Comparison of Similarity Measures in Context of Rules Clustering

Agnieszka Nowak-Brzezińska*, and Tomasz Rybotycki†
*Institute of Computer Science, Silesian University, Poland
Bankowa 12, 40-007 Katowice, Poland
Email: agnieszka.nowak@us.edu.pl
†IBS PAN, Doctoral Study
Newelska 6, 01-447 Warszawa, Poland

***Abstract*—This paper introduces five similarity measures, very well known in literature, but not because of using them to compare rules between themselves and choose the most similar one. Rules in knowledge bases are a very specific type of data representation and it is necessary to compare them carefully. The goal of the paper is to analyze the influence of using different similarity measures on the number of clusters, or the size of the representatives of the created clusters of rules. The results of the experiments are presented in Section III in order to discuss the significance of the analyzed measures and methods of rules creating.**

***Index Terms*—similarity measures, rules clustering, hierarchical clustering, knowledge-based systems.**

I. Introduction

The aim of the paper is to present the results of the comparison of different similarity measures in the context of rules clustering [4], [6]. Rules (as a form of *if* ... *then* chains) are very specific type of data. They can be given apriori by a domain expert or generated automatically from dataset using some algorithm. If they are generated from a dataset, with some (usually large) number of objects from a given domain, they are usually short (small number of premises in their left hand side). That is because the algorithms used for their creating usually aim to find the representation which is simple and easy to interpret. Shorter lenghts of the rules require their carefully comparison. Therefore, it is very important to choose a proper similarity measure when we want to compare rules between themselves. Finding a pair of the most similar rules is a crucial step in every clustering algorithm which is necessary to be used when we would like to manage large set of rules in a given knowledge base (KB). Searching within groups of rules is much faster than searching every single rule, one by one - especially when the number of rules is really huge. The answer to the question ,,why is it so important to find an exact rule in a given KB as quickly as possible" is quite simple. Decision support systems (DSS), which are usually based on rule-based KBs, use rules to extract new knowledge - this process is called the inference process. Instead of searching every rule one by one it is possible to find the most relevant group of rules (the representative of such group of rules matches the given information in the best way) and decrease the time necessary to give an answer to a user of such a system. In [14] the authors propose modularization of large KBs using methods for grouping the rules. The results obtained in the authors's previous research [4] show that in an optimal case, it was possible to find a given rule when only few percent of the whole KB was searched. The quality of groups of rules' representatives is very important. That is why it is necessary to choose the best possible clustering algoritm and the algorithm which creates the optimal descriptions for groups of rules. An optimal represenatation of groups of rules is also important for other reasons. When groups of rules have a good representation it helps the knowledge-engineer to manage such KB and knowledge saved in rules (it is easier to find rules with a given premise and to remove or modify if it is necessary).

978-1-5090-5795-5/17/$31.00 © 2017 IEEE

The analysis of the rules' similarity can be based on either premises or conclusions of the rules. In this reasearch rules are divided into a number of groups based on similar premises. This approach is dictated by the fact that conditional parts of the rules are generally longer than their decisional parts and thus, make clustering more complex, accurate and interesting. Moreover, the authors direct their research toward the forward (data driven) inference process where the premises of the rules are the basis of searching. To minimize the number of rules that need to be processed before a knowledge based system (KBS) produces an answer to a given input, instead of searching within the whole set of rules (as it is in case of traditional inference processes), only representatives of groups would be compared with the set of facts and/or hypotheses to be proved. The most relevant group of rules is selected and an exhaustive searching is done only within this group. This way, given a set of rules, new knowledge may be derived using a standard forward chaining inference process, which can be described as follows: each cycle of deductions starts with matching the condition part of each rule with known facts. If at least one rule matches the facts asserted into the rule base, it is fired. It is really crucial to find and describe all the factors which influence clustering results and interference efficiency as it would help in designing or partitioning of KBs in order to maximize KBS's effectiveness.

Similarity measures play an important role in finding pairs of rules (or further, in the clustering process/forming the

groups of rules) and deciding about the order of clustering the rules (or groups of rules). Thus, to identify a suitable similarity measure, it becomes necessary to compare the results of using different measures and choose the one which allows to get the optimal results. For this reason the next section (Section II) describes the motivation behind using a hierarchical type of the clustering algorithm. It includes the description of a rule-based form of knowledge representation.

The rest of the paper is organized as follows. In Section II the description of cluster analysis algorithm, used for rules clustering, was described. This section focuses on both similarity measures used for comparring the rules between themselves and hierarchical clustering algorithms. Section III contains the results of the experiments while the Section IV presents the summary of the proposed approach.

II. Clustering Algorithms

A cluster is a set of objects in which each object is closer (or more similar) to every other objects in the cluster than to any object not in the cluster. Moreover, a clustering algorithm aims to find a natural structure or relationship in an unlabeled data set. There are several categories of clustering algorithms. Some of the algorithms are hierarchical and probabilistic. In the paper the authors present comparison experiments related to 5 different similarity measures (and 4 different clustering methods) of rules cliustering. The starting condition is carried out by setting every object as a separate cluster. In each step, the two most similar objects are merged and a new cluster is created with a proper representative for it. The classic hierarchical algorithm combines the two most similar elements until one group containing all the elements is created. In case of rules clustering, there is no reason (usually) to joining rules or groups of them of their similarity is to small, especially when they do not have any common premises and conclusions. That is why, the auhors proposed to stop the clustering when a given condition is met. There are many possible ways for defining the stop condition. For example, it can be reaching the specified number of groups, or reaching the moment in which the highest similarity is under minimal required threshold (which means the groups of rules are now more differential than similar to one another). The ultimate clustering process can be described as below: the goal of clustering is to maximize *intra-cluster* similarity and minimize *inter-cluster* similarity. Both the similarities, *intra-cluster* and *inter-cluster* are important but the goal of this research is to analyze the meaning of the first one.

The pseudocode of the hierarchical clustering algorithm - namely Classic AHC (agglomerative hierarchical clustering) algorithm [7] - is presented as Pseudocode 1.

Pseudocode 1. Classic AHC Algorithm.
Input: stop condition *sc*, ungrouped set of objects *s*
Output: grouped tree-like structure of objects

1) Place each object *o* from *s* into a separate cluster.
2) Build similarity matrix *M* that consists of every clusters pair similarity value.
3) Using *M* find the most similar pair of clusters and merge them into one.
4) Update *M*.
5) **IF** *sc* was met end the procedure.
6) **ELSE REPEAT** from step 3.
7) **RETURN** the resultant structure.

The main advantage of hierarchical clustering is that it does not impose any special methods of describing similarity between the clusters.

The most important step is the second one, in which the similarity matrix M is created based on the selected similarity measure and a pair of the two most similar rules (or groups of rules) are merged. In this step two parameters are given by a user: the similarity measure (intra-cluster similarity method) and the clustering method (inter-cluster similarity method). Eventually, both of them result in achieving different clusterings. For this reason the authors decide to compare similarity measures in this research. In order to do that, the authors have choosen nine different similarity measures and repeated the clustering algorithm many times for every similarity measure while changing the number of groups [1]. The results of this research are presented in Section III.

A. Rules Clustering Algorithms

Many papers show the results of clustering a large set of data [13] but rarely for such a specific type of data like rule-based knowledge representation. Clustering algorithms allow to organize the rules in a smart way [7]. To achieve groups of similar rules it is necessary to propose some method of deciding which rules are the most similar in a given step of the clustering process. Because rules are a specific type of data, usually attributed with short descriptions, the differences between rules are difficult to notice. Hence, it is so important to find a similarity measure which is able to find all the differences and, as a result, decides about the order of rules clustering in an optimal way.

Natural way of knowledge representation makes rules easily understood by experts and knowledge engineers as well as people not involved in the expert system building. In this research the authors have generated rules automatically from datasets, using an algorithm which builds so-called minimal rules [2]. Every rule contains two parts: conditional (with at least one premise) and decisional (usually with one conclusion). Sometimes (what makes the analysis more complicated) conclusion of one rule may be a condition in others. In this case it is said that such rules form a chain, and during the inference process they are all processed as a cause and effect chain. Sometimes the rules' attributes are weighted therefore the importance of some rules (given as an ordered set of attributes) is higher than others because of difference in weights of their attributes and/or their lengths. Moreover, the conditional and decisional part of a rule can be also treated differently from

[1] In this work clustering is stopped when given number of clusters is generated.

[2] It means that the rules achieved in this way, have got a short description and cover as many data from the original dataset as possible.

each other, for example conditional part may have a higher priority (greater weights for premises than for a conclusion). All these circumstances make the rules very specific kind of knowledge representation.

B. Similarity Analysis

In the literature there are numerous methods of describing similarity between objects [2] that can be modified to work with rules as well. The similarity measure used to find a pair of rules or groups of rules that are the most similar in a given moment is called the *intra-cluster similarity measure*. The authors studied the following five measures: Simple Matching Coefficient (SMC), based on it - the Jaccard Index sometimes also called the weighted similarity or the weighted similarity coefficient [5]. Gower measure (widely known in the literature) [12] and two inspired by the retrieval information systems: Occurence frequency measure (OF) and so-called inverse occurence frequency measure (IOF). It is crucial to answer the question if a given similarity measure influences the shape of grouped KB's structure. Measuring similarity or distance between two data points is a core requirement for several data mining and knowledge discovery tasks that involve distance computation. The notion of similarity or distance for categorical data is not as straightforward as for continuous data. When data consists of objects that aggregate both types at once the problem is much more complicated. It is necessary to find a measure. that could deal with this case.

1) Notations: Having a set of attributes A and their values V, rules premises and conclusions are built using pairs (a_i, v_i), where $a_i \in A, v_i \in V_a$. A pair (a_i, v_i) is called a descriptor. In a vector of such pairs, i-th position denotes the value of the i-th attribute of a rule. It is important to note that most of the rules do not consist of all attributes in A, thus, constructed vectors (describing the rules) are of different lengths. The frequency (f) of a given descriptor d is equal to 0 if d is not included in any of the rules from a given knowledge base KB (if $x \notin KB$ then $f(x) = 0$). A data set might contain attributes that take many values as well as attributes that take only few of them. A similarity measure might give more importance to attributes with smaller sets of values, while partially ignoring the others. It may also work the other way around. Almost all similarity measures assign a similarity value between two rules r_j and r_k belonging to the set of rules R as follows:

$$S(r_j, r_k) = \sum_{i=1}^{N} w_i s(r_{ji}, r_{ki})$$

where $s_i(r_{ji}, r_{ki})$ is the per-attribute (for i-th attribute) similarity between two values of descriptors of the rules r_j and r_k. The quantity w_i denotes the weight assigned to the attribute a_i and usually $w_i = \frac{1}{d}$, for $i = 1, \ldots, d$. In all the definitions, the s_{ijk} denotes the contribution provided by the k-th variable, and w_{ijk} is usually 1 or 0 depending upon whether or not the comparison is valid for the k-th variable; if differential variable weights are specified it is the weight of the k-th variable or 0 if the comparison is not valid.

The simplest measure is SMC (Simple Matching Coefficient) [3] - which calculates the number of attributes that match in the two rules in the following way:

$$s_{SMC}(r_{ji}, r_{ki}) = s_{jki} = 1 \texttt{if} r_{ji} = r_{ki} \texttt{ else } 0.$$

The range of the per-attribute SMC is $\{0; 1\}$. It treats all types of attributes in the same way. Unfortunately it tends to favour longer rules thus, it is better to use the $Jaccard$ measure, which is similar to SMC however it is more advanced as it also divides the result by the number of attributes of both objects so longer rules are not favoured any more. It can be defined in the following form:

$$s_{Jaccard}(r_{ji}, r_{ki}) = s_{jki} = \frac{1}{n} \texttt{ if } r_{ji} = r_{ki} \texttt{ else } 0$$

where n is number of attributes considered. Deriving from *retrieval information systems*, two measures could be also used to check the similarity between rules (or clusters of rules). The inverse occurrence frequency (IOF) measure assigns a lower similarity to mismatches on more frequent values while the occurrence frequency (OF) measure gives opposite weighting for mismatches when compared to the IOF measure, i.e., mismatches on less frequent values are assigned a lower similarity and mismatches on more frequent values are assigned a higher similarity. The definition of them is as follows:

$$s_{IOF}(r_{ji}, r_{ki}) = \begin{cases} 1 & \text{if } r_{ji} = r_{ki}; \\ \frac{1}{1+\log(f(r_{ji}))\cdot\log(f(r_{ki}))} & \text{if } r_{ji} \neq r_{ki}. \end{cases}$$

and

$$s_{OF}(r_{ji}, r_{ki}) = \begin{cases} 1 & \text{if } r_{ji} = r_{ki}; \\ \frac{1}{1+\log\frac{N}{f(r_{ji})}\cdot\log\frac{N}{f(r_{ki})}} & \text{if } r_{ji} \neq r_{ki}. \end{cases}$$

The $Gower$ similarity coefficient is the most complex of the all used inter-cluster similarity measures as it handles numeric attributes and symbolic attributes differently. For ordinal and continuous variables it defines the value of s_{jki} as $s_{jki} = 1 - \frac{|r_{ji} - r_{ki}|}{range(r_i)}$, where: $range(r_i)$ is the range of values for the i-th variable. For continuous variables s_{jki} ranges between 1, for identical values r_{ji} = r_{ki} and 0 for the two extreme values r_{max} - r_{min}. It is vital to mention that some of the aforementioned inter-object similarity measures are intended for categorical attributes only. In these cases, a similarity between the numerical attributes of two rules is calculated in the same way as in Gower's measure. If r_i is numeric then the value of s_{jki} is calculated in the following way:

$$s_{Gower}(r_{ji}, r_{ki}) = s_{jki} = 1 - \frac{|r_{ji} - r_{ki}|}{range(r_i)}$$

If r_i is categorical and $r_{ji} = r_{ki}$ then $s_{jki} = 1$ else it is equal to 0.

[3] If both compared objects have the same attribute and this attribute has the same value for both objects then add 1 to a given similarity measure. If otherwise, do nothing. To eliminate one of the problems of SMC, which favours the longest rules, the authors also used the Jaccard Index.

TABLE I
DATA GATHERED DURING EXPERIMENTS.PART I.

	Total	arythmia	audiology	autos	balance
ClustersN	16.5 ± 14.1 4-49	12.5 ± 2.5 10-15	6.9 ± 3.0 4-10	7.8 ± 2.4 4-10	19 ± 9.1 10-28
AttrN	$5.84E+01 \pm 93$ 5-280	280	70	26	5
RulesN	234 ± 176 42-490	154	42	60	290
NodesN	452 ± 343 74-970	295.5 ± 2.5 293-298	77.2 ± 3.0 74-80	112.2 ± 2.4 110-116	560 ± 9.1 550-560
BCS	155 ± 144 $12-480$	111.5 ± 42.0 $23-145$	30.0 ± 7.9 $12-39$	37.9 ± 14.3 $12-57$	170 ± 95 $21-280$
BRL	35.4 ± 50.4 $3-150$	147.4 ± 1.9 $145-154$	66.8 ± 0.5 $66-68$	10.7 ± 0.6 $10-12$	4.0 ± 0.01 $4-4$
U	6.9 ± 9.0 $0-41$	5.8 ± 4.7 0-14	3.6 ± 2.7 0-9	3.8 ± 3.0 0-9	6.9 ± 9.0 0.01-27
BRS	36.4 ± 51.81 4-170	152.1 ± 4.9 147-165	67.0 ± 0.5 66-68	11.6 ± 1.8 10-18	4.0 ± 0.01 4-4
ARS	29.9 ± 45.6 2.4-150	134.0 ± 11.1 101.3-148.9	49.8 ± 9.1 29.8-64	8.5 ± 1.7 5.7-11.5	3.6 ± 0.43 3-4
wARS	$2.2 \pm 8.45E-01$ 1.1-4.6	2.1 ± 0.2 1.9-2.8	1.5 ± 0.32 1.1-2.4	3.2 ± 0.6 2.3-4.6	1.4 ± 0.15 1.3-1.6

TABLE II
DATA GATHERED DURING EXPERIMENTS.PART II.

	Total	Breast cancer	diab	diabetes
ClustersN	16.5 ± 14.1 4-49	11 ± 1.0 10-12	29 ± 19 10-48	29.5 ± 19.7 10-49
AttrN	$5.84E+01 \pm 93$ 5-280	$10 \pm 0.0E-01$ 10-10	9.0 ± 0.1 9-9	9.0 ± 0.0 9-9
RulesN	234 ± 176 42-490	130 ± 0.1 130-130	480 ± 0.1 480-480	490 490-490
NodesN	452 ± 343 74-970	240 ± 1.0 240-240	940 ± 19 920-960	950.5 ± 19.7 931-970
BCS	155 ± 144 $12-480$	76 ± 32 $21-120$	320 ± 130 $37-470$	336.6 ± 135.2 $45-479$
BRL	35.4 ± 50.4 $3-150$	9.0 ± 0.01 $9-9$	4.9 ± 0.46 $3-5$	4.9 ± 0.3 $4-5$
U	6.9 ± 9.0 $0-41$	5.1 ± 3.4 0.01-10	11 ± 13 0.1-36	12.5 ± 14.4 0-41
BRS	36.4 ± 51.81 4-170	9.0 ± 0.01 9-9	5.4 ± 0.49 5-6	5.5 ± 0.6 5-7
ARS	29.9 ± 45.6 2.4-150	7.1 ± 1.1 5.4-8.9	3.4 ± 0.77 2.4-4.9	3.3 ± 0.8 2.4-4.8
wARS	$2.2 \pm 8.45E-01$ 1.1-4.6	1.4 ± 0.22 1.1-1.9	2.8 ± 0.63 1.8-3.8	2.8 ± 0.7 1.9-3.8

III. EXPERIMENTS

In this section, an experimental evaluation of 5 similarity measures on 7 different KBs [3] is presented. Decision rules were generated from the original data using RSES software and $LEM2$ algorithm [1]. The smallest number of attributes was 5, the greatest 280. The smallest number of rules was 42, the greatest 490. All the details of analyzed datasets are included in Tables I and II.

The meaning of the columns in Tables I, II, III and IV is following:

$AttrN$ - number of different attributes occuring in permises or conclusions of rules in given KB.
$RulesN$ - number of rules in examined KB.
$ClustersN$ - number of nodes in dendrogram representing resultant structure.
U - number of singular clusters in resultant structure of grouping.
BRS - biggest representative size - number of descriptors used to describe longest representative.
ARS - average representative size - average number of descriptors used to describe cluster's representatives.
$wARS$ - weighted average representative size (AttrNumber) - division of average number of descriptors used to describe cluster's representative in give data set and number of attributes in this data set.
BRL - biggest representative length - number of descriptors in biggest cluster's representative.
BCS - biggest cluster size - number of rules in the cluster that contains the most of them.

TABLE III
CHARACTERISTICS OF CLUSTERS' REPRESENTATIVES VS. SIMILARITY MEASURES.

p	Ns	Ns	Ns	Ns	Ns	Ns
sim	BCS	BRL	U	BRS	ARS	wARS
Gower	157.23 ± 147.43	35.68 ± 51.20	8.05 ± 9.92	36.38 ± 52.26	30.72 ± 47.12	2.22 ± 0.91
	$18.0 - 479.0$	$4 - 154$	$0 - 41$	$4 - 159$	$2.39 - 146.47$	$1.1 - 3.9$
IOF	157.89 ± 145.83	35.27 ± 50.56	7.96 ± 10.07	36.30 ± 51.65	29.75 ± 45.75	2.26 ± 0.92
	$14 - 479$	$3 - 147$	$0 - 41$	$4 - 156$	$2.39 - 144.9$	$1.1 - 4.6$
OF	157.68 ± 146.59	35.16 ± 50.43	7.73 ± 10.08	36.25 ± 51.51	28.69 ± 44.69	2.26 ± 0.89
	$12 - 479$	$3 - 147$	$0 - 41$	$4 - 155$	$2.39 - 147.4$	$1.1 - 4.2$
SMC	155.63 ± 143.62	35.43 ± 50.78	5.57 ± 6.72	36.32 ± 52.43	30.20 ± 45.57	2.11 ± 0.75
	$12 - 479$	$4 - 151$	$0 - 31$	$4 - 165$	$2.70 - 148.9$	$1.1 - 4.2$
Jaccard	145.16 ± 141.09	35.36 ± 50.69	5.21 ± 7.71	36.59 ± 52.77	30.49 ± 46.66	2.06 ± 0.75
	$12 - 477$	$4 - 148$	$0 - 41$	$4 - 160$	$2.70 - 145$	$1.1 - 4.2$
Total	154.72 ± 143.97	35.38 ± 50.37	6.91 ± 9.03	36.37 ± 51.75	29.97 ± 45.64	2.18 ± 0.85
	$12 - 479$	$3 - 154$	$0 - 41$	$4 - 165$	$2.39 - 148.9$	$1.1 - 4.6$

The performance of different similarity measures was evaluated in the context of knowledge mining using informations like: the number of rules clusters ($ClustersN$), the number of ungrouped rules (U - Ungrouped objects), the sizes of the biggest cluster (BCS - Biggest cluster size) as well as its representative (BRS - Biggest representative size) and the representative the most specific (BRL - Biggest representative length). More specific means more detailed, containing a higher number of descriptors.

The optimal structure of KBs with rules clusters should contain the well separated groups of rules, and the number of such groups should not be too high. Moreover, the number of ungrouped rules should be minimal. Creating an optimal description of each cluster (representative) is very important because they are used further to select a proper group (and reject all the others) in the inference process, in order to mine knowledge hidden in rules (by accepting the conclusion of the given rule as a true fact). The results of the experiment have verified the initial hypotheses about the inter and intra cluster similarity measures. As can be seen in Tables I and II no single measure is always superior or inferior. This is obvious since each KB has different characteristics (different number of attributes and/or rules) as well as different types of attributes. The use of some measures however, guarantees achieving more general or more specific representatives for the created rules clusters. There are some pairs of measures that exhibit complementary performance, for example one performs well where the others perform poorly and vice-versa.

Figures 1 and 2 show that the decision on which similarity measure is used for clustering, influences the size of the cluster of rules' representative and the number of ungrouped rules.

Figure 1 shows that using the $Jaccard$ or the SMC measure for clustering the rules (or groups of rules) results in a smaller number of ungrouped rules than it is when we use the other measures.

Figure 2 shows that for SMC measure as well as for its modification ($Jaccard$ measure) the weighted size of the representative is smaller than for the rest of the proposed measures: $Gower$ as well as the IOF and the OF measure.

It was also crucial to check if there is a correlation (statistically significant) between the size of the KB (the

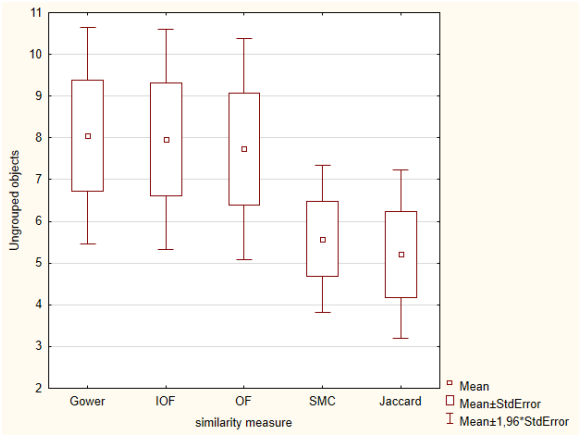

Fig. 1. Similarity measures vs. the number of ungrouped rules.

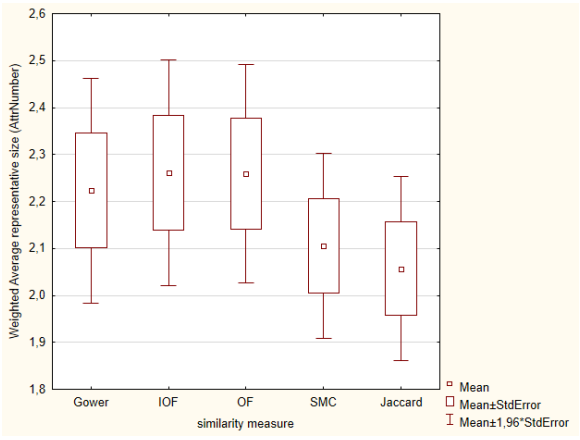

Fig. 2. Similarity measures vs. the weighted average representative's size.

number of rules, number of attributes etc.) and the parameters analyzed in this research such as: biggest cluster size, biggest representative length or the number of ungrouped rules etc. The results of this research are included in Table IV.

It can be seen in Table IV that there is a statistically significant correlation between the number of attributes in the KB ($AttrN$) and the length of the biggest representatives

TABLE IV
CORRELATION ANALYSIS FOR THE KNOWLEDGE BASE'S SIZE AND THE CHARACTERISTICS OF THE CLUSTERS' REPRESENTATIVES.

	AttrN	RulesN	NodesN
ClustersN	-0.2012 *	0.6269*	0.6015*
BCS	-0.2297*	0.8181*	0.8255*
BRL	0.9779*	-0.4026*	-0.4024*
U	-0.0965, $p = 0.107$	0.3475*	0.3302*
BRS	0.9801*	-0.3966*	-0.3966*
ARS	0.9837*	-0.3705*	-0.3704*
wARS	-0.0791, $p = 0.187$	0.3222*	0.3220*
* - statistically significant	$p < 0.05$		

(the more attributes we have the longer the representative's length). Another, quite obvious, correlation is the one between the number of rules and the number of ungrouped rules, which is positive also (the more rules to cluster the higher the number of rules that are dissimilar to the others and impossible to be merged with others). The results of the experiment also show the correlations that are not statistically significant - which shows that there is no correlation in a given data. Such a correlation is to determine for two pairs of variables: the number of ungrouped rules (U) and the number of attributes ($AttrN$) as well as the weighted average representative size ($wARS$) and the number of attributes.

IV. CONCLUSION

Readibility of produced rulesets is the main reason of clustering the rules. Buiding an informative descriptions of rules (their representatives) plays an important role in the process of an efficient exploration of the knowledge stored in the structure based on rules' clustering. The content of a given rules cluster's representative depends heavily on the set of rules that forms it. If a given cluster is build from the rules which are not too much similar to eachother, then the representative of it is not good (contains the features which do not describe every rule in a group). The results of the clustering strongily depends on the many clustering parameters as inter- and intra-cluster similarity measures. That is why in this article, the authors present the evaluation of five different similarity measures used for comparring the results of clustering of rules in KBs. The experiments have been carried out for seven different KBs from various domains and such datasets differ in many parameters. The rules have been clustered withn the AHC algorithm presented in Section II. The results obtained in the experiments are included in Tables I, II, III, IV and in figures 1,2. The most important conlusions, that can be taken from the experiments, are the following. No single measure is always superior or inferior, however, the use of some measures, guarantees achieving more general or more specific representatives for the created rules clusters. For SMC and $Jaccard$ measure, the weighted size of the representative is smaller than for the other analyzed measures ($Gower$, IOF and OF). There is a statistically significant correlation between the number of attributes in the KB ($AttrN$) and the length of the biggest representatives (the more attributes we have the longer the representative's length). Another, quite obvious, correlation is the one between the number of rules and the number of ungrouped rules, which is positive also (the more rules to cluster the higher the number of rules that are dissimilar to the others and impossible to be merged with others).

REFERENCES

[1] Bazan J.G., Szczuka M.S., Wróblewski J. A new version of rough set exploration system. Rough Sets and Current Trends in Computing. Springer-Verlag, Berlin, pp. 397 404, 2002.

[2] Boriah S., Chandola V., Kumar V. Similarity Measures for Categorical Data: A Comparative Evaluation. Proceedings of the 8th SIAM International Conference on Data Mining, pp. 243-254, 2008.

[7] Dubes R., Jain A.K. Clustering techniques: The user's dilemma. Pattern Recognition. vol. 8, nr 4, 1976.

[9] Goodall D.W. A new similarity index based on probability. Biometrics. vol.22, pp. 882-907, 1966.

[12] Gower J.C. A general coefficient of similarity and some of its properties. Biometrics, vol.27, International Biometric Society, Washington, pp.857-871, 1971.

[8] Lee O., Gray P. Knowledge base clustering for KBS maintenance. Journal of Software Maintenance and Evolution, vol.10, nr 6, pp. 395-414, 1998.

[3] Lichman M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], University of California, 2013.

[4] Nowak-Brzezińska A. Mining rule-based knowledge bases. Advanced Technologies for Data Mining and Knowledge Discovery, CCIS, Springer, Volume 613, pp. 94-108, 2016.

[5] Nowak-Brzezińska A., Rybotycki T. Visualization of medical rule-based knowledge bases. Journal of Medical Informatics & Technologies, Vol.24, pp. 91-98, 2015.

[6] Nowak-Brzezińska A. Mining Rule-based Knowledge Bases Inspired by Rough Set Theory, Fundamenta Informaticae 148, pp. 3550, 35, DOI 10.3233/FI-2016-1421, IOS Press, 2016.

[13] Przybyła-Kasperek M., Wakulicz-Deja A. Global decision-making system with dynamically generated clusters. Information Sciences Volume 270, 172191, 2014.

[14] Simiński R. Multivariate approach to modularization of the rule knowledge bases, chapter in: ManMachine Interactions 4. Series: Advances in Intelligent Systems and Computing, vol.391, Springer, 2016.

[10] Wierzchoń S. T., Kłopotek M.A. Algorithms of Cluster Analysis. Wydawnictwo IPI PAN, Warsaw, 2015.