PROPOSING A SIMILARITY MEASURE IN CASE BASED REASONING FOR PRODUCTS SELECTION
An Experimental Evidence

Fadi Amroush
Department of Economic Theory and History, University of Granada, Granada, Spain

Keywords: Decision Support Systems, Case Based Reasoning, Products Selections.

Abstract: This paper presents a novel similarity measure to design a Decision Support System for products selection using Case Based Reasoning "CBR". The presented approach combines a novel local similarity measure with Nearest Neighbour Matching Function which is used as a typical evaluation function to compute the nearest-neighbour matching case in CBR. This paper suggests using this similarity measure in CBR in order design our model in products selection to help users to find the optimal product according to their preferences. The nature of this local similarity measure is to give more reality measure used by people in selecting products instead of the traditional one proposed by (Xiao-tai et al., 2004). We illustrate the significance of our proposed measure experimentally. The paper shows that our approach has been followed by about 80% of subjects.

1 INTRODUCTION AND LITERATURE REVIEW

The product selection will become more important nowadays especially as online products can be used to give online consumers a better choice of products than can be found in traditional shops. Selecting what products you'll buy is one of the most common decisions a consumer makes, mostly the consumer selects products based on his personal preferences or hot-item lists consideration. Let us assume that you want to buy a laptop, and there are many types of products different in prices and characteristics. The main question is how are you going to select the best product -laptop- that meets your preferences within the budget you allocated, and without paying more for characteristics you don't need?

In this work, we focus on proposing a decision support system to help people in selecting the optimal product using CBR technique. The concept of optimality has first appears in 1890, it means "most favourable", that is meaning to look for the best possible compromise solution to a problem, when there are several competing considerations, not all of which can be simultaneously maximized. The solution is induced by applying CBR steps through the process of retrieving the stored cases, calculating the similarity ratio between these cases and the new case, and then selecting the most similar case. The novel idea here is to suggest a new similarity metric for products selection and test its significance experimentally.

Using the Case based Reasoning "CBR" techniques in the decision making process is one of many methods raised with the appearance of data mining techniques (Kolodner, 1991). CBR is implemented in large scale in many arias, Yang et al (2009) presented in his paper a Case Based Reasoning Decision Support System (CBR-DSS) that assists contractors in solving mark up estimation problem. This proposed CRR-DSS uses successful cases of previous completed projects to derive solution to new project mark up estimation problem, the principle of the CBR-DSS was to analogy new project with previous projects. Schmitt and Bergmann (1999) suggested applying CBR technology for Product Selection and Customization in Electronic Commerce Environments. Another study is given by Lin et al (2010) which focused on strategy selection for product service system design, in this study CBR is utilized to provide suggestions for finding this appropriate strategy. Ricci and Werthner (2002) adapted a case based querying for travel planning recommendation, it adapts CBR to

Amroush F..
PROPOSING A SIMILARITY MEASURE IN CASE BASED REASONING FOR PRODUCTS SELECTION - An Experimental Evidence.
DOI: 10.5220/0003746504990502
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 499-502
ISBN: 978-989-8425-95-9

provide personalized recommendations system based on previous system experience, and it applies query refinement methods helping to adjust queries according to the data available in a given product catalogues

2 OUR PROPOSED SIMILARITY MEASURE

In this section we will propose our similarity measure and how we use it in our model.

2.1 Nearest Neighbour Function

Nearest-neighbour retrieval is one of the most Nearest-neighbour retrieval is a famous approach that computes the similarity between stored cases and new input case based on weight features. Nearest-neighbour retrieval is a simple approach that computes the similarity between stored cases and new input case based on weighted features. Traditionally, the similarity between queries, Q and a case, C is defined as the sum of the similarities of its constituent features multiplied by their relevance weights, as in the following equation Eq1.

$$similarity\ (CaseI\ ,CaseR\ ) = \frac{\sum_{i=1}^{n} w_i \times sim\ (f_i^I, f^R)}{\sum_{i=1}^{n} w_i} \quad (1)$$

Eq1. Nearest-neighbour evaluations function.

Where w_i is the importance weight of a characteristic, *sim* is the similarity function and f_i^I and f_i^R are the values for characteristic *i* in the input and retrieved cases respectively.
We have started from the typical evaluation function which is proposed by Kolondner, 1991 in equation (2). Local similarity *Sim* is calculated rationally as following (Xiao-tai et al., 2004):

$$sim(f_i^I, f_i^R) = 1 - \frac{\left|f_i^I - f_i^R\right|}{k_i} \quad (2)$$

Where k_i is the value scale of the characteristic i,

From our point of view, this function is not suitable for products selections, because when a product has a characteristic supported more than it is needed, it will be considered like than it is not supported.

In this paper we have enhanced this function to be more realistic for products selection. We suggest for product selections if the product has more support for a specific characteristic so that is not bad and also should not affect others. Thus the deference local similarity is not suitable.

we propose The following definition of the local similarity between two vectors "The ratio of the minimum level that the consumer wants to have in his product to the degree of support that the product has for this characteristic, but without exceeding 1". Let us assume:

- P: The degree of support a product has for the characteristic.
- C: The minimum level the consumer wants to have in the product for the characteristic.
  Local *Sim* can be computed as follows:

```
If C==0 then Sim = 0 else
   If C>P then sim = 1
    Else  Sim = P/C
```
(3)

2.2 Example

Let us assume that there are three products P1, P2, and P3, which have the following main characteristics A, B, and C. On the other hand there are given preferences "R" as following, the minimum level the consumer wants to contain in the product is as following for the same A, B, and C.

- The characteristic A is: Essential thus, has 10/10.
- The characteristic B is: Desirable thus, has 5/10.
- The characteristic C is: Not important thus, has 0/10.

Now there are the following products P1, P2 and P3 which have the degree of support of the characteristics A, B and C as following table2.

Table 1: The degree of support for three products.

	A	B	C
P1	7	5	0
P2	5	10	5
P3	10	1	0
R	10	5	0

2.2.1 Applying Our Proposed Local Similarity Metric

Applying Eq1, Eq3 for calculating full similarity:

```
Sim(P1,R)=(0.7)(0.66)+(1)(0.33)+(0)(0)=0.792.
Sim(P2,R)=(0.5)(0.66)+(1)(0.33)+(1)(0)= 0.66.
Sim(P3,R)=(1)(0.66)+(0.2)(0.33)+(0)(0)=0.726.
```

The sorting is P1, P3, P2, so in our approach P1 is the best option.

2.2.2 Applying Traditional Approach

Now we will Follow the traditional approach by applying Eq2, Eq3 for calculating full similarity:

```
Sim(P1,R)=(0.7)(0.66)+(1)(0.33)+(0)(0)=0.792.
Sim(P2,R)=(0.5)(0.66)+(0.5)(0.33)+(1)(0)=0.49
Sim(P3,R)=(1)(0.66)+(0.6)(0.33)+(0)(0)=0.858.
```

The sorting is P3, P1, P2, so in traditional approach P3 is the best option.

3 EVALUATION

Finding a concrete manifestation of the term "similarity" is usually the task of a knowledge engineer, from the basic knowledge, there are no experiments to investigate similarity measure for products selections, However, the last decade has seen a number of approaches that aim at using Machine Learning techniques to adjust the similarity assessment in Case-Based Reasoning.

3.1 The Experiment

An Experiment has been done to investigate how Subjects select products according to their preferences of the minimum level they want their product to contain, and are they follow the informatics system advice. In the Experiment we follow the section 3.1.1 in order to compare results.

There are 4 stages in the experiment, in each stage we asked the subjects to have a decision to select one product if they want to buy according to their given preferences. These questions are including: own opinion, opinion about the others selection, opinion after proposing our similarity measure, final decision after proposing a prize of 10 or 20 Euros if the selection is common among the others, this prize to assure the he will select seriously what he think is the best, not doing as lottery. Here is the forth stages, and theirs question to the subjects.

3.2 Running the Experiment

The experiment consisted of two sessions, in each session, there were twenty three participants. All of the sessions were conducted at Laboratory de Experimental Economic ¨EGEO¨ in the Faculty of Economic and Business, at University of Granada, Spain. Spanish is the language of experiment. The subjects were recruited from undergraduate courses and some post graduated in business and economics at the same university. The experiment was computerized using the Web based program developed by the author. There were two treatments differ in the amount of the prize. Each subject earned 5euros as show up fees and additional prize (20 Euros or 20 Euros, nothing). In average earned 16.95 Euros.

3.3 Results

The number of subjects who select P1 in stage 4 is increased in the second treatment, the other factors is similar, so there is no deference between the two sessions - treatments-, and there is no significant factors, we will handle with whole group in addition to see gender effect, males and females.

As mentioned above, three have 4 Stages, here figure5-which shows how many subjects select each product in each stage.

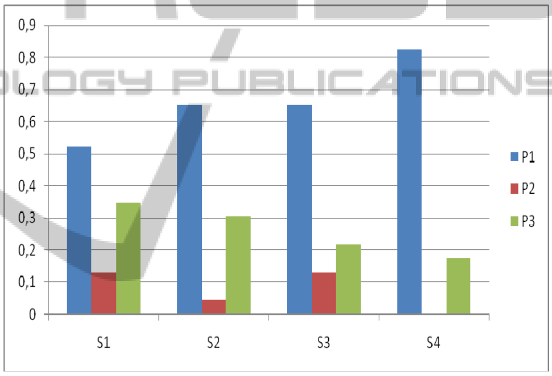

Figure 1: Number of subjects who select each product in each stage.

It is clear to notice that .number of subjects that selecting P1 is increasing stage after stage, as figure3 shows. Here is how the informatics system information affects decisions of subjects in paid stage S4.

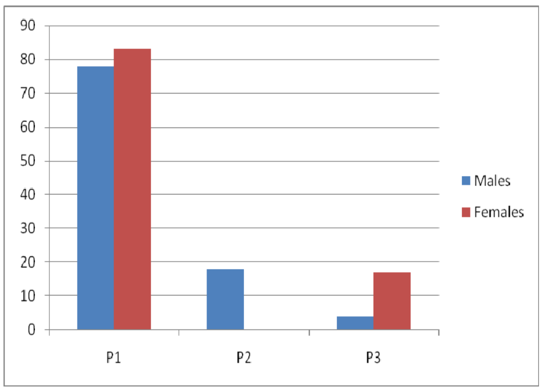

Figure 2: How many subjects select products at final stage.

3.4 Discussion

The results give an experimental evidence for our proposed similarity measure which find out that P1 is the best whereas the traditional one finds out that P3 is the best. About 80% of total subjects agree with our proposed similarity measure. (63% follow it regardless of the prize).

There is no age effect, Age is not important for males: The average of male's age in both with selecting P1 and without is 21.3, whereas the average of Female's age is 21.87, and who selected P1 22,05. The amount of prize is not affecting subjects in selecting more P1.

4 CONCLUSIONS

Characteristic based product selection well be famous especial on the Internet, and with integrating with e-commerce, so e- tailors must provide systems to support online products selection. Case-based reasoning is an approach that can provide a solutions to the problem of Products selection, all based on a knowledge representation and similarity metric.

In the context of CBR, we present in this paper a decision support model for products selections, we have presented a novel local similarity metric for products selection and compare it with traditional one. The evidence presented indicates the effective of our proposed similarity advise, and showed that subject follow it when it is presented as informatics advise. An experimental study is conducted to investigate how people select products, we reported the results and how subjects change their selections.

ACKNOWLEDGEMENTS

I acknowledge and warmly appreciate the tremendous support from my supervisors (Nikolaos Georgantzís, López Herrera, Antonio Gabriel) from Universidad de Granada, Thanks for their helpful comments.

REFERENCES

Aamodtn, A., Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations and system approaches, AICom - Artificial Intelligence Communications, *IOS Press*, Vol. 7: 1, (pp. 39-59).

Babka, O., Whar, S. Y. (1997). Case-based reasoning and decision support systems, Intelligent Processing Systems, ICIPS '97. *IEEE International Conference*, Issue Date: 28-31 Oct 1997, vol.2, (pp. 1532 - 1536).

Breese, J. S., Heckerman, D. (1995). Decision-Theoretic Case-Based Reasoning, *The Proceedings of the 5th International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, January, 1995.

Coyle, L., Doyle, S., Cunningham, P. (2004). Representing Similarity for CBR in XML, the *proceedings of the 7 the European Conference on Case Based Reasoning*, ECCBR, 2004, PP. 119-127.

Empirical Similarity Itzhak Gilboa, Offer Lieberman, David Schmeidler, *THE PINHAS SAPIR CENTER FOR DEVELOPMENT*, Discussion Paper No.3-2006, June, 2006.

Gayer, G., Gilboa I. (2007). Offer Lieberman, Rule-Based and Case-Based Reasoning in Housing Prices, the B.E. *Journal of Theoretical Economics*, Volume 7, Issue 1 2007 Article 10.

Gilboa, I., Schmeidler, D. (1995). Case-Based Knowledge and Planning. Discussion Papers 1127, Northwestern University, *Centre for Mathematical Studies in Economics and Management Science*.

Khoshgoftaar, T. M., Seliya, N ., Sundaresh N. (2006) An empirical study of predicting software faults with case-based reasoning, *Software Quality Journal - SQJ*, vol. 14, no. 2, pp. 85-111. DOI: 10.1007/s11219-006-7597-z

Kolodner, J. L. (1991) Improving human decision making through case based decision aiding, *AI Magazine*, 12(2), pp .52–68.

Kumar, V., Viswanadham, N. (2007). A CBR-based Decision Support System Framework for Construction Supply Chain Risk Management, *Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering,* Scottsdale, AZ, USA, Sept 22-25, 2007.

Kun, H L., Li, H S., Shing, S L., & Yung T L. (2010). Strategy selection for product service systems using case-based reasoning, *African Journal of Business Management,* Vol. 4(6), pp. 987-994, June 2010.

McDermott, J. (1980). R1: an expert in the computer system domain, *Proc National Conference on Artificial Intelligence*, (pp.269-71).

Schmitt, S., Bergmann, R. (1999). Applying Case-Based Reasoning Technology for Product Selection and Customization in Electronic Commerce Environments, *Global Networked Organizations Twelfth International Bled Electronic Commerce Conference Bled*, Slovenia, June 7 - 9, 1999.

Turban, E. (1995). Decision Support and Expert Systems: Fourth Edition, *Prentice Hall International*, Inc., London.

Yang, Z., Deng, F., Liu, W., & Fang, Y. (2009). A CBR method for CFW prevention and treatment, Expert Systems with Applications, 36 (2009), pp.5469–5474.

Yin, Z., Li Y. (2010). Intelligent Decision Support System for Bridge Monitoring. In *Proceedings of the 2010 International Conference on Machine Vision and Human-machine Interface* (MVHI '10). IEEE Computer Society, Washington, DC, USA, 491-494. DOI= 10.1109/MVHI.2010.203 - http://dx.doi.org/10. 1109/ MVHI.2010.203.