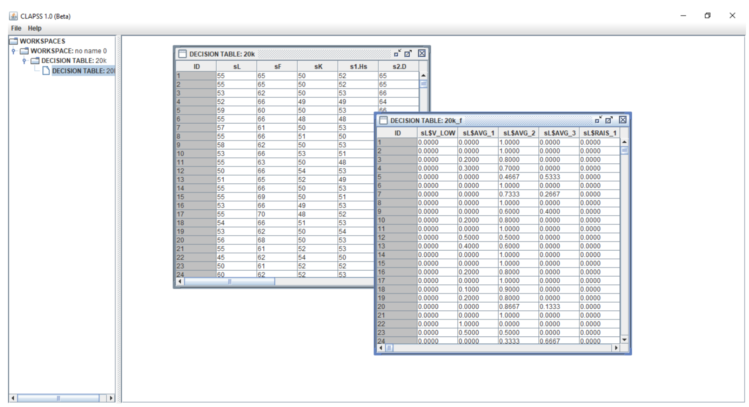

Fig. 1. CLAPSS (the graphical user interface).

[11].

The general usage of methods, based on fuzzy sets, implemented in CLAPSS is shown in Figure 2. Information/decision systems (also those fuzzified) can be imported from/exported to other machine learning software systems, RSES [1], WEKA [4], ORANGE [2], as it was depicted in Figure 2.

For methods based on fuzzy sets, CLAPSS offers, first of all, a tool for fuzzification of attribute values in information/decision systems. The fuzzification process can be done in three ways: graphical, scripting, and external.

For graphical fuzzification, Membership Function Creator (MFC) has been developed (see Figure 3). MFC enables the user to:

- determine linguistic values and membership functions (their shapes and parameters) associated with them,
- manually modify membership functions created earlier (for example, characteristic points or slopes can be moved),
- see calculated values of the fuzzified attribute (these values are automatically updated if some changes in membership functions are made).

After the fuzzification process of the selected attribute, a script (a special scripting language was designed for CLAPSS) is generated. The script, consisting of membership function definitions for each attribute to be fuzzified, can also be created manually, i.e., in a scripting way (see an example below).

```
ATTR[0]->fuzzification(lingvalues={low=(trapezoidal,0.0000,0.0000,1.0000,3.0000),
medium=(triangular,1.0000,2.5000,4.0000),high=(trapezoidal,2.0000,4.0000,5.0000,5.0000)});
```

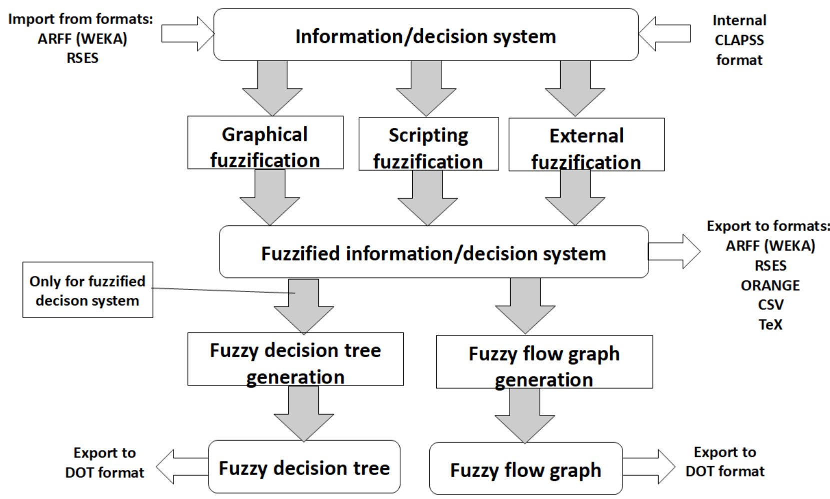

Fig. 2. CLAPSS (a general scheme of the usage of the fuzzy set tool).

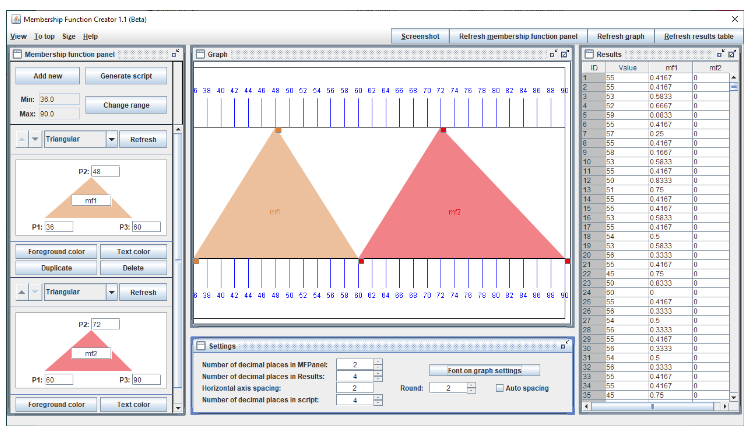

Fig. 3. CLAPSS (Membership Function Creator).

```
ATTR[1]->fuzzification(lingvalues={low=(trapezoidal,0.0000,0.0000,1.0000,3.0000),
medium=(triangular,1.0000,2.5000,4.0000),high=(trapezoidal,2.0000,4.0000,5.0000,5.0000)});
ATTR[2]->fuzzification(lingvalues={low=(trapezoidal,0.0000,0.0000,1.0000,3.0000),
medium=(triangular,1.0000,2.5000,4.0000),high=(trapezoidal,2.0000,4.0000,5.0000,5.0000)});
```

Fuzzification of attribute values in the external tool is also possible. Then, the user can import a fuzzified information/decision system into CLAPSS.

Further steps which can be performed in CLAPSS for fuzzified information/decision systems can be as follows:

– Generation of fuzzy decision trees. Fuzzy decision trees are generated using the algorithm based on cumulative information estimations of initial data [5].
– Generation of fuzzy flow graphs [11]. Fuzzy flow graphs are generated using the fuzzy cardinality (power) of linguistic values (cf. [6]).

The visualization of fuzzy decision trees and fuzzy flow graphs is possible due to the option for exporting them to the DOT format [3].

The practical usage of CLAPSS was presented in case of analysis and classification of MMPI (Minnesota Multiphasic Personality Inventory) data. Fuzzified decision systems were used among others in determining the importance of ranges of MMPI scales [10] and classification of MMPI profiles using fuzzy decision trees [9].

3 Conclusions

In the paper, we have briefly presented an important part of the CLAPSS system concerning implemented methods based on fuzzy sets. CLAPSS is constantly being developed. One of the main directions in further developing of CLAPSS, in the considered area, is to add other types of membership functions and to add graph algorithms for fuzzy flow graphs to extract rules, episodes, etc.

Acknowledgments

This work was partially supported by the Center for Innovation and Transfer of Natural Sciences and Engineering Knowledge at the University of Rzeszów.

References

1. Bazan, J.G., Szczuka, M.S.: The Rough Set Exploration System. In: Peters, J., Skowron, A. (eds.) Transactions on Rough Sets III, Lecture Notes in Artificial Intelligence, vol. 3400, pp. 37–56. Springer-Verlag, Berlin Heidelberg (2005)
2. Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., Zupan, B.: Orange: Data mining toolbox in Python. Journal of Machine Learning Research 14, 2349–2353 (2013)
3. Ellson, J., Gansner, E.R., Koutsofios, E., North, S.C., Woodhull, G.: Graphviz and Dynagraph - static and dynamic graph drawing tools. In: Jünger, M., Mutzel, P. (eds.) Graph Drawing Software, pp. 127–148. Springer, Berlin, Heidelberg (2004)
4. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
5. Levashenko, V., Zaitseva, E.: Usage of new information estimations for induction of fuzzy decision trees. In: Yin, H., Allinson, N., Freeman, R., Keane, J., Hubbard, S. (eds.) Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2002). LNCS, vol. 2412, pp. 493–499. Springer Berlin Heidelberg (2002)

6. Mieszkowicz-Rolka, A., Rolka, L.: Flow graphs and decision tables with fuzzy attributes. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Żurada, J.M. (eds.) Artificial Intelligence and Soft Computing (ICAISC 2006). LNCS, vol. 4029, pp. 268–277. Springer, Berlin, Heidelberg (2006)
7. Pancerz, K.: On selected functionality of the classification and prediction software system (CLAPSS). In: Proceedings of the International Conference on Information and Digital Technologies (IDT'2015). pp. 278–285. Zilina, Slovakia (2015)
8. Pancerz, K., Grochowalski, P., Paja, W.: On selected data preprocessing procedures with the classification and prediction software system (CLAPSS). In: Proceedings of the International Conference on Information and Digital Technologies (IDT'2016). pp. 219–226. Rzeszow, Poland (2016)
9. Pancerz, K., Levashenko, V., Zaitseva, E., Gomuła, J.: Experiments with classification of MMPI profiles using fuzzy decision trees. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) Proceedings of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS'2018), Annals of Computer Science and Information Systems, vol. 15, pp. 125–128. Poznań, Poland (2018)
10. Pancerz, K., Paja, W., Sarzyński, J., Gomuła, J.: Determining importance of ranges of MMPI scales using fuzzification and relevant attribute selection. In: Howlett, R., Toro, C., Hicks, Y., Jain, L. (eds.) Proceedings of the 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES-2018), Procedia Computer Science, vol. 126, pp. 2065–2074. Belgrade, Serbia (2018)
11. Pancerz, K., Lewicki, A., Sarzyński, J.: Discovering flow graphs from data tables using the classification and prediction software system (clapss). In: Mihálydeák, T., Min, F., Wang, G., Banerjee, M., Düntsch, I., Suraj, Z., Ciucci, D. (eds.) Rough Sets, LNCS, vol. 11499, pp. 356–368. Springer International Publishing, Cham (2019)
12. Pawlak, Z.: Rough Sets. Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)