



Evolving Fuzzy Membership Functions for Soft Skills Assessment Optimization

Antonia Azzini, Stefania Marrara^(✉), and Amir Topalović

Consortium for the Technology Transfer (C2T),
via Nuova Valassina, Carate Brianza, MB, Italy
{antonia.azzini,stefania.marrara,amir.topalovic}@consorzioctt.it

Abstract. This work proposes the design of a decision support tool able to guide the choices of any company HR manager in the evaluation of the profiles of PhD candidates. This paper is part of an ongoing research in the field of PhD profiling. The novelty here is an evolutionary fuzzy model, based on the Membership Functions (MFs) optimization, used to obtain the soft skills candidate profiles. The general aim of the project is the definition of a set of fuzzy rules that are very similar to those that a HR expert would otherwise have to calculate each time for each selected profile and for each individual skill.

Keywords: Fuzzy logic · Evolutionary algorithms · Membership function optimization

1 Introduction

Since many years, those who have been facing a PhD course in Europe have been aware that a good percentage of them will not find a permanent position in the Academy, but will have to migrate to companies and public/private organizations that are not always ready to understand and enhance the research experience [6].

One of the problems at the basis of this misunderstanding between PhDs and companies is encountered during the recruitment phase. Most job matching portals are based on keyword searches in a candidate's CV, but the taxonomy used in job offers (also called *job vacancies*) is set on employers' vocabulary and usually does not match the words that a PhD would use to describe her experience. Consequently, it is widely recognized that there is a need to define a system that can support an Human Resources (HR) team in recruiting doctoral candidates.

In the analysis of the profiles of possible candidates for a job vacancy, the identification of skills is a very important but costly activity in terms of time for the HR team. To overcome this problem, some works in the literature adopt machine learning and fuzzy based approaches to manage, for example, employability [10], which, together with skills, takes into account personal attributes for the development of teaching strategies.

This work is part of the project “*SOON - Skills Out of Narrative*”, which aims at designing a decision support tool able to guide the choices of any company HR manager in the evaluation of the profiles of PhD candidates.

The novelty of this work, compared to previous works in the same project [3–5], is focused on obtaining the soft skills candidate profiles by using an evolutionary fuzzy approach.

Since fuzzy models highly depend on the sets of terms underlying each variable and the fuzzy inference method adopted, although easy to design, they are often difficult to be tuned.

In this direction the evolutionary algorithm employed in this work aims at optimizing the Fuzzy Inference System (FIS) model by creating the population of Membership Functions (MFs). The interval values representing the MFs are encoded in the chromosome of the evolutionary algorithm as real numbers. The chromosome of each individual represents the diffusion of the MFs generated.

This approach is in fact able to compute a set of fuzzy rules that are very similar to those that a HR expert would otherwise have to calculate each time for each selected profile and for each individual skill.

The core of this work is the definition of an evolved model, whose behavior will be completely customized because computed on a dataset of pairs (profile, set of soft skills) that represents the decision-making behavior of a certain HR Manager rather than another.

The remaining of the paper is organized as follow. After a brief summary of the related works in Sect. 2, while the taxonomy is defined into Sect. 3. The soft skills profiles are defined into Sect. 4. Then, the evolutionary encoding of the MFs, together with the overall architecture of the approach and the details of the evolutionary algorithm, are reported into Sect. 5. Preliminary experiments are presented and discussed, together with the obtained results, in Sect. 6, while final remarks are reported in Sect. 7.

2 Related Work

The automatic extraction of meaningful information from unstructured texts has been mainly devoted to support the e-recruitment process [12], e.g., to help human resource departments to identify the most suitable candidate for an open position from a set of applicants or to help a job seeker in identifying the most suitable open positions. For example, the work described in [16] proposes a system which aims to analyze candidate profiles for jobs, by extracting information from unstructured resumes through the use of probabilistic information extraction techniques as Conditional Random Fields [11].

Differently, in [18] the authors define Structured Relevance Models (SRM), a context based extension of relevance-based language models for modeling and retrieving semi-structured documents, and describe their use to identify job descriptions and resumes vocabulary, while in [4], a methodology based on machine learning aimed at extracting the soft skills of a PhD from a textual, self-written, description of her competencies is described. In that paper a neuro-fuzzy controller defining a set of inference rules, similar to those setted by a HR

expert, is proposed. Anyway, in that case the most critical aspect refers to the definition of the fuzzy MFs.

In fact, MFs are a crucial part of the definition as they define the mappings to assign meaning to input data. The MFs map crisp input observations of linguistic variables to degrees of membership in fuzzy sets to describe properties of the linguistic variables. Kroeske, for example, explains in [9] how suitable MFs are designed depending on the specific characteristics of the linguistic variables as well as peculiar properties related to their use in optimization systems.

Other works have been presented in this direction. Some of them consider the evolution of fuzzy rules, while some others are focused on the MFs deployment. A last set of approaches consider the optimization of both memberships and rules: in [8] the authors present a genetic algorithm to optimize the MFs used in determining fuzzy association rules. On the other hand, the genetic tuning algorithm implemented in [14] aims at optimizing both MFs and rules for the optimization of a fuzzy temperature controller. Herrera [7] proposes a review of classical models compared with the most recent trends for Genetic Fuzzy Systems, while an approach based on the fuzzy rules evolution has been more recently explained by Mankad and colleagues [13].

Other examples of evolutionary algorithms, as Differential Evolution and Evolutionary Strategies, are instead more natural for continuous optimization. In the work carried out in [17], two kinds of computational intelligence techniques were used to create the framework: a fuzzy logic system (FS) as the decision maker and evolutionary computation as the model parameter optimizer. In particular, FS membership function parameters have been optimized by using a differential evolution (DE) algorithm to find optimal model parameters. Also Adhikari and colleagues presented in their work [1] a fuzzy adaptive differential evolution (DE) for path planning. The path-planning problem was formulated as a multi-objective unconstrained optimization problem, with the aim of minimizing the costs as well as finding the shortest path. DE has been used for optimization with a fuzzy logic controller used to find the parameter values of DE during the optimization process. Evolutionary Strategies are then proposed by Santika and colleagues in [15] was based on an Evolution Strategy method to determine the appropriate rules for Sugeno FIS having the minimum forecasting error. Near to the work carried out into this project, the Mean Square Error (MSE) has been used to evaluate the goodness of the result (in their case a forecasting). The numerical experiments showed the effectiveness of the proposed optimized Sugeno FIS for several test-case problems, as well as the capability to produce the lower MSE comparable to those achieved by other well-known methods in the literature.

3 The Soft Skills Taxonomy

In the last years, recognizing, evaluating and in case enhancing soft skills in employees has been a hot topic in literature. *Soft or Transferable skills* enhance people future employability, adaptability and occupational mobility. Since FYD's mission is to promote researchers employment in companies, the core of the

project is identifying the skills that can be transferred from the Academy experience. There is a lack of consistent theory for defining and classifying various skills, and there is no generally accepted transferable skills taxonomy. The European project team decided to distinguish three categories of skills: (1) soft skills; (2) generic hard skills; (3) specific hard skills. Specific hard skills are characterized by their lower level of transferability, whereas soft skills and generic hard skills are skills with high transferability across sectors and occupations and can be identified as transversal skills. Focusing on researchers our attention was centered on capturing the soft skills that support the innovation activity. Having these skills, which can be transferred from one context to another, is a good basis for accumulation of specific skills required by a given job expected in managing a robust innovation pipeline and portfolio to deliver new growth opportunities. Therefore, our approach classifies the researcher soft skills into 6 categories: *carefulness*, i.e., the candidate is careful to look at or consider every part of something to make certain it is correct or safe; *creativity*, i.e., the ability to produce original and unusual ideas, or to make something new or imaginative; *unexpected/emergency* i.e., the ability to deal in an effective way with something that happens suddenly or expectantly and needs fast action in order to avoid harmful results; *uncertainty*, i.e., the ability to deal with a situation in which something is not known, or with something that is not known or certain; *communication*, i.e., the ability to communicate with people; and *networking*, i.e., the process of meeting and talking to a lot of people, esp. in order to get information that can help you.

Each skill category is divided into several classes, each class representing a particular soft skill. A detailed description of the taxonomy is available in [4].

4 PhDs Profiles Definition

As described in [1,2], in the “*SOON - Skills Out of Narrative*” project each PhD is required to provide two textual descriptions of her experience, the *curriculum vitae (CV)* and a questionnaire composed by 13 free text questions called *experience pills*, or simply *pills*, in which she is loosely guided to describe her soft skills. The approach extracts the skills from the text provided by cv and pills and creates a formal representation of the researcher, the profile.

In literature several techniques that can be applied to create the PhD profile are presented, at present the vector based representation developed by the Information Retrieval researchers for documents representation is adopted. In the vector based model a document D is represented as an m -dimensional vector, where each dimension corresponds to a distinct term and m is the total number of terms used in the collection of documents.

From this basis, in this approach a profile RP is composed by two vectors, $RP = (H, S)$ where H is the vector representing the hard skills of the PhD, while S represents her soft skills. The hard skills vector H is not focus of this paper (see [4]). The soft skills vector S is written as (x_1, \dots, x_n) , where x_j is the weight of skill s_j and n is the number of skills defined in the soft skills proprietary taxonomy described in [4]. If the profile does not contain a skill then the corresponding weight is zero.

PhD Soft Skills Representation. Every item of the vector S is a linguistic 2-tuple value [7] representing the degree the PhD possesses that soft skill. Note that a positional notation is used: $S = (s_1, s_2, \dots, s_k)$, where $s_j \in S$, with $j = \{1, \dots, 60\}$, describes the linguistic degree assigned to the j -th skill of the PhD.

The vector of soft skills S is computed by taking into account two contributions. The first contribution to S is a vector HR of 60 skills, which represents the assessment the HR operator performs during an interview with the candidate. To allow a flexible assessment, but avoiding at the same time an excessive overhead for the HR operator, this vector adopts a representation with 5 labels (L^5) plus the NC value (NC =not classified) to describe each skill. Note that during an interview the HR operators explicitly assess only a few skills (usually 6 or 7), all other skills are set to NC by default.

$$- L^5 = \{l_0 = \text{VeryLow} = VL, l_1 = \text{Low} = L, l_2 = \text{Medium} = M, l_3 = \text{High} = H, l_4 = \text{Full} = F\}$$

The second contribution to S is the vector ML of 60 skills that represents the automatic assessment of the candidate performed by the machine learning based classifier, presented in [3,4]. After a preprocessing phase in which the raw text is divided into sentences, each sentence is analyzed to extract the skills. The ML classifier analyses the textual self description each PhD is required to provide when she enrolls to the project. In order to allow a high flexibility for the ML vector we adopt a representation with 11 labels (L^{11}) to assess each skill (s_j).

$$- L^{11} = \{L_0 = \text{Null} = N, l_1 = \text{VeryVeryLow} = VVL, l_2 = \text{VeryLow} = VL, l_3 = \text{Low} = L, l_4 = \text{AlmostMedium} = AM, l_5 = \text{Medium} = M, l_6 = \text{MoreThanMedium} = MM, l_7 = \text{AlmostHigh} = AH, l_8 = \text{High} = H, l_9 = \text{VeryHigh} = VH, l_{10} = \text{Full} = F\}$$

The final vector S employs 11 labels as the vector ML .

The result obtained from the vector product of the labels of the two HR and ML vectors (added to the null case rule, i.e. $(5 * 11) + 1$) represents the total number of rules that are generated and then evaluated by a FIS model.

5 The Approach

The architecture of the tool able to extract the soft skills profiles for a PhD is shown in Fig. 1.

The PhD registering to the project is asked to provide a textual description of her transverse competencies: the questionnaire called “pills”.

The *Text Preprocessing & Skills ML Extraction* module is in charge to extract a soft skills vector (in Figure is the *Soft-Skills ML Vector*) from the textual content of the “pills”. In this approach the HR Operator, during the interview, compiles a report regarding her evaluation of the soft skills of the candidate via a simple interface (in Figure *HR Interface*) that guides the compilation of the *Soft-Skills HR Vector*. Please note that even if a complete evaluation of the candidate would require an assessment of each of the 60 soft skills available in the adopted

taxonomy, this is not necessary here, and the operator is required to give an explicit assessment only to the few skills (usually 6/7) he really saw during the interview. The others skills in the *Soft-Skills HR Vector* are automatically set to NC (*not classified*).

The two soft skills vectors are used by the *Evolutionary Module* to compute the final *Soft Skills RP Vector* that composes the Researcher Profile (RP). Section 5.2 detail how the *Evolutionary Module* works.

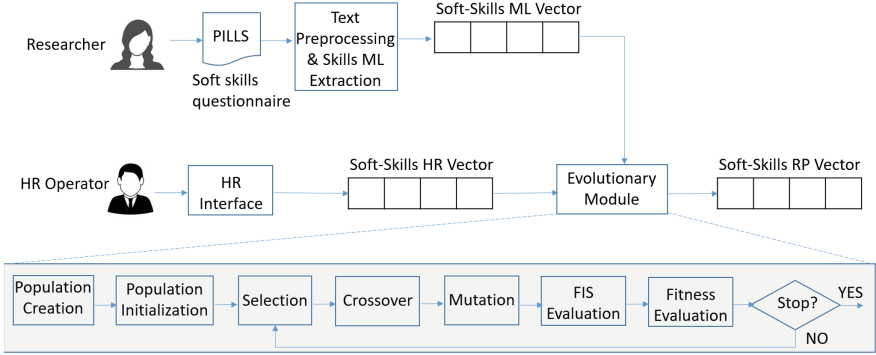


Fig. 1. Architecture of the HR supporting tool.

5.1 Evolutionary Encoding of Fuzzy Membership Functions

In the evolutionary module an individual represents the overall space coverage as its chromosome encodes the MFs associated to the different linguistic degrees in the fuzzy partition considered by the FIS module. In particular each individual chromosome encodes the MFs for each of the two input vectors that define a candidate soft skills profile, i.e., the *ML* and the *HR* vector.

Each MF is represented in the chromosome by the triple (a, c, b) , where (a, b) represent the edges of the triangle base, while c represents triangle center. Note that almost all the MFs are represented by non-symmetric triangles. An example of an individual encoding is shown in Fig. 2. All the MFs are randomly initialized in $[-0.5, 0.5)$.

5.2 Evolutionary Algorithm

The evolutionary phase of the algorithm starts with a population of N individuals, with randomly generated chromosomes, and, on the basis of a first fitness evaluation, the individuals are evolved by the evolutionary cycle till the stop condition is reached, as shown in the grey selection of Fig. 1.

In each generation the worst individuals of the population are replaced with the new ones generated after crossover and mutation. The obtained results are used to calculate the performances (i.e. the fitness) of the individuals. The evolutionary process is stopped accordingly to one of these conditions: (1) no new

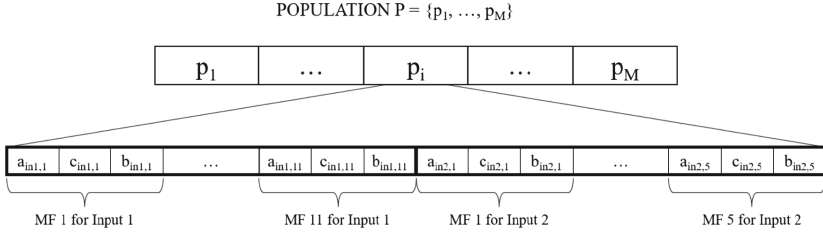


Fig. 2. Individual chromosome representation.

best value was given in the past n iteration; where n is set to 25, or (2) the maximum number of generations is reached.

Selection: the Tournament Selection is implemented by choosing the best individual among a group of x individuals, randomly selected, from the overall population. In this work x is set to 15. The selection is repeated in order to obtain the 10 best individuals that will form the couples of parents used by the crossover operator.

Crossover: a Two-Point Crossover is implemented in this work, with $pcross$ probability. It's applied over the five couples of individuals obtained from the selection operator. The two-point crossover is carried out by considering two different cutting points, one for each of the inputs, respectively the ML and the HR vectors, applied over each chromosome of the two parents.

Mutation: is applied, with a probability $pmut$, to each MF of the offspring generated from the crossover operator. The elements involved in the mutation are, hence, the vectors $\langle a, b, c \rangle$ that define each MF. The mutation is applied with the same probability on each element of the vector and on their respective combinations with a mutation rate $pmut_{rate}$.

After each mutation all MFs are checked and those that are included into larger ones and, with centers c_1 and c_2 shifted by a small distance δ , are deleted, since their contribution becomes negligible with respect to the overall individual behavior (Delete Mutation).

Fitness Evaluation: the fitness of each individual, $fitness = \frac{1}{MSE}$, refers to a maximization problem, and it is calculated over the result produced by the FIS Evaluation module according to the Mean Square Error (MSE). MSE measures the similarity between the estimated model (the soft skill S vector) and the expected one (created by hand by the HR team and modeled as a Mamdani fuzzy system) as explained by the Eq. 1.

$$MSE = \frac{1}{N} \sum_{j=1}^N (S_j - O_j)^2 \tag{1}$$

N corresponds to the population dimension, S_j is the output calculated by the evolving fuzzy approach, while O_j is the output defined by the fuzzy system modeled by hand.

6 Preliminary Experiments and Discussion

This approach represents the first step of an ongoing research project that aimed at defining an automatic approach for the representation of the human expert knowledge in the recruiting process. A first set of experiments has been carried out in order to test the defined architecture.

A benchmark dataset was prepared by analyzing about 600 questionnaires gathered by the HR team cooperating with the project, by dividing each questionnaire answer in sentences, and by labeling each sentence with the most appropriate soft skill. This analysis step gathered around 15,000 labeled sentences. Unfortunately not all the soft skills categories detailed in the taxonomy were populated enough to perform a classifier training task. The most populated category was “communication” with around 5000 sentences randomly partitioned in the 9 soft skills belonging to this area, and stating the classes of the ML classifiers.

For this reason a preliminary evaluation was performed only on this area, with the aim to assess the benefits of using an evolutionary approach applied to a fuzzy system in this context. The benchmark creation activity is currently an ongoing work, with the aim to complete the labeling of all questionnaires available now in the full collection.

The evolutionary process runs over a population of 50 individuals randomly created, for 200 generations, while the dataset was used to evolve. After the selection of the best 5 couples individuals, chosen by the tournament, the crossover is applied to the selected parents with a *pcross* probability set to 0.5. The mutation of the generated offspring is applied with a *pmut* probability set to 0.2., with a *pmut_{rate}* set to 0.14. The fitness curve of the best individual resulting from the application of the fuzzy evolutionary approach over the test set is reported into Fig. 3. It is possible to observe how the function, which is a maximization, results to be increasing over the execution time, until the maximum number of generations is reached.

The output has been validated by comparing the results obtained by the evolving fuzzy approach with a Mamdani fuzzy system manually defined by the HR manager as shown in Table 1. The input values correspond respectively, to the encoding of the labels reported into the ML and HR input vectors, while the output produced corresponds to the ownership intensity of the analyzed soft skills.

The obtained results suggest that the Evolving Fuzzy approach follows the behaviour of the HR expert whenever she is not available, with an accuracy of the best individual equal to 93.9% and an averaged *MSE* equal to 6.1%, thus representing a good automation of her reasoning when assessing profiles.

Moreover, the evaluations carried out on the skills by the evolving fuzzy approach show a cautious and prudent attitude, especially when the assessment made by the human expert is really different from that made by the evolutionary approach.

Moreover, the experiments carried out in this work highlight that, as shown by comparing Figs. 4 and 5, the elimination of all the MFs with a negligible behavior, allows to reduce the overall number of rules that have to be evaluated

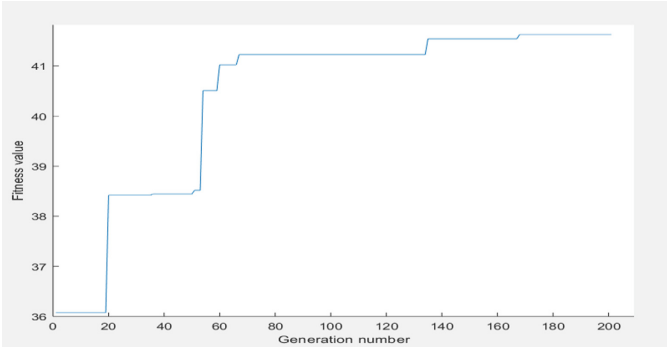


Fig. 3. Fitness function of the best individual calculated over the test set.

Table 1. Comparative results between the mamdani and the evolutionary fuzzy approach.

Output	Inputs (ML, HR)								
	(-0.3, -0.3)	(-0.3, 0)	(-0.3, 0.3)	(0, -0.3)	(0, 0)	(0, 0.3)	(0.3, -0.3)	(0.3, 0)	(0.3, 0.3)
Mamdani	-0.2241	-0.2000	0.0009	-0.1000	0.0009	0.1000	0.0545	0.1545	0.3241
Evo-fuzzy	-0.2040	-0.2000	-0.0907	-0.0356	-0.0339	0.0410	0.1679	0.1419	0.2731

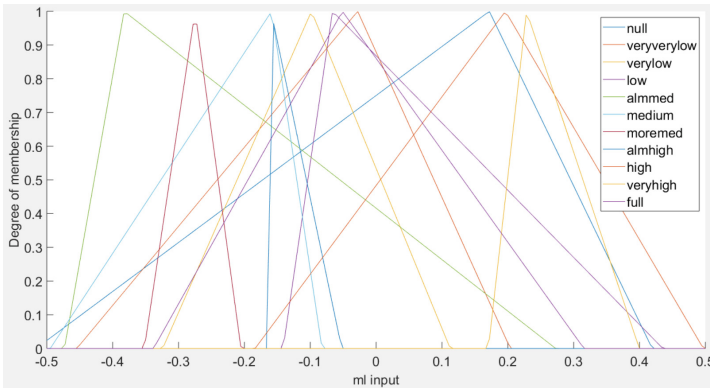


Fig. 4. Triangular evolved membership functions distribution without delete mutation.

each time by the fuzzy system, by also reducing the computational costs of the entire approach. In this case the total number of triangles was reduced from 11 to 8, and the total number of FIS rules from 56 to 41. The evolutionary approach has also allowed the triangular MFs not to saturate at the extremes of the crisp values, but to obtain discrete values.

Preliminary evaluations show that MFs mostly disappear in the set of ML inputs, by simplifying the corresponding individual encoding. In this sense the evolutionary algorithm is able to optimize the MFs by approximating the behavior of a human expert.

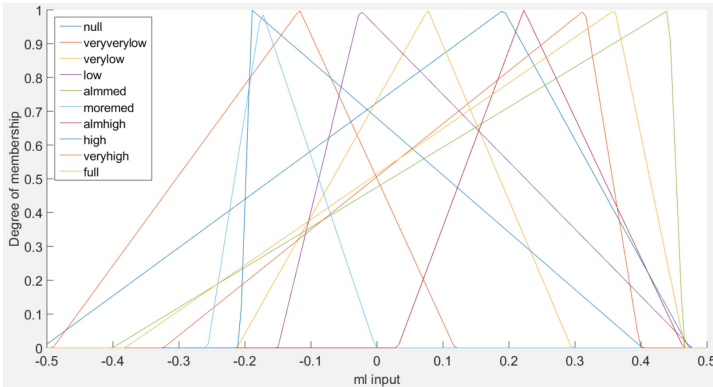


Fig. 5. Triangular evolved membership functions distribution with delete mutation.

With respect to our previous work [5], in which it was necessary to have knowledge of the output of the Neuro-Fuzzy model in order to build the FIS model, the evolutionary fuzzy approach here presented requires only an expected profile of the PhD (manually created by the HR team), while the FIS model is built by means of an evolutionary algorithm. Moreover, a smaller number of MFs reduces the computational complexity of the FIS model.

7 Conclusion

In this paper an evolutionary fuzzy approach is used to compute the vector of soft skills that composes a profile. The vector is the output of two contributes: one obtained by human assessment, and one obtained by extracting the soft skills from a textual self description. The evolutionary tuning algorithm employed in this works aims at optimizing the fuzzy membership functions. Preliminary results show that the evolutionary fuzzy approach well approximates the reasoning of a human expert when assessing profiles. Moreover, differently from the previous approach [4], the evolutionary approach allows to automatically define the FIS system without any previous knowledge of the desired output. Future work will investigate the application of the proposed approach in other contexts of HR recruitment, besides PhD, thus collecting more datasets to better test the hypothesis of creating a good “HR manager bot”.

References

1. Adhikari, D., Kim, E., Reza, H.: A fuzzy adaptive differential evolution for multi-objective 3D uav path optimization. In: Proceedings of the IEEE Conference on Evolutionary Computation (CEC), June 2017
2. Azzini, A., Galimberti, A., Marrara, S., Ratti, E.: A taxonomy of researchers soft skills (2017)

3. Azzini, A., Galimberti, A., Marrara, S., Ratti, E.: A classifier to identify soft skills in a researcher textual description. In: Sim, K., Kaufmann, P. (eds.) *EvoApplications 2018*. LNCS, vol. 10784, pp. 538–546. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77538-8_37
4. Azzini, A., Galimberti, A., Marrara, S., Ratti, E.: SOON: supporting the evaluation of researchers' profiles. In: Uden, L., Hadzima, B., Ting, I.-H. (eds.) *KMO 2018*. CCIS, vol. 877, pp. 3–14. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95204-8_1
5. Azzini, A., Marrara, S., Topalovic, A.: A neuro-fuzzy approach to assess the soft skills profile of a PhD. In: Fullér, R., Giove, S., Masulli, F. (eds.) *WILF 2018*. LNCS (LNAI), vol. 11291, pp. 134–147. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12544-8_11
6. Box, S.: *Transferable skills training for researchers supporting career development and research: supporting career development and research*. OECD Publishing (2012)
7. Herrera, F.: Genetic fuzzy systems: status, critical considerations and future directions. *Int. J. Comput. Intell. Res.* **1**(1), 59–67 (2005)
8. Kaya, M., Alhaji, R.: Utilizing genetic algorithms to optimize membership functions for fuzzy weighted association rules mining. *Appl. Intell.* **24**(1), 7–15 (2006)
9. Kroeske, J., Ghandar, A., Michalewicz, Z., Neumann, F.: Learning fuzzy rules with evolutionary algorithms—an analytic approach. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) *PPSN 2008*. LNCS, vol. 5199, pp. 1051–1060. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87700-4_104
10. Kumar, S., Kumar, R., Sharma, V.: Adaptive neural fuzzy inference system for employability assessment. *Int. J. Comput. Appl. Technol. Res* **3**, 159–164 (2014)
11. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning, ICML*, vol. 1, pp. 282–289 (2001)
12. Lee, I.: Modeling the benefit of e-recruiting process integration. *Decis. Support Syst.* **51**(1), 230–239 (2011)
13. Mankad, K., Sajja, P., Akerkar, R.: Evolving rules using genetic fuzzy approach - an educational case study. *Int. J. Soft Comput. IJSC* **2**(1), 35–44 (2011)
14. Mudholkar, R., Somnatti, V.: Genetic tuning of fuzzy inference system for furnace temperature controller. *Int. J. Comput. Sci. Inf. Technol.* **6**(4), 3496–3500 (2015)
15. Santika, G., Mahmudy, W., Naba, A.: Rule optimization of fuzzy inference system sugeno using evolution strategy for electricity consumption forecasting. *Int. J. Electr. Comput. Eng.* **7**, 2241–2252 (2017)
16. Singh, A., Rose, C., Visweswariah, K., Chenthamarakshan, V., Kambhatla, N.: PROSPECT: a system for screening candidates for recruitment. In: *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, pp. 659–668. ACM (2010)
17. Vorachart, V., Takagi, H.: Evolving fuzzy logic rule-based game player model for game development. *Int. J. Innov. Comput. Inf. Control* **13**(6), 1941–1951 (2017)
18. Yi, X., Allan, J., Croft, W.B.: Matching resumes and jobs based on relevance models. In: *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 809–810. ACM (2007)