An Effective Fuzzy Recommender System for Fund-raising Management

Published Version:
L. Barzanti, S.G. (2020). An Effective Fuzzy Recommender System for Fund-raising Management. Singapore : Springer [10.1007/978-981-13-8950-4_8].

Availability:
This version is available at: https://hdl.handle.net/11585/787071 since: 2021-01-07

Published:
DOI: http://doi.org/10.1007/978-981-13-8950-4_8

An effective fuzzy Recommender System for fund raising management

Luca Barzanti[1], Silvio Giove[2] and Alessandro Pezzi[3]

[1] Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126, Bologna, Italy
luca.barzanti@unibo.it
[2] Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121, Venice, Italy
sgiove@unive.it
[3] School of Economics and Management, University of Bologna, P.le della Vittoria 15, 47121, Forlì, Italy
a.pezzi@unibo.it

Abstract. In the social Economics field that deals with the Non Profit Organizations (NPO's), the fund raising is a crucial activity that requires the management of a great number of quantitative and qualitative information regarding Donors and Contacts (*i.e.* potential donors). This data is normally stored in a structured Data Base (DB) by each NPO, and it is clear that their effective processing by Data Science methods significantly improves the performances of the fund raising campaigns. For this reason, the use of rigorous mathematical methods and Decision Support Systems (DSS) has been playing a very important role in this context.

The process of fund raising is very complex and in part different depending on the characteristics of each Organization. However, a common important feature is the role of the Contacts, therefore the method for turning the Contacts into actual Donors contextualized in the so called "giving pyramid" is crucial from a strategic point of view. Recently a Recommender System (RS) has been proposed to optimize the Contacts' management, by computing the similarity of each Contact with respect to the Donors.

In this contribution, we enhance and complete this model by considering both a large DB and two significant extensions of the model, obtaining in this way an effective and whole fuzzy RS. With respect to the DB, the availability of information is effectively exploited. As for the algorithm, a proper similarity measure is defined, based on the specificity of the context. Moreover, a complete estimation of the Contacts' characteristics is taken into account, by considering not only the frequency but the averaged amount of the gift as well, in the context of a non-parametric approach. The experimental results show the effectiveness of the proposed system.

Keywords: fund raising management, similarity measures, fuzzy recommender systems, non-parametric estimation.

1 Introduction

A crucial support for NPOs, which operate in the context of Social Economics, is the fund raising (FR) activity, in which the resources for the mission of the Association are collected, see (Andreoni, 2005), (Rosso, 2004). FR strategies are therefore crucial for the achievement of the mission and, specifically, to reaching the goal of the current campaign, see (Nudd, 2003), (Sargeant, 2001). Quantitative methods employing DB technologies have been studied and developed in the pertaining literature for making these strategies more effective, see (Flory, 2001-a), (Flory, 2001-b), (Kercheville, 2003). The effective use of the information on Donors and Contacts (i.e. potential donors), which is in fact normally managed by an organized DB, is crucial for optimizing the resources for the campaign by selecting the most promising Donors/Contacts for the considered context. For a few years, the literature in the area of mathematical models and DSS has dealt with the fund raising problem, determining an evolution, a strengthening and a specialization of the proposed methods and algorithms. With respect to the first two objectives, see (Verhaert, 2012), (Barzanti, 2013). Moreover, the consideration of the specific NPO's characteristics and their consequences in the modeling process is developed e.g. in (Barzanti, 2009) and (Barzanti, 2012). In these contributions and in the operative literature, see e.g. (Melandri, 2017, p. 7), it is also documented that the operative word of Associations has been using the results that have been achieved in this field. Moreover, in (Moro, 2017) it is showed the analogy between the fund raising process and some bank activities and the consequent correspondence of the employed methodologies.

One of the goals of a loyalty campaign is to involve new people in the mission of the Association by their first donation. In operational language, the goal is to make some Contacts going up from the ground of the so called "giving pyramid", see (Melandri, 2017), to the first level. From a strategic point of view, it is particularly important that a Contact becomes Donor at the first gift request, and therefore it is fundamental to solicit the gift in a campaign that is suitable for that Contact.

In this paper, we focus our attention on such an essential aspect of the fund raising process, by determining a *recommender system*, see (Jananch, 2001), a branch of *Artificial Intelligence*[1], that identifies the more suitable Contacts to reach in the current campaign. In fact, reaching a person obviously has a cost (that depends on the type of request) and the budget constraint requires a choice. For this purpose, a suitable *similarity measure* can be used for matching the profile of the Contacts with the profile of all the regular Donors in the Association DB who usually donate for similar campaigns. In particular, by using the results obtained in Econometrics (Duncan, 1999), (Lee, 1999), and more specifically in (Cappellari, 2011), (Duffy, 2007), the variables of the personal profile that influence the gift probability are selected.

The problem of the *Contacts' management* was before considered in (Barzanti, submitted), where a first attempt of a recommender system has been implemented. Although a relevant effort has been performed for modeling the process, the numerical

[1] Refer to (Russell, 2003) for an exhaustive introduction to algorithms and aim of Artificial Intelligence.

experiments were limited to a few Contacts and Donors. Conversely, in the present paper we focus on a large structured DB, that implies a relevant improvement of the significance of the results. The large amount of information is also exploited by the consideration of all the Donors' historical data, in particular of the gift volume, in addition to the frequency. This also allows the estimation of an amount for each Contact and the completion of the analysis, through the joined computation and elaboration of both the significant quantities.

Moreover, a specific measure of similarity is developed, that takes into account with appropriate weights the most relevant variables of the personal profile, which are significant for the pertaining literature for measuring the gift probability.

The numerical experiments are developed with the contribution of ASSIF, the Italian fund raiser association, and Philanthropy, a University fund raising research center, that ensure the reliability of the proposed approach.

2 Contacts characterization

In this Section, we consider the Contacts characterization. In the Subsection 2.1, we analyze the similarity between a Donor and a Contact, based on a set of personal data, namely the financial situation, the age and the qualification. The most important criterion, the financial situation, is considered as necessary for the similarity, and thus a novel similarity function is proposed[2]. Subsequently, in Subsection 2.2 we compute, for each, an estimation of the Contact's expected gift Volume and Frequency, using a non-parametric approach based on the similarity between the considered Contact and each Donor in the DB. This non-parametric method is inspired by the Fuzzy Nearest Neighbor techniques, see (Aman, 2013), (Keller,1985).

2.1 The personal characterization for Donors and Contacts

The pertaining literature suggests that the propensity to gift depends on some personal parameters, like the financial situation (the most significant), the age, the risk aversion and other, see e.g. (Cappellari, 2011) and (Duffy, 2007). In this contribution, among others, we consider as most representative the financial situation (*Fin*), the age (*Age*) and the qualification (*Qual*), that are:

1) *Fin* (Real), the global income amount.
2) *Age* (real).
3) *Qual* (label), with 4 ordered values: P (PhD), B (Bachelor), H (High School), O (Other).

The measure of the other parameters, as for the risk aversion, can indeed be difficult, while their influence on the gift attitude can be debatable (for instance, the presence of

[2] For general properties and a list of possible similarity measures, the interested reader can refer to (Couso, 2013), (Beg, 2009).

children can be source of effects of opposite sign). About the possible influence of such parameters on the propensity to gift, a future and deeper analysis is advisable, but it is beyond the purpose of this contribution. For this reason, in this paper we consider only the three most important parameters, *Fin*, *Age*, *Qual*, even if the inclusion of other variables would not change the methodological framework. The *Fin* parameter is clearly the most significant, given that a strong correlation exists among the personal richness and the gift probability (and amount). Moreover, this represents a necessary condition, given that financial scarcity is normally a serious obstacle for the charity. In absence of similarity for *Fin* between a Donor and a Contact, the entire similarity is null, and cannot be compensated by high values of similarity between *Age* and *Qual*. The personal variables can be collected into a vector, and thus the *i*-th Donor can be represented as the vector:

$$W(i) = \{Fin(i), Age(i), Qual(i), f(i), v(i), r(i)\}$$

being (F, V, ρ) the frequency, the (average) volume of past gifts and the statistical robustness. The reader can refer to (Barzanti, submitted) for a detailed computation of (F, V, ρ). The first three components of the vector W_i are used to compare the *i*-th Donor with a Contact, using a suitable similarity measure, given that the Contact personal data are known. For this purpose, it is convenient to consider a given Contact and, for its value of the *Fin* parameter, to define a fuzzy moving window, centered in the Contact value. The same is done for the *Age* parameter, but for *Fin* the amplitude depends on the value of the variable itself. Let us define the following moving fuzzy window:

$$MOV(a,b,c,d) = \begin{cases} 1, & b \le x \le c \\ \dfrac{x-a}{b-a}, & a \le x < b \\ \dfrac{d-x}{d-c}, & c < x \le d \\ 0, & x < a, \text{ or } x > d \end{cases} \tag{1}$$

The fuzzy window for *Fin* is then defined as follows, given that, for the considered *j*-th Contact, is $Fin_j = \lambda$:

$$G_j(x) = MOV(a_j, b_j, c_j, d_j) \text{, where}$$

$$a_j = \lambda(1-\beta), \quad b_j = \lambda(1-\alpha), \quad c_j = \lambda(1+\alpha), \quad d_j = \lambda(1+\beta) \tag{2}$$

and $0 < \alpha < \beta < 1$ define the upper and lower amplitude of the moving fuzzy window, both monotonically dependent on the current value of $Fin_j = \lambda$. Then, the *Fin* similarity between the *j*-th Contact and the *i*-th Donor, $SimFin(i,j)$, is $SimFin(i,j) =$

$G_j(Fin_i)$. A similar formulation can be applied to the *Age* parameter, but, in this case, the amplitude can be considered as a constant. Thus, if $Age_j = \mu$:

$$H_j(x) = MOV(e_j, f_j, g_j, h_j)$$
$$e_j = \mu - \gamma,\ \ f_j = \mu - \delta,\ \ g_j = \mu + \delta,\ \ h_j = \mu + \gamma \tag{3}$$

and $SimAge(i,j) = H_j(Age_i)$ with $\delta < \gamma$. Finally, with respect to the third variable *Qual*, formed by ordered classes, we consider the similarity between two classes as a two entries function:

$$SimQual(i, j) = L_j(Qual_i, Qual_j) \tag{4}$$

This requires the assessment of 6 parameters (all the possible combinations of two different classes).
Finally, the complete similarity between the Donor *j* and the Concact *i* is:

$$Sim(D_i, C_j) = SimFin(i, j)\frac{(\omega_1 + \omega_2 SimAge(i, j) + \omega_3 SimQual(i, j))}{\omega_1 + \omega_2 + \omega_3} \tag{5}$$

where $\omega_1,\ \omega_2,\ \omega_3$ are suitable weights ($\omega_1,\ \omega_2,\ \omega_3 > 0$, $\omega_1 + \omega_2 + \omega_3 = 1$).

2.2 Non-parametric estimation of Volume and Frequency

A possible way to produce a numerical estimation of an unknown distribution from a set of data is based on the *Nearest Neighbor* function approach, see (Keller, 1985), (Bhatia, 2010). Roughly speaking, the probability estimation is based on an average combination of the occurrences of the items, weighted by the similarity degree between each item and the case. In some cases, a *kernel function* is formed by the set of items that are *similar* to the current case, see (Wand, 1995), (Canestrelli, 2007). The unknown probability is estimated using, as conditioning parameters (weights), the kernel-based similarities of the case study to each item inside a suitable neighbor. The closer the item, the greater the weight, given that more "importance" (credibility) is given to the most similar items. In (Fan, 1996), the reader can find a complete and detailed explanation of the kernel-based estimation methods, while in (Keller, 1985) a fuzzy *K*-Nearest Neighbor is described. From the quoted literature, if $B = \{b_1, b_2, ..., b_n\}$ is a set of input vectors (*incomplete* pattern, see (Giove, 1999)), with corresponding output value $Y = \{y(b_1), ..., y(b_n)\}$, the expected value of the probability distribution of an (other) item *b*, conditioned to *B*, can be computed as:

$$E(F) = \frac{\sum_{b \in B} Sim(a,b) \cdot y(b)}{\sum_{b \in B} Sim(a,b)} \tag{6}$$

We use this formulation to compute both the (expected value of) Frequency and Volume, using as similarity the function described in Sec. 2.1, corrected by the statistical robustness of the Donor. Namely, for the same reason as for the similarity, we give less weight to those Donors who are less robust, in the sense that their own history about the gifts is limited to few requests (and corresponding answers, if any). Thus, let $F(j)$, $V(j)$ be the estimated values of (average) Frequency and Volume of the Contact j:

$$F(j) = \frac{\sum_{i=1}^{n} Sim(i, j) \cdot r(i) \cdot f(i)}{\sum_{i=1}^{n} Sim(i, j) \cdot r(i)} \tag{7}$$

$$V(j) = \frac{\sum_{i=1}^{n} Sim(i, j) \cdot r(i) \cdot v(i)}{\sum_{i=1}^{n} Sim(i, j) \cdot r(i)} \tag{8}$$

Moreover, to avoid a meaningless estimation, we impose that at least K Donors exist, with similarity and robustness greater than a specified threshold, otherwise the Contact is classified as *undecidable* and is inserted in a separated list. Then, if we define $SR(i) \in \{Sim(i, j) \cdot r(i)\}$ as the ordered similarities multiplied by the robustness, i.e. $SR(1) \geq SR(2) \geq ... \geq SR(n)$, the rule classifies a Contact as undecidable if:

$$\sum_{i=1}^{K} SR(i) < \sigma \tag{9}$$

being $\sigma > 0$ a specified threshold (i.e. $\sigma = K$). Finally, for all the promising Contacts, a proxy of the estimation's robustness is computed as the sum of all the similarities:

$$R(j) = \sum_{i=1}^{n} Sim(i, j) \cdot r(i) \tag{10}$$

The considered Contact is thus characterized by the 3-ple: $F(j), V(j), R(j)$. This information can be used for the Contacts' evaluation and/or screening, based on the system's User own preferences, and thus constitutes the engine of a Decision Support System in the field of the Non Profit Organizations. The numerical results show the good performances of the proposed approach.

3 Computational results

The numerical experiments have been performed in collaboration with ASSIF[3] and Philanthropy Centro Studi[4]; this way, a real context is considered, in particular with respect to the simulated DB construction, implemented by the criteria of the giving pyramid for a medium-sized Organization. A SQL Server DB with 10000 Donors and 2000 Contacts is considered, while the system is designed in Visual Basic, using SQL language.

The Decision Maker (DM) sets up the similarity parameters, with reference to the moving fuzzy windows in formulas (2), (3), the assessment required by (4) and the weights in (5). The thresholds of undecidability (formula (9)) and robustness are also set (see below). Fig. 1 shows the whole graphical interface.

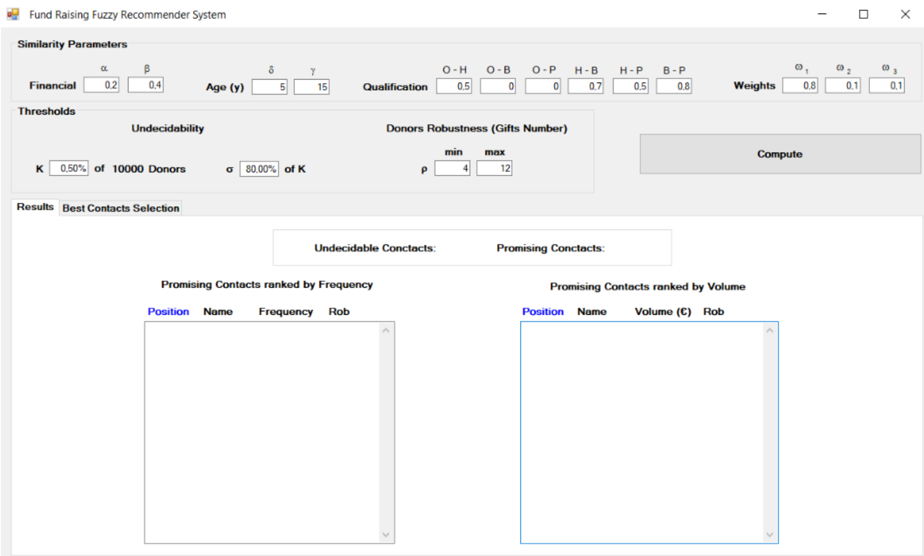


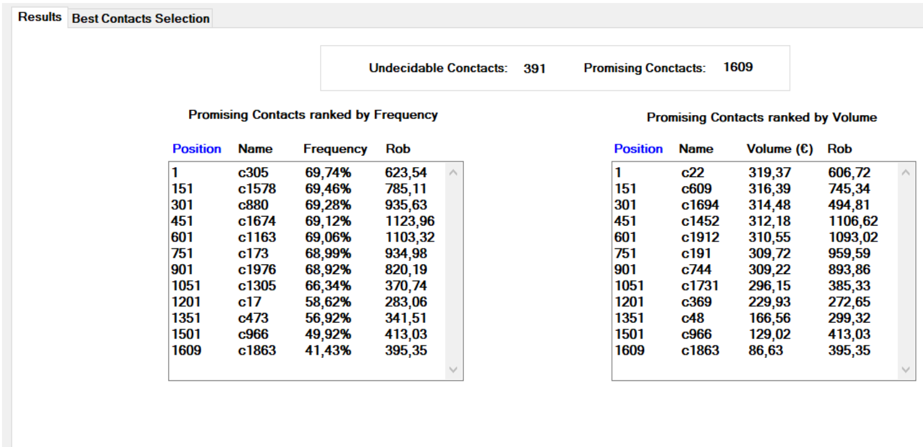
Fig. 1. The user interface

The parameters are set like in Fig. 1. In particular, notice that, for the undecidability, K is set as (the integer part of) a percentage of the DB's numerosity and σ is a specified percentage of K, while for the Donors' robustness ρ, the thresholds are set in terms of gifts number, referring to (Barzanti, submitted), as pointed out in Section 2.1.

The results of the estimated Contacts' gift Frequencies and Volumes (with reference to formulas (7) and (8)) and of the robustness index (*Rob*) for the estimation (formula (10)) are presented in Fig. 2.

[3] The Italian fund raiser association

[4] Research center on non profit, fund raising and social responsibility operative in the University of Bologna



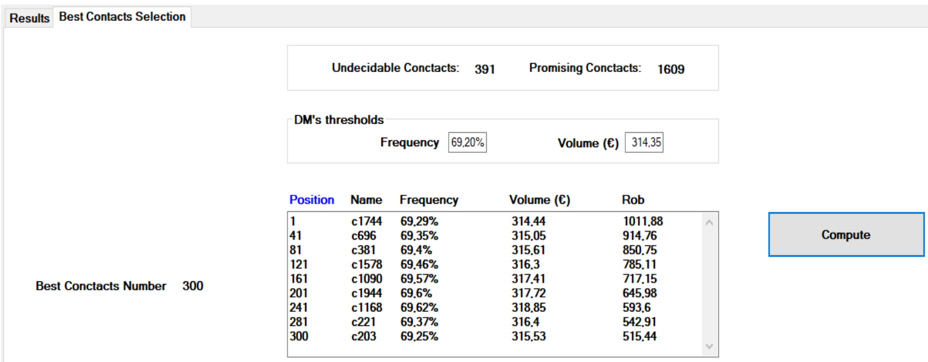
Position	Name	Frequency	Rob
1	c305	69,74%	623,54
151	c1578	69,46%	785,11
301	c880	69,28%	935,63
451	c1674	69,12%	1123,96
601	c1163	69,06%	1103,32
751	c173	68,99%	934,98
901	c1976	68,92%	820,19
1051	c1305	66,34%	370,74
1201	c17	58,62%	283,06
1351	c473	56,92%	341,51
1501	c966	49,92%	413,03
1609	c1863	41,43%	395,35

Position	Name	Volume (€)	Rob
1	c22	319,37	606,72
151	c609	316,39	745,34
301	c1694	314,48	494,81
451	c1452	312,18	1106,62
601	c1912	310,55	1093,02
751	c191	309,72	959,59
901	c744	309,22	893,86
1051	c1731	296,15	385,33
1201	c369	229,93	272,65
1351	c48	166,56	299,32
1501	c966	129,02	413,03
1609	c1863	86,63	395,35

Fig. 2. Rankings of the estimated gift Frequencies and Volumes, with the corresponding robustness index, for the promising Contacts. Number of undecidable Contacts.

For both the Frequency and the Volume, a ranking of promising Contacts is showed. Moreover, the number of undecidable Contacts is exhibited. Note that the DM can choose how deep to explore the rankings.

In Fig. 3, the result of a composite selection query (ordered by *Rob*) is displayed, where those Contacts are selected, that exceed both the two prefixed thresholds by the DM, one for each estimated characteristic (Frequency and Volume). For the sake of brevity, these Contacts are called “best Contacts”.



Position	Name	Frequency	Volume (€)	Rob
1	c1744	69,29%	314,44	1011,88
41	c696	69,35%	315,05	914,76
81	c381	69,4%	315,61	850,75
121	c1578	69,46%	316,3	785,11
161	c1090	69,57%	317,41	717,15
201	c1944	69,6%	317,72	645,98
241	c1168	69,62%	318,85	593,6
281	c221	69,37%	316,4	542,91
300	c203	69,25%	315,53	515,44

Fig. 3. The best Contacts

Notice that the DM can set the thresholds in order to obtain a prefixed number of Contacts to reach (a unit cost is involved, see the Introduction), in this case 300, displayed in the Figure with a step of 40, in order to satisfy the budget constraint of the considered fund raising campaign.

4 Conclusion

In this paper, we developed an effective recommender system for fund raising management, which extends into an actual context a previous approach, by using a large structured DB. This extension involves a significant improvement of the basic method, by considering both a proper similarity measure and the entire available information (including the Volume), in the context of a non-parametric approach. The numerical study, realized in collaboration with two Italian operative research centers on fund raising, gives evidence of the effectiveness of the proposed approach.

References

1. Aman Kataria, Singh, M. D.: A Review of Data Classification Using K-Nearest Neighbor Algorithm. International Journal of Emerging Technology and Advanced Engineering Website 3(6), 354-360 (2013).
2. Andreoni, J.: Philantropy. In Kolm, S. C. and Ythier, J. (eds), Handbook the Economics of Giving, Altruism and Reciprocity 2, Elsevier, 1201-1269 (2006).
3. Barzanti, L., Gaspari, M., Saletti, D.: Modelling decision making in fund raising management by a fuzzy knowledge system. Expert Systems with Applications 36, 9466-9478 (2009).
4. Barzanti, L., Giove, S.: A decision support system for fund raising management based on the Choquet integral methodology. Expert Systems 29 (4), 359-373 (2012).
5. Barzanti, L., Giove, S., Pezzi, A.: A Recommender System for fund raising management (submitted).
6. Barzanti, L., Mastroleo, M.: An enhanced approach for developing an expert system for fund raising management. In Segura, J. M. and Reiter, A. C. (eds), Expert System Software: Engineering, Advantages and Applications. NYC Nova Science Publishers, 131-156 (2013).
7. Beg, I., Ashraf, S.: Similarity measures for fuzzy sets. Appl. and Comput. math, 2(8), 192-202 (2009).
8. Beliakov G., Calvo T., James S.: Aggregation of Preferences in Recommender Systems. Recommender Systems Handbook, Springer, New York, 705-734 (2011).
9. Bhatia, N, Vandana, A.: Survey of Nearest Neighbor Techniques. International Journal of Computer Science and Information Security 8(2), 302-305 (2010).
10. Bridge, D., Goker, M., McGinty, L., Smyth, B.: Case-based recommender systems. The Knowledge Engineering Review 20(3), 315-320 (2006).
11. Canestrelli, E., Canestrelli, P., Corazza, M., Filippone, M, Giove, S., Masulli, F.: Local learning of tide level time series using a fuzzy approach. In Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17 (2007).
12. Cappellari, L., Ghinetti, P., Turati, G.: On time and money donations. Journal of Socio-Economics 40(6), 853-867 (2011).
13. Couso, I., Garrido, L., Sànchez, L.: Similarity and dissimilarity measures between fuzzy sets: A formal relational study. Information Sciences 229, 122-141 (2013).
14. Duffy, J., Ochs, J., Vesterlund, L.: Giving little by little: Dynamic voluntary contribution game. Journal of Public Economics 91(9), 1708-1730 (2007).
15. Duncan B.: Modeling charitable contributions of time and money. Journal of Public Economics 72, 213-242 (1999).

16. Fan, J., Gijbels, I.: Local polynomial modelling and its applications. Chapmann & Hall, London (1996).
17. Flory P.: Fundraising databases. DSC London (2001-a).
18. Flory P.: Building a fundraising database using your PC. DSC London (2001-b).
19. Giove, S., Pellizzari, P.: Time series filtering and reconstruction using fuzzy weighted local regression. Soft Computing in Financial Engineering, Physica-Verlag, Heidelberg, 73-92 (1999).
20. Jananch D., Zanker M., Felfernig A., Friedric G.: Recommender Systems: an introduction. Cambridge University Press (2001).
21. Keller, J. M., Gray, M. R., Givens, Jr., J. A.: A Fuzzy K-Nearest Neighbor Algorithm. IEEE Transactions on Systems, Man, and Cybernetics Smc-15(4), 580-585 (1985).
22. Kercheville, J., Kercheville, J.: The effective use of technology in nonprofits. In Tempel, E. (eds), Hank Rosso's achieving excellence in fund raising; John Wiley & Sons, 366-379 (2003).
23. Lee, L., Piliavin J. A., Call, V. R.: Giving time, blood and money: similarities and differences. Social Psychological Quarterly 62 (3), 276-290 (1999).
24. Melandri, V.: Fundraising. Civil Sector Press. Toronto – Canada (2017).
25. Moro, S., Cortez, P., Rita, P.: A divide-and conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. Expert Systems DOI: 10.1111/exsy.12253, pp. 1-13 (2017).
26. Nudd, S. P.: Thinking strategically about information. In Tempel, E. (eds.), Hank Rosso's achieving excellence in fund raising. John Wiley & Sons, 349-365 (2003).
27. Rosso, H., Tempel, R., Melandri, V.: The Fund Raising Book. Bologna: ETAS (in Italian) (2004).
28. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. 2nd edn. Prentice Hall (2003).
29. Sargeant, A.: Using Donor Lifetime Value to Inform Fundraising Strategy. Nonprofit Management & Leadership, John Wiley and Sons Inc 12(1), 25-38 (2001).
30. Verhaert, G. A., Van den Poel, D.: The role of seed money and threshold size in optimizing fundraising campaigns: Past behavior matters!. Expert Systems with Applications 39, 13075-13084 (2012).
31. Wand, M. P., Jones, M. C.: Kernel Smoothing. Chapmann & Hall, London (1995).