Linguistic Summaries Generation with Hybridization Method Based on Rough and Fuzzy Sets

Iliana Pérez Pupo[1](✉) , Pedro Y. Piñero Pérez[1] ,
Rafael Bello[2] , Luis Alvarado Acuña[3] ,
and Roberto García Vacacela[4]

[1] Departamento de Investigaciones en Gestión de Proyectos,
Universidad de las Ciencias Informáticas, 54830 La Habana, CP, Cuba
{iperez,ppp}@uci.cu
[2] Centro de Investigación en Informática, Universidad Central Marta
Abreu de las Villas, Santa Clara, Cuba
rbellop@uclv.edu.cu
[3] Departamento de Ingeniería de la Construcción, Universidad Católica
del Norte, Antofagasta, Chile
lualvar@ucn.cl
[4] Facultad de Especialidades Empresariales, Universidad Católica de Santiago
de Guayaquil, Guayaquil, Ecuador
roberto.garcia@cu.ucsg.edu.ec

Abstract. In this paper authors propose a new algorithm for linguistic data summarization based on hybridization of rough sets and fuzzy sets techniques. The new algorithm applies rough sets theory for feature selection in early stages of linguistic summaries' generation. The rough sets theory was used to reduce on significant way, the amount on summaries obtained by others algorithms. The algorithm combines lower approximation, k grade dependency and fuzzy sets to get linguistic summaries. The results of proposed algorithm are compared with association rules approach. In order to validate the algorithm proposed, authors apply both qualitative and quantitative methods. Authors used two databases in order to validate the algorithm; theses databases belong to "Repository of Project Management Research". The first database is associated to personality traits and human performance in software projects. The second database is associated to analysis of revenue assurance in different organization. Considering quantitative approach, the algorithm proposed, obtains better results than the algorithm based on association rules; while regards execution time, the best algorithm was the algorithm based on association rules, because rough sets theory was high time-consuming technique.

Keywords: Linguistic data summarization · Rough sets · Project management · Human resources

1 Introduction

Linguistic data summarization was first introduced by Yager in 1982's [1] and it has been applied in real scenarios such as: autonomy of things, medicine and others real scenarios. Different investigations associated to this technique have been developed in the last two decades following three main work lines:

- Conceptualization of linguistic summaries and its structure.
- Indicators to evaluate the quality of linguistic summaries.
- Algorithms to generate linguistic summaries from data.

About the structure, the summaries are classified considering different "protoforms" [2, 3]:

- Classic protoforms, to summarize attributes [1, 4].
- Time series protoforms [5].
- Events representation protoforms [6].

But the most used are protoforms with the following syntax:

- Overviews whose structure is *Qy's are S*, which describe relationships such as the following:
  T(Most employees have low pay) = 0.7.
- Summaries structured as *QRy's are S*, describing relationships such as:
  T(Most young employees have low pay) = 0.7.

Kacprzyk and Zadrożny classified in [7] six protoforms that described the structure of summaries and the queries for their search, see Table 1.

Table 1. Classification of protoforms of LDS [7].

Type	Protoform	Given	Sought
0	QRy's are S	All	Validity T
1	Qy's are S	S	Q
2	QRy's are S	S and R	Q
3	Qy's are S	Q and structure of S	Linguistic values in S
4	QRy's are S	Q, R and structure of S	Linguistic values in S
5	QRy's are S	Nothing	S, R and Q

About the indicators to measure the quality of the summary, several authors have been proposed different *T* indicators. For example, in [1] Yager proposed six indicators, called as *T* values *T1, T2, T3, T4, T5, T6,* as follows:

- Degree of truth (*T1*): called the measure of validity of the summary, provides an indication of how compatible the linguistic summary is with the database.
- Degree of imprecision (*T2*): is important validity criterion, measure of both uncertainty and vagueness concepts. This indicator depends on the form of the summary, not on the database.
- The degree of coverage (*T3*): measures how many objects in database are supported for linguistic summary.
- Degree of appropriateness (*T4*): This degree describes how characteristic is the summary for the particular database. It degree permits to distinguish between trivial summaries, having full validity (truth), and really important summaries. The summary found reflects an interesting, not fully excepted relation in our data [8].
- The length of an overview (*T5*) measure of the length or summaries, how many elements conform the summary.
- An indicator for resume the quality evaluation of a particular linguistic overview (*T6*), is defined as the weighted average of the previous 5° of validity.

There are other measures, for example in [5] Kacprzyk and Wilbik proposed a set of indicators specifically for time series scenarios. In [9], the authors proposed a set of indicators to extend the Yager's indicators based on degree of indeterminacy information in summaries. About the algorithms to generate summaries there are different trends too, such as:

- Linguistic summaries generated form sql queries [10].
- Generation of summaries by using association rules [11, 12].
- Generation of summaries through meta-heuristics [13].
- Generation of summaries by using clustering techniques [6].
- Other approaches that combine previous works [14–17].

But most of algorithms, to generate summaries reported in bibliography, not use appropriately, the information associated to the attributes relationships. In order to improve the summaries' generation methods, authors of this work proposed a new algorithm for linguistic data summarization based on hybridization of rough sets and fuzzy sets.

This work is organized in the following sections. The second section presents a brief analysis of rough sets concepts and its adoption in the algorithm proposed. The third section presents the results of algorithm in a human resource problem. Finally, the conclusions of the work are presented.

2 A New Algorithm for Linguistic Data Summarization

In linguistic summaries generation is very important to discover the attributes relationships. The linguistic summaries consist on filters and summarizers, in general these components can be represented from an information system $S = (U, A \cup D)$, where

filters belong to set A while summarizers belong to decision attributes D. In this sense, the authors of this paper propose the application of rough sets theory to discover the attributes relationships. The authors adopted some concepts of rough sets theory in the new algorithm, in the next paragraphs we explain main concepts of this theory.

The rough sets theory was proposed in 1986 by Pawlak for application in data inconsistency. Usually, rough sets are used in two alternatives: to discreet data [18] based on equivalence relationships or to extended indiscernibility relationships [19, 20]. Different extensions of rough sets applications were reported in [18, 21, 22].

Given an information system $S = (U, A \cup D)$, let $X \subseteq U$ a set of objects and $B \subseteq A$, a selected set of attributes, from the information contain in B, X can be approximate like following:

- The lower approximation of X with respect to B is:

$$B_*(X) = \{x \in U : B(x) \subseteq X\} \tag{1}$$

- The upper approximation of X respect to B is:

$$B^*(X) = \{x \in U : B(x) \cap X = \Phi\} \tag{2}$$

- The boundary region we can define as:

$$BNB(X) = B*(X) - B*(X) \tag{3}$$

- The negative region of decision d with respect to B is:

$$NEGB(X) = U - B*(X) \tag{4}$$

- Indiscernibility relation: defines an equivalence relation $INDB$ [23, 24], and this relation is denoted by:

$$IND(B)B = \{(x, y) \in U \times U : a(x) = a(y) \text{ for every } a \in B\} \tag{5}$$

- The positive region of decision d with respect to B is:

$$POS_B(d) = \cup\{B_*(X) : X \in U/IND(d), d \in D\} \tag{6}$$

Other useful concept useful is k grade dependency, that we explain in next paragraph.

Definition 1: Intuitively, a set of decision attributes D, depends totally on a set of B attributes, denoted by $B \Rightarrow D$, if all the values of the D attributes are univocally determined by the values of the attribute in B. In other words, D depends totally on B, if there is a functional dependency between the values of D and B [23]. D depends on B in a k grade where $k \in [0,1]$, and denoted by $B \Rightarrow_k D$, see Eq. (7). If $k = 1$ then D depends totally on B, while if $k < 1$ then D depends partially on B.

$$k = \frac{|POS_B(D)|}{|U|} \tag{7}$$

Where:

$$POS_B(D) = \bigcup_{X \in \frac{U}{D}} B_*(X) \tag{8}$$

2.1 LDS_RoughSet Algorithm

In this section we propose an algorithm for the construction of linguistic summaries of data, generating them from hybridization of rough sets and association rules. The authors established a linguistic variable associated to quantifiers for the construction of the summaries, see Fig. 1:
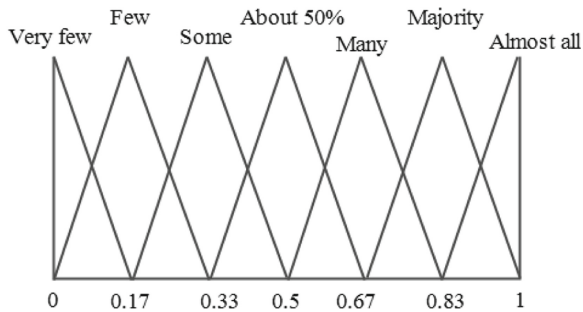

Fig. 1. Linguistic variable associated to quantifiers.

The proposed algorithm and its parameters are presented below.

```
Algorithm's name: LDS_RoughSet.
  Inputs:
    Information system S = (U, A ∪ D)
    U: dataset for analysis.
    A, D: attributes in the information system.
    Fuzzifiers: Set of fuzzy variables and its fuzzy sets
      for each attribute aᵢ ∈ A ∪ D.
    αₖ: alpha  cut  k  to  limit  the  low  the  rough  set
      dependency.
    Q: linguistic variable that represent the quantifiers
      of the summaries.
  Begin
    1. Transform S = (U, A ∪ D) in linguistic dataset by
       considering fuzzifiers set and maximal membership
       principles.
    2. A_ItemSet = attributes_to_sets(A∪D)
    3. D_ItemSet = A_ItemSet
    4. Stackset.push(A_ItemSet)
    5. CS = {}
    6. while not Stackset.empty
     6.1.   A_ItemSet = Stackset.pop
     6.2.   for each B ⊆ A_ItemSet
     6.3.      for each X ⊆ D_ItemSet: B ∩ X = ∅
     6.4.
```

$$POS_B(D) = \bigcup_{X \in U/D} B_*(X)' \quad k = \frac{|POS_B(D)|}{|U|}, \quad B \Rightarrow_k D, \ k \geq \alpha_k$$

```
     6.5.         for each Oₜ(B,X)∈ POSB(D)
     6.6.            if (there is total dependency
                              B⇒ₖ₌₁X in Oₜ(B,X))
     6.7.               CS = CS ∪ Oₜ(B,X)
     6.8.            else
     6.9.               Stackset.push({B ∪ X})
     6.10.           end
                  end
               end
            end
         end while
    7. LSUM = BuildSummaries(CS, Q)
    8. Calculate T values for each summary s ∈ LSUM
    9. Reorder LSUM considering T values
    10. Return LSUM
```

In step 2, each element $a \in A \cup D$ it is transformed in a set {a} and insert into `A_ItemSet` set, example: `{a₁, a₂, a₃}=>{{a₁},{a₂},{a₃}}`.

In step 4, a stack called *Stackset* is used for working with item sets to build the summaries.

In step 5, *CS* variable is a set to storage the candidate summaries.

In step 6.4 the positive region $POS_B(D)$ is calculated according to Definition 1 and Eqs. (7), (8). The $k \geq \alpha_k$ condition, help to find summaries between partial dependency attributes, but considering the sets of attributes that have minimum relationship level. Later, in 6.6 step, $O_t \in U$, $O_t(B,X)$ represents the values of attributes $a \in B \cup X$ in O_t object; if there is total dependency in O_t context, then $O_t(B, X)$ attributes values will be used as candidates summaries.

Afterward, in 6.9 step the algorithm searches other attributes combination to generate summaries with more filters. But no superset attributes of any low dependency attributes set, should be considered for summary generation. In order to prune the attributes combination, just must be consider the O_t objects that $O_t \in POS_B(D)$.

Finally, the summaries are sorted according to *T* values and then, submitted for evaluation of an experts group in this thematic area. The active learning method is used to identify and validate the best summaries from semantic point of view. This step is important in the final selection of summaries for decision-making.

3 Results

In order to validate the algorithm, authors applied qualitative and quantitative methods as follow in this section. Authors consider project management scenarios because projects are practically organized in all areas of society with a high social and economic impact [26, 27]. The demonstration of linguistic summaries applicability in project management help to show the high applicability of these algorithm in wide areas of human activity.

The qualitative evaluation of the algorithm was based on discovering of relationships between personality traits and human performance in software projects, see Sect. 3.1. Particularly in software projects, human resources are the main resource because these projects depend in large extent on professional skills, creativity and motivation of their resources.

The quantitative evaluation of the algorithm was based on the comparison of *LDS_RoughSet* algorithm with an algorithm to generate linguistic summaries based on association rules (*LDS_AssociationRules*), see Sect. 3.2.

3.1 Qualitative Evaluation Based on Real Case Study Application

In this section authors present the application of algorithm to discovering of relationships between personality traits and human performance in software projects. The principal motivations were:

- In order to achieve an adequate selection and conformation of teams, it is important to elaborate a personality profile according to the daily situations of the personnel. Personality is composed of several cognitive characteristics and behavioral trends that determine the similarities and differences in thoughts, feelings and behaviors of individuals [11].
- There are several tests that are used extensively and for multiple purposes. Most of these tests were standardized in correspondence with different populations and, there is some consensus on the application and interpretation of results that they provide.
- In order to acquire human resources for software project, authors consider the combination of: sociological, technical and quality of life test, is very important.
- In general, it is considered that, the characterization of these resources with respect to learning styles and personality traits, is essential for the formation of balanced teams, for the increase of efficiency and effectiveness in the development of projects. But most research only focuses on explaining the importance of psychological characteristics´ analysis, but does not establish mechanisms to identify relationships between personality traits and job performance in projects.

The experiment was applied to a population with 62 professionals for whom information on job performance is available in different roles and projects over a period of 3 years. Each person in the experiment population complete four questionnaires to known about its personality traits [28]:

- Instrument: Questionnaire on Leadership Styles.
- Instrument: Questionnaire on leadership styles using word computation.
- Instrument: Personality Inventory 16 PF Form C [29].
- Instrument: BFQ, Big Five Questionnaire [30].

The information obtained for each test was extended with the performance evaluations of the respondents and authors conform four datasets [31]. Finally, the algorithm was applied to each dataset and the following linguistic summaries were obtained:

Results in analysis of dataset "An Instrument Questionnaire on Leadership Styles":

1. The specialists with high performance in programmer role are characterized by being passive and task oriented. In addition, they can perform tasks in architect role.
2. The project members with high performance in a third role as implementer are characterized as passive people in normal conditions and can perform tasks in programmer role.
3. The specialists with an average performance in a second role as analyst are people-oriented under normal conditions and can perform tasks in programmer role.

Results in analysis of dataset "B Instrument: Questionnaire on leadership styles using word computation":

1. Programmers with high performance are characterized by being passive and people-oriented under both normal and stressful conditions. Under normal conditions, they have a technical mix, so they consider themselves to be exact, precise, calm and logical people, they complete important tasks following proven methods and do not like to take risks.
2. The specialists who work in second role as architects, with high performance, are characterized by being passive in stressful conditions, and can perform tasks in programmer role.
3. The specialists who work in second role as analysts, with average performance, under normal conditions are passive and people-oriented. Under stress conditions they are also passive. They can perform tasks as programmers.
4. The specialists who work in a third role as implanter, with a medium performance, are people-oriented under normal conditions and tasks-oriented under stress conditions.
5. The specialists who work in quality with high performance are people-oriented under normal conditions and tasks-oriented under stressed conditions.

Results in analysis of dataset "C Instrument: Personality Inventory 16 PF Form C".

1. The specialists with high performance in programmer role work in group and strengthen their ego. In a group, they consider themselves suspicious, complicate themselves and they act with premeditation. In addition, they have a lot of strength in their ego, so they are characterized by being emotionally stable, calm, mature, realistic, balanced and able to maintain solid group morale.
2. The specialists who work in quality with high performance, agreed as group on normal values in animation, sensitivity, abstraction and socialization.
3. High implanter: They point to cunning, are considered cunning, calculating, insightful, subtle and lucid people. His approach is intellectual and unsentimental.
4. The specialists with an average performance in a second role as analysts, agreed as group on normal values in animation, apprehension or security and socialization.
5. The specialists with high performance in a second role as architect, agreed as group on normal values in perfectionism and attention to standards.

Results in analysis of dataset "D Instrument: BFQ, Big Five Questionnaire":

1. The specialists with high performance in programmer role, are characterized by being moderately meticulous, precise, responsible, orderly and able to master their emotions. They are also unsympathetic and tolerant.
2. The specialists with high performance in quality tasks are characterized by being moderately creative, informed and open to cultural interests. Also they are quite responsible, orderly, cooperative and affectionate.
3. Professionals with a high degree of implanting competence are considered responsible and orderly people and can take on programmer role.
4. The specialists with high performance in a second role as architects are characterized by being moderately responsible, orderly and diligent. With little peace and quiet and patience. In addition, they are very inactive and can perform tasks in programmer role.

5. The specialists with high performance in a second role as analysts are characterized by being moderately creative, knowledgeable, understanding and tolerant. In addition, they have some positive bias in their responses, so they tend to deny their personal shortcomings or they are particularly naive.
6. Specialists with average performance in a second role as analysts are characterized by being moderately creative, informed, meticulous, precise, open to new ideas and values different from their own. Also, they are unsympathetic, tolerant and affectionate.

From the analysis of the summaries obtained, it is identified that the runs carried out on the different databases yielded the following common results:

1. The specialists with high performance in programmer role, agreed as group on normal values in animation, security, self-sufficiency and extroversion. In addition, they are passive people under normal conditions.
2. The specialists with average performance in a second role as analysts, agreed as group on normal values in animation, safety, extroversion and attention to standards.

These results were presented to the respondents, and they were asked to evaluate them without specifying the test that generated them. It was concluded that the majority of respondents considered the Big Five test to be the most appropriate for their personal characteristics. On the other hand, the Management Styles Test, that use computer with words techniques for the evaluation, gave better results than the variant using discrete variables. This research results were used in processes of acquisition and formation of software development teams in the organization where this research was applied.

In this investigation, it was possible to identify the characteristics associated to personality traits, as well their relationship with high performance in a given role; that is why we can predict, with some certainty, from the results of a personality analysis test, in which position a new employee will have better results. However, we emphasize that these results must be combined with professional skills for the correct assignment of roles.

3.2   Quantitative Evaluation from Comparison with Other Algorithm

In this section authors compare *LDS_RoughSet* algorithm and *LDS_AssociationRules* algorithm based on association rules techniques. Authors apply the two algorithms to database "200226_gp_eval_proyfinal" from "Repository of Project Management Research" [31]. This database contains 202 records and 8 attributes: 4 nominal attributes and 4 numeric attributes. Each record represents an organization with information about province location, economical affectation types and amount of economical affectation in several moneys.

For comparison, authors propose the following metrics: number of summaries obtained, execution time and a set of statistical metrics for each *T* value. For each *T* value (*T1, T2, T3, T4 T5*) of linguistic summaries authors calculate: mean, standard deviation, minimum and maximum values.

Table 2. Linguistic summaries evaluation using *T* indicators.

Algorithm		*T1*	*T2*	*T3*	*T4*	*T5*
LDS_RoughSets	Min	0.968	0.027	0.833	0.005	0.25
	Max	0.99	0.067	1	0.97	0.5
Execution time 2.6987 ms	StdDev	0.0058	0.011	0.029	0.31	0.062
Summaries amount 46	Mean	0.987	0.031	0.99	0.72	0.48
LDS_AssociationRules	Min	0.775	0.027	0.338	0	0.063
	Max	0.99	0.257	1	0.88	0.5
Execution time 2.2607 ms	StdDev	0.053	0.068	0.14	0.31	0.147
Summaries amount 76	Mean	0.96	0.057	0.94	0.56	0.409

Table 2 shows that the algorithm proposed obtains better results than the algorithm based on association rules regards the following metrics: summaries' amount, *T1, T3, T4* and *T5*. But the algorithm based on association rule is better in execution time; because rough sets theory is high time-consuming. There are not significant differences between algorithms regards *T2* indicator.

The analysis of *T* indicators revel that in future works traditional *T* indicators could be extended by considering other elements like indeterminacy and falsity.

4   Conclusions

The proposed procedure allows the identification of relationships between personality traits and the performance evaluation index in the roles assigned in software projects. This method has been tested in software projects, but its conception allows its application in several scenarios.

In the application of personality instruments, participants reported Big Five questionnaire as suitable instrument for their characteristics.

In the experiment it was found that specialists with high performance in programmer role, is characterized by being moderately meticulous, accurate, responsible, orderly and able to master their emotions. However, they are not very tolerant and project managers need to be aware of these characteristics in order to facilitate communication within the project and avoid interpersonal conflicts.

The research results allow the identification of personal characteristics of professionals that facilitate the communication with them of managers and avoid conflicts in work teams.

It is also identified that personality traits suitable for analyst role are creative, informed, meticulous, precise, open to new things, ideas and values different from their own and these characteristics help their work performance and exchange with clients.

Considering quantitative analysis, the proposed algorithm obtains better results than the algorithm based on association rules in the most of the indicators. In particular, *LDS_RoughSet* algorithm was superior regards indicators: summaries' amount, *T1, T3, T4* and *T5*. While, the algorithm based on association rules was the best considering execution time, because of rough sets theory is high time-consuming. There are not significant differences between algorithms regards *T2* indicator.

References

1. Yager, R.R.: A new approach to the summarization of data. Inf. Sci. **28**(1), 69–86 (1982). https://doi.org/10.1016/0020-0255(82)90033-0
2. Ramos-Soto, A., Martin-Rodillab, P.: Enriching linguistic descriptions of data: a framework for composite protoforms. Fuzzy Sets Syst. **26** (2019). https://doi.org/10.1016/j.fss.2019.11.013
3. Hudec, M., Bednárová, E., Holzinger, A.: Augmenting statistical data dissemination by short quantified sentences of natural language. J. Off. Stat. **34**(4), 981–1010 (2018). https://doi.org/10.2478/jos-2018-0048
4. Kacprzyk, J., Zadrożny, S.: Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. Inf. Sci. **173**(4), 281–304 (2005). https://doi.org/10.1016/j.ins.2005.03.002
5. Kacprzyk, J., Wilbik, A.: Linguistic summaries of time series: on some extended aggregation techniques. Stud. Mater. Polskiego Stowarzyszenia Zarzdzania Wiedza **2010**(31), 326–337 (2010)
6. Wilbik, A., Dijkman, R.M.: On the generation of useful linguistic summaries of sequences. In 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 555–562. IEEE (2016). https://doi.org/10.1109/fuzz-ieee.2016.7737736
7. Kacprzyk, J., Zadrożny, S.: Linguistic database summaries using fuzzy logic, towards a human-consistent data mining tool (20), 10 (2009)
8. Kacprzyk, J., Zadrozny, S.: Linguistic data summarization: a high scalability through the use of natural language? In: Scalable Fuzzy Algorithms for Data Management and Analysis: Methods and Design, pp. 214–237. IGI Global (2010)
9. Pérez Pupo, I., Piñero Pérez, P.Y., García Vacacela, R., Bello, R., Santos Acosta, O., Leyva Vázquez, M.Y.: Extensions to linguistic summaries indicators based on neutrosophic theory, applications in project management decisions. Neutrosophic Sets Syst. **22**, 87–100 (2018)
10. Kacprzyk, J., Zadrożny, S., Dziedzic, M.: A novel view of bipolarity in linguistic data summaries. In: Kóczy, L.T., Pozna, C.R., Kacprzyk, J. (eds.) Issues and Challenges of Intelligent Systems and Computational Intelligence. SCI, vol. 530, pp. 215–229. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-03206-1_16
11. Pérez Pupo, I., García Vacacela, R., Piñero Pérez, P., Sadeq, G., Peña Abreu, M.: Experiencias en el uso de técnicas de soft-computing en la evaluación de proyectos de software. Rev. Invest. Oper. **41**(1), 106–117 (2020)
12. Wilbik, A., Kaymak, U., Dijkman, R. M.: A method for improving the generation of linguistic summaries. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE): pp. 1–6. IEEE (2017). https://doi.org/10.1109/fuzz-ieee.2017.8015752
13. Donis-Díaz, C.A., Bello, R., Kacprzyk, J.: Using ant colony optimization and genetic algorithms for the linguistic summarization of creep data. In: Angelov, P., et al. (eds.) Intelligent Systems'2014. AISC, vol. 322, pp. 81–92. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-11313-5_8
14. Kaczmarek-Majer, K., Hryniewicz, O., Dominiak, M., Święcicki, Ł.: Personalized linguistic summaries in smartphone-based monitoring of bipolar disorder patients. Presented at the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019). Atlantis Press (2019). https://doi.org/10.2991/eusflat-19.2019.56
15. Jain, A., Popescu, M., Keller, J., Rantz, M., Markway, B.: Linguistic summarization of in-home sensor data. J. Biomed. Inf. **96** (2019). https://doi.org/10.1016/j.jbi.2019.103240

16. Smits, G., Nerzic, P., Lesot, M.-J., Pivert, O.: FRELS: fast and reliable estimated linguistic summaries. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6. IEEE (2019). https://doi.org/10.1109/fuzz-ieee.2019.8858836

17. Smits, G., Nerzic, P., Pivert, O., Lesot, M.-J.: Efficient generation of reliable estimated linguistic summaries. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–8 (2018). https://doi.org/10.1109/fuzz-ieee.2018.8491604

18. Pawlak, Z., Skowron, A.: Rough sets: some extensions. Inf. Sci. **177**(1), 28–40 (2007). https://doi.org/10.1016/j.ins.2006.06.006

19. Vanderpooten, D.: Similarity relation as a basis for rough approximations. Adv. Mach. Intell. Soft Comput. **4**, 17–33 (1997)

20. Bello, R.: Uncertainty Management with Fuzzy and Rough Sets: Recent Advances and Applications, vol. 377. Springer, Heidelberg (2019). http://www.springer.com/series/2941

21. Wang, G.: Extension of rough sets under incomplete information systems. In: 2002 IEEE World Congress on Computational Intelligence. In: 2002 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2002. Proceedings (Cat. No.02CH37291), vol. 2, pp. 1098–1103 (2002). https://doi.org/10.1109/fuzz.2002.1006657

22. Zhai, J., Zhang, S., Zhang, Y.: An extension of rough fuzzy set. J. Intell. Fuzzy Syst. **30**(6), 3311–3320 (2016). https://doi.org/10.3233/IFS-152079

23. Pawlak, Z.: Rough sets. Int. J. Comput. Inform. Sci. **11**(5), 341–356 (1982). https://doi.org/10.1007/BF01001956

24. Bello, R., Verdegay, J.L.: Rough sets in the soft computing environment. Inf. Sci. **212**, 1–14 (2012). https://doi.org/10.1016/j.ins.2012.04.041

25. Wasilewska, A.: Apriori algorithm. Lecture Notes (2007). https://www3.cs.stonybrook.edu/cse634/lecture_notes/07apriori.pdf

26. PMI: A guide to the project management body of knowledge (PMBOK guide) Sixth Edition/Project Management Institute. Project Management Institute, Inc., Newtown Square, Pennsylvania 19073-3299, USA (2017)

27. Varajão, J., Colomo-Palacios, R., Silva, H.: ISO 21500: 2012 and PMBoK 5 processes in information systems project management. Comput. Stand. Interfaces **50**, 216–222 (2017)

28. López, P.: Procedimiento para la aplicacion de test de personalidad como apoyo a la gestion de recursos humanos en proyectos informáticos (Maestría en Gestión de Proyectos). Universidad de las Ciencias Informáticas, La Habana, Cuba (2017)

29. Furnham, A.: Personality and occupational success: 16PF correlates of cabin crew performance. Pers. Individ. Differ. **12**(1), 87–90 (1991)

30. Caprara, G.V., Barbaranelli, C., Borgogni, L., Perugini, M.: The "big five questionnaire": a new questionnaire to assess the five factor model. Pers. Individ. Differ. **15**(3), 281–288 (1993)

31. Pérez, P.P., Pupo, I.P., Rivero Hechavarría, C.C., Lusardo, C.R., Sosa, R.G., López, S.T.: Repositorio de datos para investigaciones en gestión de proyectos. Rev. Cubana Ciencias Inf. **13**(1), 176–191 (2019). https://gespro.uci.cu/projects/