

Fuzzy Intelligent Queries on the 500 Biggest Portuguese Companies

Claudia S. Castelhana, Susana R. Barros, Rita A. Ribeiro, Ana M. Moreira
Departamento de Informática,
Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa
2825 Monte Caparica, Portugal
Phone: +351-1-294 85 36
Fax: +351-1-294 85 41
Email: {rr|amm}@di.fct.unl.pt

ABSTRACT: This paper presents a fuzzy querying system to perform intelligent queries on the 500 biggest non-financial Portuguese companies' database. Queries are performed in fuzzy natural language and the system uses fuzzy reasoning to provide answers. The fuzzy reasoning process uses concepts of the representation language PRUF, Zadeh (1978). The objective is to provide an intelligent user-friendly querying system capable of answering questions such as "is company X *very profitable*?" or "are most private companies dynamic?".

KEYWORDS: modifiers, quantifiers, fuzzy natural language.

1. INTRODUCTION

Every year, the biggest non-financial 500 companies are ranked by the Portuguese magazine Exame (1996). The companies are ordered by their net sales volume and the 500 selected are those with net sales above 27.4 million dollars, i.e. they are the 500 biggest non-financial Portuguese companies. The magazine also presents a wide range of other financial and economic indicators of those biggest firms.

The objective of this paper is to describe a user-friendly fuzzy querying system, developed to help users obtain intelligent information about data of the 500 biggest non-financial Portuguese companies. The querying system is based on fuzzy natural language concepts, Zadeh (1978) (1983) and accepts four types of questions: modification (type I), composition (type II), quantification (type III) and qualification (type IV). Each type of question has a pre-defined syntax structure and its main components are subject, verb and the fuzzy attribute. The attributes can be affected by modifiers or qualifiers who expand or concentrate the concept, as for example "*very profitable*".

The querying system described is based in a larger project, developed by the first two authors under supervision of the last two, to obtain the BS degree in Computer Science, Castelhana & Barros (1998). The full application was developed in Portuguese, but here a translated version is presented.

2. THE DATABASE

The data is given in a table composed of 48 different economic and financial attributes. The attributes range from net sales, sales growth, capital, and number of workers to other economic and financial indicators. Some of the attributes were grouped according to the main four market player perspectives, government, managers, stockholders and banks, as follows:

- A. *Economic contribution*, i.e. contribution to the national economy. The indicators used are: a.1) Gross added value — sum of the net sales, production fluctuations, subsidies and net extraordinary profits; a.2)

Gross_added_value/net_sales measures how much a company contributes to the national economy per escudo (Portuguese currency) sold.

- B. *Dynamism*, i.e. firms dynamism and efficiency. The attributes used are: b.1) Sales growth, given by the ratio sales_95/sales_94; b.2) Net profits growth (given by the ratio net_profits_95/net_profits_94) measures the dynamism and the capacity to maintain or increase the market quota; b.3) Assets turnover (given by net_sales/assets) represents the degree of efficiency of available resources; b.4) Productivity (given by the ratio gross_added_value/number_of_workers) measures the degree of efficiency of human resources.
- C. *Profitability*. The indicators used are: c.1) return on investment gives the profit per unit of capital invested in the company; c.2) return on equity (given by the ratio of net_profits/owners_equity) measures the profitability of the owners capital; c.3) Profit margin on sales is given by the ratio of profits_after_taxes/sales; c.4) Sales profitability is measured with current_profits/sales.
- D. *Financial equilibrium*, i.e. the financial health of companies. The attributes used are: d.1) Indebtedness (given by the ratio liabilities/net_assets) measures the capacity of the firm to contract loans (the bigger the worse); d.2) Solvency (given by the ratio owners_equity/liabilities) measures the long-term capacity to fulfil commitments; d.3) Financial autonomy (given by the ratio owners_equity/net_assets) measures the participation of the owners equity in financing of the company activities (complement of Indebtedness); d.4) General liquidity (given by the ratio assets/current_liabilities) measures the capacity to fulfil the short-term commitments; d.5) Cash flow measures the auto-financing capacity of the company.

The reason for this grouping is linked, as mentioned, with obtaining information on the different perspectives of managers, bankers, stockholders and government, as well as, with the idea of offering a general view, to new stock market investors, of the best Portuguese firms. Obviously, the user can make either an individual query on each attribute or on a group.

It should also be noted that there were huge financial and economical differences between public and private companies, hence we opted to create two different tables (public and private companies) to be queried separately. It would not make much sense to aggregate data on different scales.

The attributes were fuzzified using an open interval trapezoidal function because, when we plotted the values contained in the table, it seemed the most appropriate type of function. Figure 1 shows the plotted values for gross_added_value/net sales, where the X-axis contains the ordering number of private companies and the Y-axis contains the actual indicator values.

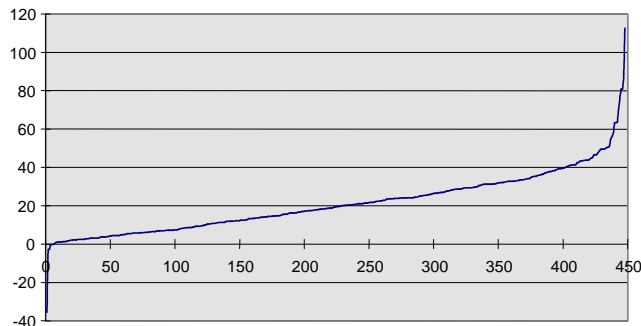


Figure 1. Gross added value / net sales

From the figure above, the fuzzy set corresponding to gross_added_value/net_sales is constructed as follows:

$$m(x) = \begin{cases} 0 & x \leq 36 \\ \frac{x - 36}{80 - 36} & 36 \leq x \leq 80 \\ 1 & x \geq 80 \end{cases}$$

All other fuzzy attributes were constructed in the same fashion.

3. THE QUERYING SYSTEM

The querying system uses an approximate reasoning process, based in the language PRUF, Zadeh (1978) to perform fuzzy queries on the database. PRUF (Possibilistic Relational Universal Fuzzy), is a representation language that allows the translation of natural language sentences, with their embedded imprecision, and the use of qualifiers or quantifiers such as *very*, *little*, *many* and so forth.

A natural language parser was constructed to define the syntactic structure of a proposition (query) and to analyze its semantic consistency. At the syntactic level four translation rules were used, Zadeh (1978):

Type I – Modification rule, "X is mF", that is:

<company_name> is <modifier><attribute>

Type II – Composition rule, "X is mF and/or Y is mF" or "if X is mF then Y is mF" or "X is {much, less} F than Y", that is :

<company_name_1> is <modifier><attribute>{and/or}<company_name_2> is <modifier> <attribute>

or

if <company_name_1> is <modifier><attribute> then <company_name_2> is <modifier> <fuzzy_set>

or

<company_name_1> is {much/less} <attribute> then <company_name_2>

Type III – Quantification rule, "QX are mF", that is:

<quantifier> <company_name> are <modifier><attribute>

Type IV – Qualification rule, "X is mF is {true, possible, probable}", that is:

<company_name> is <qualifier><attribute> is {true/possible/probable}

At the semantic level the parser validates the proposition grammar to detect invalid propositions, as for example using a single subject with a plural verb or an incoherent question like "if X is *dynamic* then Y is *profitable*, is true?" (The incoherence is due to the fact that both indicators should belong to the same group to be comparable).

As observed in the rules, the attributes can have modifiers, qualifiers or quantifiers attached to them. Modifiers and quantifiers either concentrate or enlarge the strength of an attribute. For example, the query "X is *very* productive" is more difficult to achieve a high membership value than the query "X is productive".

Modifiers, qualifiers and quantifiers are also represented as trapezoidal fuzzy sets in the system. The modifiers available are {*approximately*, *many*, *more or less*, *few*, *quite*, *very*, *low*, *high*}, the quantifiers are {*some*, *many*, *approximately-half*, *most*, *none*, *few*, *all*} and qualifiers are {*true*, *false*, *probable*, *improbable*, *possible* and *impossible*}.

Modifiers act as filters for attributes, i.e. considering a membership value of attribute F to be 0.8, if we say *very*F, the final membership value will be 0.5 because value 0.8 has membership value of 0.5 on the fuzzy set *very*. Quantifiers behave as a filtering process to the percentage of population on the universe that satisfies one or more attributes. For example the query "are many companies profitable?" will trigger a counting of the percentage of companies that are profitable in the database and that value is filtered through the fuzzy set *many* in the same way as for a modifier. Qualifiers apply the filtering process to the whole proposition in the same fashion.

Details about the translation rules can be found in Zadeh (1978) and details about the formulation of the modifiers, qualifiers and quantifiers can be found in Ribeiro (1993) (1995).

4. QUERIES AND ANSWERS

The query system interface is depicted in Figure 2. As can be observed, the interface is built to minimize the possibility of introducing errors because the user can select with a click modifiers, qualifiers, attributes, indicators, select a region or a company from private to public ones. Further, examples of each of the four types of query are available to help inexperienced users to formulate their queries.

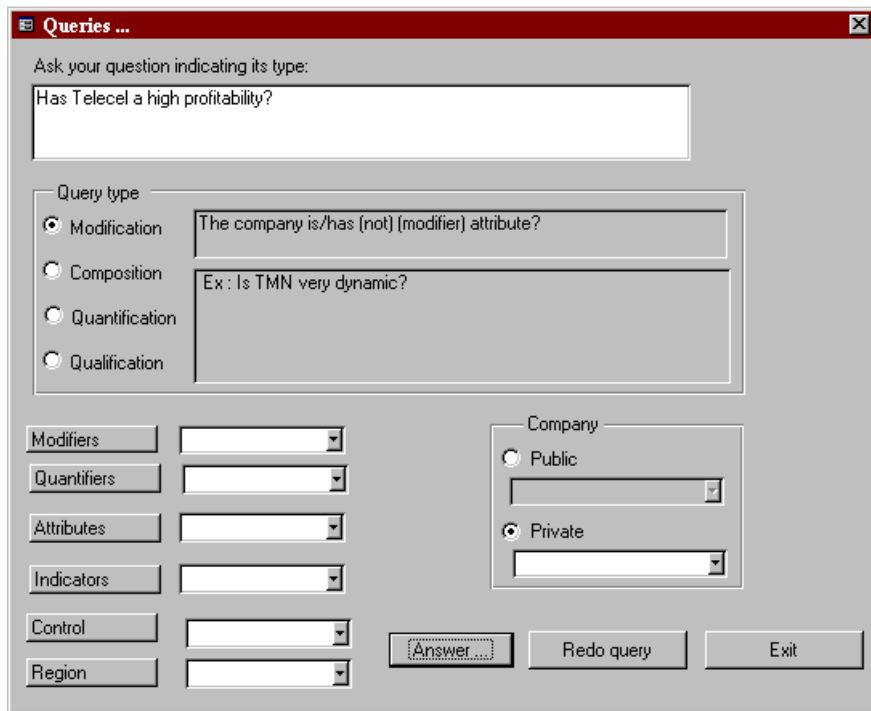


Figure 2. Intelligent queries interface

The answers provided by the system, which are a numerical measure of the formulated query, are shown in its numerical form and are also translated into a natural language evaluation. Examples of queries and answers, for each type of query are:

Type I – Modification

Question: "Is Anderson Consulting profitable?" (Without modifier)

Answer: The profitability is good (67%) because:

- the return on investment is good (67%)
- the return on equity is very big (86%)
- the profit margin sales is average (59%)
- the sales profits are average (57%).

Type II - Composition

Question: "Is Makro more dynamic then Maconde?"

Answer: It is quite true (75%) because:

- in sales growth is false (0%)
- in net profits growth is true (100%)
- in assets turnover is true (100%)
- in productivity is true (100%)

Type III – Quantification

Question: "Do some German companies have economic contribution?" (to Portugal's economy)

Answer: The economic contribution is small (25%) because:

- Gross added value is null (0%)
- Gross_added_value/net_sales is medium (52%).

Type IV – Qualification

Question: "Is it true that TMN has financial equilibrium?"

Answer: It is more or less true (40%) because:

- in the cash flow is not-much true (35%)
- in indebtedness is quite true (57%)

- c) in general liquidity (no data available!)
- d) in solvency is not-much true (25%)
- e) in financial autonomy is quite true (41%)

This small subset of queries is representative of the kind of questions the system can answer. Of course, many other queries using other modifiers, qualifiers and quantifiers can be asked to the system.

5. CONCLUSIONS

We presented a fuzzy intelligent query system capable of handling various types of questions in a natural language form. The query system allows questions on different market perspectives, such as managers, bankers, stockholders and government, as well as, a general overview about the main economic and financial data of the best 500 non-financial Portuguese firms.

We believe this type of intelligent interface can be very useful for companies that wish to provide a really user-friendly service to the community. Increasingly, interfaces to databases should be simplified to allow easy and fast access to the users.

The fuzzy reasoning used here to build intelligent interfaces is based in Zadeh's work on natural languages; it is well-suited for this approach. The four query-types used cover a large spectrum of possible queries to be made in the non-financial Portuguese company's database.

Since the queries are context-dependent, the query-types need to be adapted for other databases. However, the generality of the proposed approach indicates that the adaptation is not difficult.

REFERENCES

- Castelhano, C. S. e Barros, S. M. R. 1998. "Query's Inteligentes". Report of the final year project. Computer Science Department, faculty of Science and Technology, New University of Lisbon, Portugal.
- EXAME magazine, 1996. Special edition. Number 1, October/November.
- Ribeiro, R. A. 1993. "Application of Support Logic Theory to Fuzzy Multiple Attribute Decision Problems", PhD Thesis, University of Bristol, UK.
- Ribeiro, R. A. and J. F. Baldwin, 1995. "A Multiple Attribute Fuzzy Decision Support System: Two Applications". In: *Fuzzy Logic and Soft Computing*, Editors: B. Bouchon-Meunier, R. Yager and L. Zadeh, Advances in Fuzzy Systems-Applications and Theory Vol 4. World Scientific, pp 452-461.
- Zadeh, L. A. 1978. "PRUF-a Meaning Representation Language for Natural Languages". *International Journal of Man-Machine Studies*, Vol. 10, pp. 395-460.
- Zadeh, L. A. 1983. "A Computational Approach to Fuzzy Quantifiers in Natural Languages". *Computational Mathematics Applications*, Vol. 9, pp. 149-184.