PERGAMON

Expert Systems with Applications 26 (2004) 217–224

Expert Systems with Applications

www.elsevier.com/locate/eswa

An expert system for job matching of the unemployed

A. Drigas*, S. Kouremenos, S. Vrettos, J. Vrettaros, D. Kouremenos

Department of Technological Applications, National Center for Scientific Research 'DEMOKRITOS', P.O. Box 15310 Gr. Ag. Paraskevi Attikis, Athens, Greece

Abstract

This paper presents an expert system for evaluation of the unemployed at certain offered posts. The expert application uses Neuro-Fuzzy techniques for analyzing a corporate database of unemployed and enterprises profile data. The process of matching an unemployed with an offered job is performed through a Sugeno type Neuro-Fuzzy inference system. Large training sets of old historical records of unemployed (belonging to the same social class), rejected or approved at several posts, (provided by the Greek General Secretariat of Social Training) were used to define the weights of the system parameters. New instances of rejected or approved cases arriving become part of the training set. Retraining is performed after a standard amount of new cases available. The system output is a measure of the unemployed suitability for the certain job (evaluation mark).

Keywords: Expert systems; Neuro-fuzzy; Job matching

1. Introduction

Integrating neural networks and fuzzy logic creates powerful expert decision systems. In recent years, the research area of hybrid and neural processing has seen a remarkably active development. Furthermore, there has been an enormous increase in the successful use of hybrid intelligent systems in many diverse areas such as speech/natural language understanding, robotics, medical diagnosis, fault diagnosis of industrial equipment, education (Vrettaros, 1996), e-commerce (Kouremenos, Vrettos & Stafylopatis, 2003), recommendation and information retrieval (Vrettos & Stafylopatis, 2001). However, expert job matching has not been previously examined in literature. The process of matching an unemployed with a certain job too often needs to be done in a more structured and thorough manner than the simple Boolean matching method used extensively on the web (A better job fast; Europe's career market on the internet, 2001; Federal job, 2001; Jobsdb, 1998–2001). The selection of qualified individuals for certain posts is a difficult task either in larger or smaller companies and requires expert decision systems. Skills Analyzer Tool (Labate & Medsker, 1993) was designed for solving management problems concerning the employee's classification into several projects. It combines neural networks and rule-based analysis to match the employees of a company with certain jobs of new projects. The above system is expert, though its hybrid techniques are old.

Collaboration filtering techniques were used in CASPER project (Rafter, Bradley & Smyth, 2000) to enforce with intelligence the search engine of JobFinder website (www.jobfinder.com).

The CASPER system is consisted of the following subsystems: a user profiling system that creates the user profile according to his/her behavior within the JobFinder site, an automated collaborative filtering engine for recommendation services and a personalised retrieval engine. Mobile Agents technology has been applied to the EMA employment agent (Gams, Golob, Karaliø, Drobniø, Grobelnik, Glazer, Pirher, Furlan, Vrenko & Križman, 1998). EMA is a typical recommender agent providing information upon demand or when it notices a relevant job for a particular user. Methods used by CASPER and EMA have been extensively used in recommendation and information retrieval and could be successfully used in job information retrieval. However, those systems are not expert and make no use of realistic cases of job matching. Retraining according to the final rejection or approval of the user by the enterprise is a critical point and

* Corresponding author. Tel.: +30-210-650-3124; fax: +30-210-650-3129.

E-mail address: dr@imm.demokritos.gr (A. Drigas).

0957-4174/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0957-4174(03)00136-2

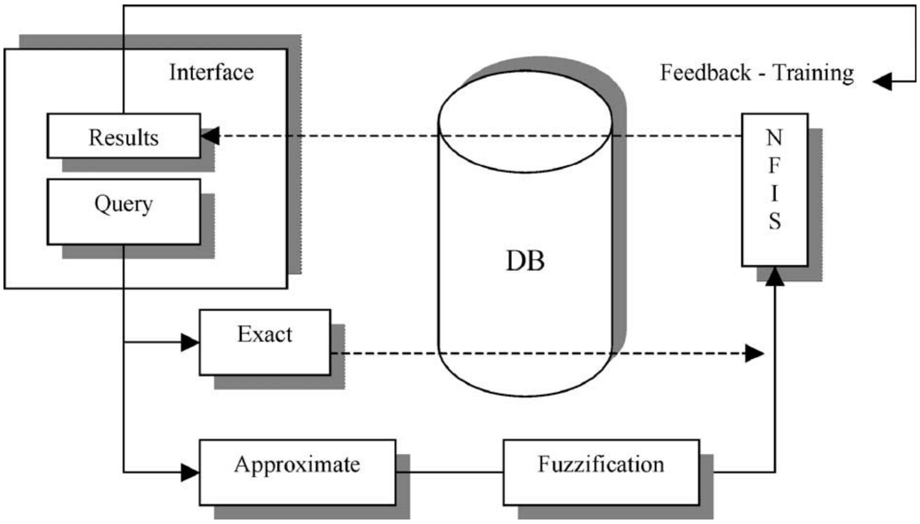


Fig. 1. The simplified architecture of the system. The exact part of the query limits the amount of records that will be evaluated by the Neuro-Fuzzy inference system. Approximate part for each record is fuzzified and the NFIS assigns a suitability measure.

should not be neglected. Not in few cases recommendation percentage was medium or even high while the enterprise employers rejected the unemployed after a typical interview. Intelligent recommender and retrieval systems do not support training and should be used for indicative retrieval of job information and not for real job matching.

The structure of the rest of our paper is the following. Section 2 sets analytically our system architecture and structure, Section 3 is a typical use case example—use and results of a certain real case of job matching. Conclusions are presented in Section 4. In Appendix A we present statistical data for a macroscopic view of the unemployed profile.

2. Description of the system

Our system has the following functional characteristics:

1. Connectivity with the corporate database that contains the unemployed, employers and offered jobs records.
2. Use of Neuro-Fuzzy techniques for the inductive (through examples) training of complex fuzzy terms. These terms contribute to the evaluation of the data and the final decision phase.
3. Supervised retraining of the Neuro-Fuzzy network when recommended by the administrator.
4. Fuzzy models that design and develop the fuzzy inference engine.
5. Combination–processing of the fuzzy elements for the final data evaluation.
6. Flexible and friendly user interface (Visual Basic input forms).

Our system is consisted of certain components (Fig. 1) that need to be integrated into the expert system for the extraction of the final output (evaluation mark of the unemployed).

The query/job opportunity is formulated using the following six fields (Fig. 2):

1. Age
2. Education
3. Additional Education (Training)
4. Previous Employment (Experience)
5. Foreign Language (English)
6. Computer Knowledge

Generally, the query is consisted of two parts: the exact and the approximate. The exact part corresponds to binary criteria that the candidate *must* fulfill. In case of multiple binary criteria, the database is filtered so that all Boolean expressions are satisfied. The approximate part of the query is fuzzified and fed as an input to a Sugeno type FIS (Jang, Sun & Mizutani, 1997; Sugeno & Kang, 1988).

The output of the FIS is the evaluation mark of the unemployed concerning a certain employment (how suitable is the unemployed for the certain job). This problem is a functional approximation problem with continuous inputs. That means, it is a problem of supervised learning. We used previous unemployed evaluation experience to train the inner parameters of the FIS.

2.1. Rule base of the system

There is a fuzzy rule for every criterion that is to be met, of the form:

Candidate's X matches X Criterion

where X can be one of the fuzzy sets (Age, Education, Training, Experience, Language, Computer Knowledge). The set of all these rules constitute the rule base of the system (Table 1).

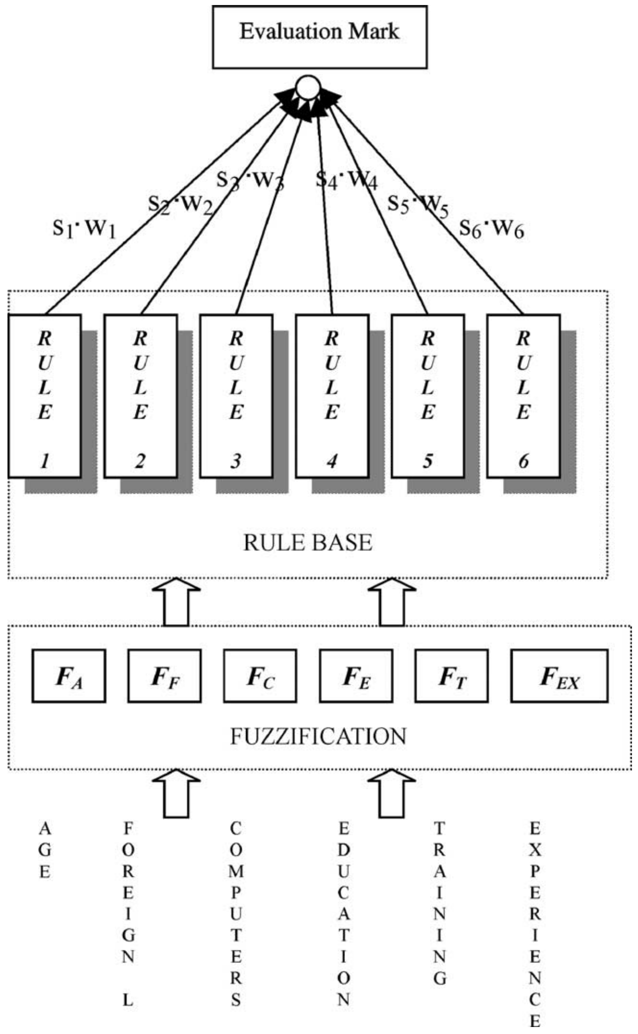


Fig. 2. The Sugeno type Neuro-Fuzzy inference system.

2.2. Fuzzification and rule evaluation

In the following paragraphs, we will describe the exact fuzzification methods of the fields that formulate the query.

We make extensive use of the notion of Fuzzy Relations (Jang et al., 1997). Binary fuzzy relations are fuzzy sets in X*x*Y which map each element in X*x*Y to a membership grade between 0 and 1. More formally: Let X and Y be two universes of discourse. Then

$$\propto_R = \{((x,y), \propto_R(x,y))|(x,y)\} \quad (1)$$

The above equation is a binary fuzzy relation in X*x*Y. When X, Y are discrete we can describe a fuzzy relation using a relation matrix of the form

$$R = \begin{matrix} \mu(1,1) & \dots & \mu(1,n) \\ \dots & \dots & \dots \\ \mu(m,1) & \dots & \mu(m,n) \end{matrix} \quad (2)$$

where $\mu(i,j)$ is the membership value for i,j.

For the fuzzification of education, additional education and previous employment fields, we developed a fuzzy relation between different specialities. This fuzzy relation indicates the relevance between different specialities. The membership values were determined after several interviews from enterprise employers about the relation between different specialities. The final value for each speciality is the average value of the interviews results. We created a relation matrix called speciality membership (Table 2) of the different specialities belonging to the same category (e.g. mechanics, engineers). We used special codes to define the categories and subcategories of the employments. In Table 2, code '45' is the general category of the employment 'Manufacture' and '4513', '4514', '4590' are the specialities of the certain employment.

2.2.1. Age

The age match membership function is based on the acknowledgment that the employers can determine age limits for each applicant (20–30, 25–25, etc.). For this, we defined a fuzzy membership function for the ages that satisfy the predefined limits which escalates smoothly between these limits (Fig. 3). In this way, capable applicants

Table 1
The rule base of the system

Candidate's age matches age criterion
Candidate's education matches education criterion
Candidate's training matches training criterion
Candidate's experience matches experience criterion
Candidate's F. language matches F. language criterion
Candidate's C. knowledge matches C. knowledge criterion

Table 2
The specialities relation matrix (speciality membership)

	4513	4514	4515	4531	4532	4533	4590
4513	1	0.9	0.5	0.2	0.1	0	0.3
4514	0.9	1	0.5	0.2	0.1	0	0.3
4515	0.5	0.5	1	0	0.25	0.2	0.4
4531	0.2	0.2	0	1	0.9	1	0
4532	0.1	0.1	0.25	0.9	1	0.4	0.1
4533	0	0	0.2	1	0.4	1	0.2
4590	0.3	0.3	0.4	0	0.1	0.2	1

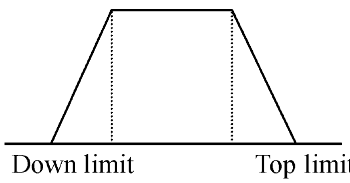


Fig. 3. The age parameter membership function.

who diverge slightly from the age limits are not rejected. So, the strength s_1 of the rule

Candidate's Age matches Age Criterion

is given by the value of Age match (membership value).

2.2.2. Education

The user education is recorded into three fields: basic education, degree and post-graduate. These qualifications are recorded as work codes with a general approved coding method. The rule:

Candidate's Education matches Education Criterion

is realized as a fuzzy relation between the fields Candidate's Education and Education Request (Table 3).

The rule strength s_2 is the product of Education Match with Profession Membership.

2.2.3. Additional education (training)

In this field, we record the possible additional user education at certain employments. This field can take only two values {Yes: 1, No: 0}. The strength s_3 of the rule:

Candidate's Training matches Training Criterion

is given as the product of Training Match {0,1} with Profession Membership.

2.2.4. Previous employment (experience)

In the same way, the previous experience of the user is recorded up to three records with the same coding method. This field can take only two values {Yes: 1, No: 0}. The strength s_4 of the rule:

Candidate's Experience matches Experience Criterion

is given as the product of Experience Match {0,1} with Profession Membership.

2.2.5. Foreign language–computer knowledge

The foreign language and the computer knowledge are recorded combined with fuzzy terms concerning the level of the knowledge (Average, Good, Very Good, Excellent) (Table 4).

Table 3
Fuzzy relation of candidate's education (C) and education request (R)

C	R		
	Basic	Degree	Post-grad
Basic	1	0.7	0.5
Degree	1	1	0.7
Post-grad	1	1	1

Table 4
Fuzzy relation of candidate's F. language (computer knowledge) and F. language request (computer knowledge request)

C	R			
	Average	Good	Very good	Excellent
Average	1	0.5	0.3	0.1
Good	1	1	0.8	0.5
Very good	1	1	1	0.8
Excellent	1	1	1	1

The strengths s_5, s_6 of the rules:

Candidate's F. Language matches F. Language Criterion
Candidate's C. Knowledge matches C. Knowledge Criterion

are given as the Foreign Language match and Computer Knowledge match, respectively.

2.3. Training

Every rule has a contribution to the final decision which is denoted using a weight w_i for field i. These weights are constrained in the interval [0–1], where 0 means least significant and 1 very significant. These weights change using machine learning algorithms on example cases (Jang, 1993). We trained the weights of the FIS through training samples collected from previous enterprises hiring cases. If proposed post is actually accepted then it is considered as a positive sample and as a negative sample otherwise. The elements produced are placed into a file, the training set of the Neuro-Fuzzy. These data (inputs/outputs) are repetitively placed into the Neuro-Fuzzy network forming its weights. The system performance is checked through a test set data formed by the employment of the certain query.

Each training sample supplies the input and the output of the network with the desirable output. During training procedure, the network parameters (weights) are adapted through the training in a way to reduce the mean square error between the real and the desirable output. By the end of the training, the network should be able to treat completely new data (generalization capability). For formally, let (s^i, d^i), $i = 1, \ldots, n$ be the training data, where $s^i = (s^i_1, \ldots, s^i_6)$ the criteria satisfaction for training case i, and d^i is 1 if the proposed post for case i was accepted and 0 otherwise. Training aims at finding a weight vector w that minimizes the constrained linear least-squares (Matlab Optimization Tool box)

$$E(w) = \frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} s^i_j \cdot w_j - d_i\right)^2 = \min \tag{3}$$

s.t.

$$\mathrm{lb}_j \leq w_j \leq \mathrm{ub}_j$$

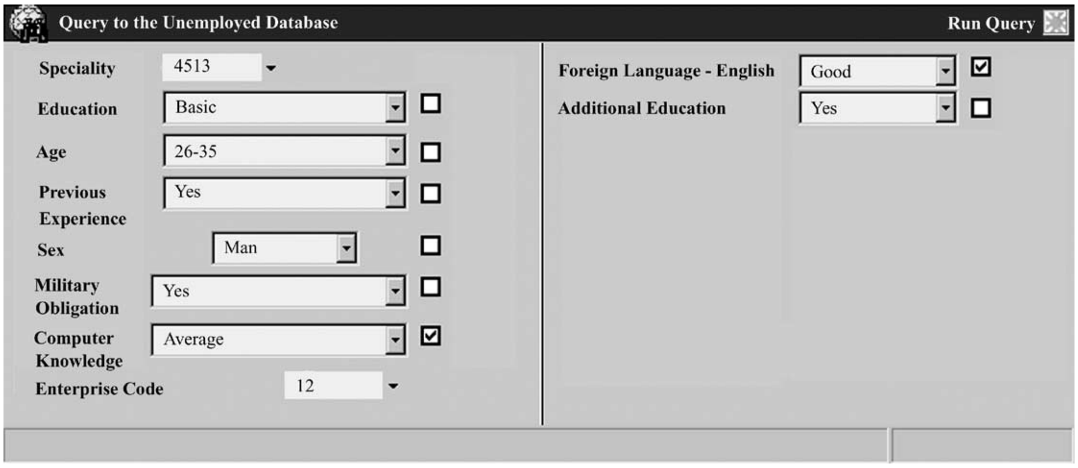


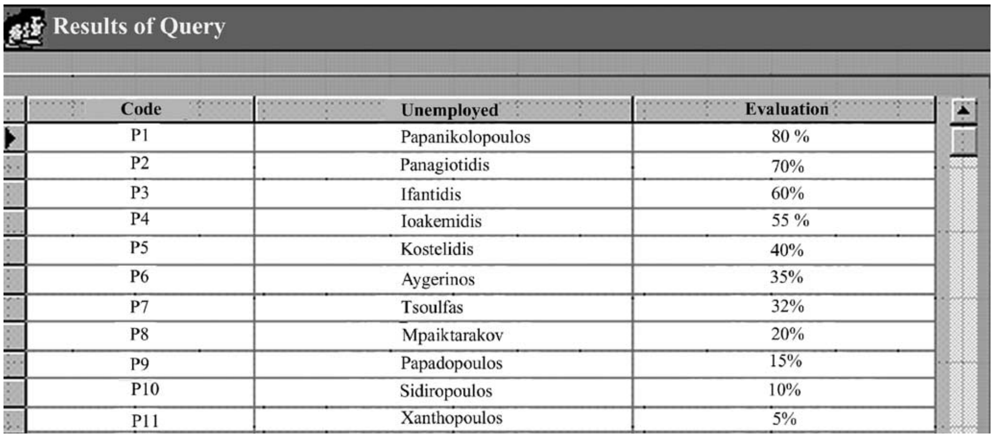
Fig. 4. The user interface of the query to the expert system. The CheckBox next to each input field enables exact or approximate logic.

swhere lb_j is the lower bound of weight j and ub_j is the upper bound. So, training is performed in such a way that provides us with inside knowledge of the decision process of the Neuro-Fuzzy network, while prior knowledge of the relative importance of criteria can also included in the system. Different combinations of lb_j and ub_j can be evaluated using the method of leave-one-out cross validation, where for a particular set of parameters $(lb_j, ub_j \; j = 1, \ldots, m)$ we repetitively train the network using $n - 1$ training data and use the data left as a test sample (Cherkassky & Mulier, 1998). The mean square error over all data when used as test samples (generalization capability) is the criterion to choose among constraints. Retraining is performed after a standard amount of new cases available (defined by the administrator).

3. Example

Till now, the General Secretariat of Social Training has registered 1.283 unemployed and 3.859 enterprises (0.34% primary, 32.83% secondary and 66.83% tertiary sector). In Appendix A we present a statistical analysis of the database regarding the evaluation criteria—inputs of the expert system. This macroscopic view of the database contributes better understanding of the unemployed special profile—the biggest percentage of registered users are resettlers, while the rest include released, detainees, underage offenders, single parent, gypsies, immigrants, former users and general population—and the expert system results. The age distribution appears in Fig. A1 where is clearly shown that most unemployed are aged between 26 and 35. Fig. A2 indicates that 63% have only basic education. Figs. A3 and A4 show that only 2% have been educated additionally and 15% have previous experience correspondently. Regarding the fields Foreign Language and Computer Knowledge, Figs. A5 and A6 show that most unemployed have average knowledge.

The present example concerns the expression of interest by an International Manufacturing enterprise for a new employee for furniture manufacturing (speciality


Results of Query

Code	Unemployed	Evaluation
P1	Papanikolopoulos	80 %
P2	Panagiotidis	70%
P3	Ifantidis	60%
P4	Ioakemidis	55 %
P5	Kostelidis	40%
P6	Aygerinos	35%
P7	Tsoulfas	32%
P8	Mpaiktarakov	20%
P9	Papadopoulos	15%
P10	Sidiropoulos	10%
P11	Xanthopoulos	5%

Fig. 5. The results of the query. Clicking on the unemployed with the highest mark, we see his contact profile.

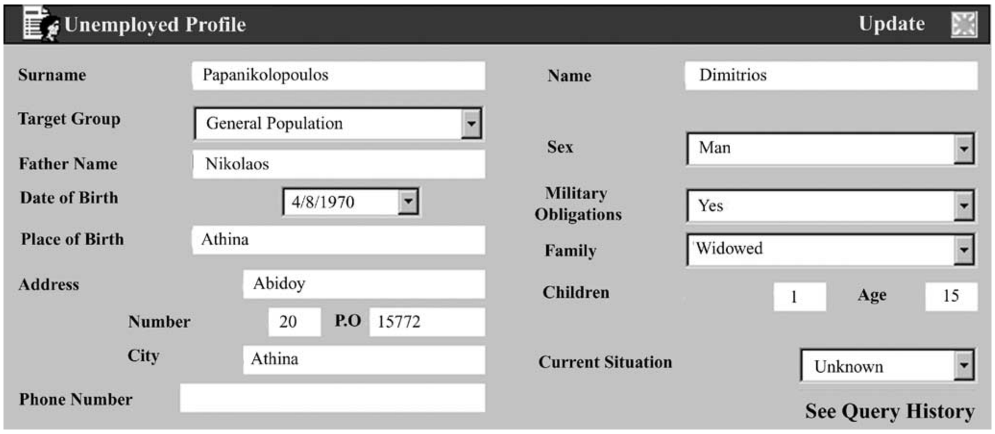


Fig. 6. The profile of the unemployed. The ‘See Query History’ link retrieves the history of the unemployed queries.

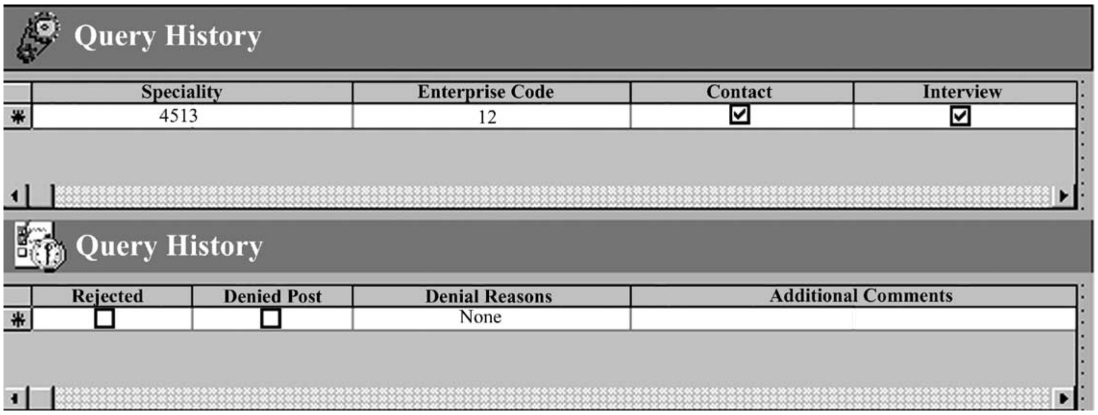


Fig. 7. The history of the relevant queries. ‘Contact’, ‘Interview’, ‘Rejected’, ‘Denied Post’ checkboxes are used to register contact with the unemployed, interview with the employer, rejection, denial of post correspondently. Denial Reasons and Additional Comments are used for informational reasons. Query History is used for retraining at the certain speciality.

code 4513). The administrator of the system adds the enterprise profile to the database. The speciality of the post and its qualifications are registered into the query form (Fig. 4). The parameters of the query consist the inputs of the expert system. Checking or not the checkbox next to each field is used to enable the use of exact or approximate logic correspondently. In the present use case study, only average Computer Knowledge and good Knowledge of English are considered as demandable and for this are checked as exact part. Pressing the ‘Run Query’ button, the administrator poses the present query to the expert system to match the unemployed with the certain post. The results of the expert system, concerning the query of Fig. 4 appear in Fig. 5 along with the evaluation mark of the unemployed (0–100%). The unemployed with the highest mark is recommended to the enterprise for interview. Clicking on the unemployed with the highest mark, we see his/her personal profile (Fig. 6) created at his registration. Clicking on ‘See query history’, we see the history of the query (Fig. 7).

4. Conclusions

In this paper, a hybrid expert system is developed for job matching. The application was applied to a special social class of unemployed described macroscopically in Appendix A. The performance of the system (percentage of applicants that were hired) was examined using weights manually selected with equal importance ($w_i = 0.5$) instead of weights learned by training. Although relative performance for every speciality is different, learning weights lead to an average 10% increase in job matching. The fields of age and previous employment were found to be very strongly weighted by the system. This fact may be due to the special profile of the unemployed (most being resettlers).

Acknowledgements

Research work of this paper is partly funded by the Greek General Secretariat for Research and Technology (GSRT), through the research project 'SINXRIMATODOTISI'.

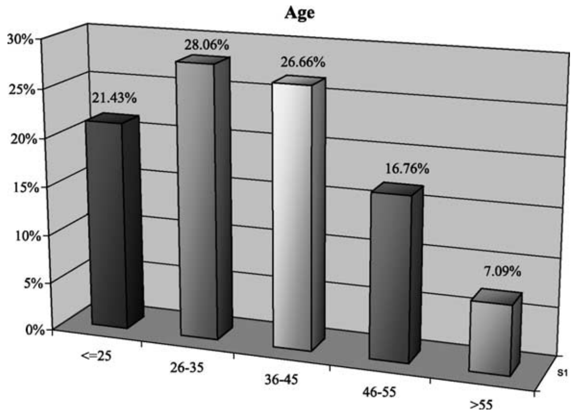


Fig. A1. Age.

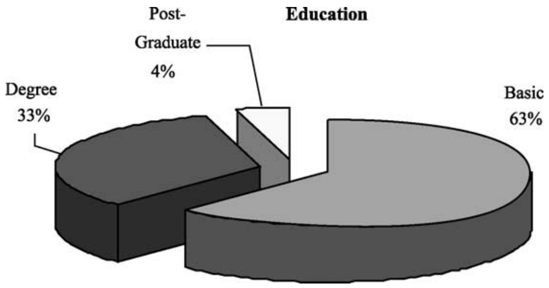


Fig. A2. Education.

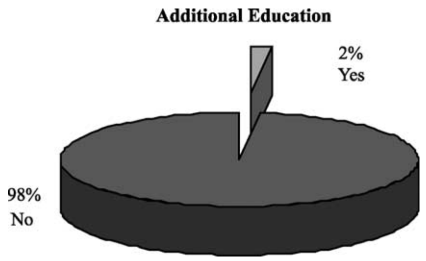


Fig. A3. Additional education.

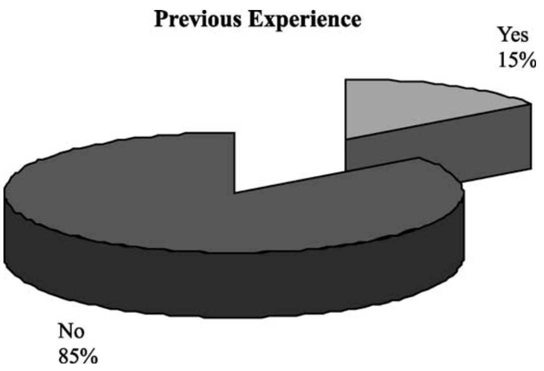


Fig. A4. Previous experience.

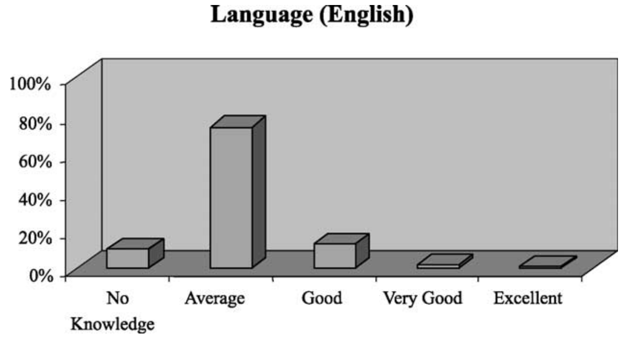


Fig. A5. Foreign language.

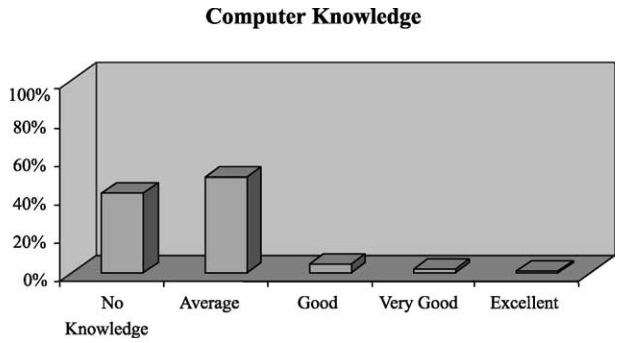


Fig. A6. Computer knowledge.

Appendix A

Figs. A1–A6

References

A better job fast http://www.myjobsearch.com.

Cherkassky, V., & Mulier, F. (1998). *Learning from data*. New York: Wiley.

Europe's career market on the internet (2001). http://jobpilot.co.th.

Federal job (2001). http://www.fedworld.gov/jobs/jobsearch.html.

Gams, M., Golob, P., Karaliø, A., Drobniø, M., Grobelnik, M., Glazer, J., Pirher, J., Furlan, T., Vrenko, E., & Križman, R. (1998). *EMA—zaposlovalni agent*, http://www-ai.ijs.si/~ema/EMA_Info-e.html.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on System, Man, and Cybernetics*, *23*(5/6), 665–685.

Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence*. *Matlab curriculum series*, Englewood Cliffs, NJ: Prentice-Hall.

Jobsdb (1998–2001). *Top jobs and careers in asia and around the globe*. http://www.jobsdb.com.

Kouremenos, S., Vrettos, S., & Stafylopatis, A. (2003). *An intelligent agent-mediated web trading environment*. *IEEE/WIC international conference on web intelligence (WI 2003)*, in press.

Labate, F., & Medsker, L. (1993). Employee skills analysis using a hybrid neural network and expert system. In *IEEE international conference on developing and managing intelligent system projects*. Los Alamitos, CA, USA: IEEE Comput. Soc. Press.

Matlab optimization toolbox: http://www.mathworks.com/products/optimization/.

Rafter, R., Bradley, K., & Smyth, B. (2000). Personalised retrieval for online recruitment services. In *Proceedings of the 22nd annual colloquium on IR research*. Cambridge, UK.

Sugeno, M., & Kang, G. T. (1998). Structure identification of fuzzy model. *Fuzzy Sets and Systems*, *28*, 15–33.

Vrettaros, J. (1996). Fuzzy connectionist systems for student modeling. In *Proceedings of the Annual International Conference on Technology and Education*.

Vrettos, S., & Stafylopatis, A. (2001). A fuzzy rule-based agent for web retrieval—filtering. In *Proceedings of first Asia–Pasific conference on web intelligence* (pp. 448–453). Maebashi City, Japan.