Intelligent Query Model for Business Characteristics

Rita A. Ribeiro and 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

ABSTRACT: This paper presents an intelligent human-oriented interface based on fuzzy logic. The queries reasoning process uses four fuzzy translation rules of the meaning representation language PRUF, proposed by Zadeh. A database with the 500 biggest non-financial Portuguese companies is used to obtain specific queries of business characteristics. Illustrative examples of the four natural language queries are presented to show the capabilities of this human-oriented interface IMACS/IEEE CSCC'99 Proceedings, Pages: 5871-5876

Key-words fuzzy natural language, modifiers, quantifiers, qualifiers, financial indicators.

1. Introduction

An important issue in developing human-oriented interfaces has been to build querying models that allow non-expert users to query databases in a natural language. An intelligent database query model should be able to answer queries such as *is IBM a dynamic compan* \Re » or *«is Andersen Consulting very profitable* ». However, since these queries include imprecise concepts such as *«dynamic»* or *«very profitable»* we need tools to handle this imprecision.

Fuzzy set theory is a useful tool to handle imprecision [1] [2]. Specifically, the meaning representation language PRUF developed by Zadeh [3] provides an interesting framework to deal with imprecise statement $\{4\}$ [5] [6] [7]. The application of fuzzy set theory in the area of extending relational databases, to allow fuzzy queries, as been widely addressed in the literature (see for exampl [4] [8] [9] [5] [10] [11]). These studies reflect the importance of developing human-oriented querying models capable of handling imprecise statements.

The objective of this paper is to describe a human-oriented fuzzy querying model, developed to help users obtain intelligent information about the 500 biggest non-financial Portuguese companies. The querying model is based on the meaning representation language PRUF (Possibilistic Relational Universal Fuzzy) proposed by Zade [**B**] [7]. Zadeh represents fuzzy statements through four types of translation rule *§*3]: modification (type I), composition (type II), quantification (type III) and qualification (type IV). Our model uses these translation rules to obtain answers. A parser was also developed to create the syntactic structures, required by the four rules and to validate their semantic consistency.

The Portuguese magazine Exame [12] provided a table with the business characteristics of the 500 biggest non-financial companies ordered by their net sales volume. The data includes financial and economic indicators of companies with net sales above 27.4 million dollars. Further, we aggregated the indicators into four main categories -- dynamism, financial health, profitability and economic contribution -- to allow different market players such as managers, bankers, stockholders and government. Obviously, the user can perform queries about a single indicator or on a group.

Our model is based in a project developed as a requirement to obtain the BSc degree in Informatics [13]. The application was developed in Portuguese but here we present a translated version.

This paper is organized in four sections. Section one is this introduction. Section two discusses in detail the query model developed. Section three presents the illustrative examples. Section four provides the conclusions of this study.

2. Querying model

The querying model includes five main modules: consultation environment, the fuzzy engine, the database, and the interface. Figure 1 depicts the general architecture of the model.



Fig. 1. Query Model

2.1. Consultation environment

The consultation environment module is composed by a knowledge base and the fuzzy natural language elements.

2.2.1. Knowledge base

The knowledge base is concerned with the attributes, the fuzzy sets, the domains of the fuzzy sets, and relations.

Attributes. The attributes used in this model were extracted from the table with the 500 non-financial Portuguese companies [, 1996 #807]. The retrieved fifteen are: Sales growth, Net profits growth, Assets turnover, Productivity, Return on investment, Return on equity, Profit margins, Sales profitability, Gross added value. Gross added value/net sales. Indebtedness, Solvency, Financial autonomy, General liquidity, Cash flow.

Fuzzy sets. They fuzzy sets of our model correspond to the fifteen economic and financial attributes described above. We used open interval triangular functions to define the fuzzy attributes, because, when we plotted the values for each attribute (increasingly ordered) it seemed a simple and appropriate type of function.

Domain of the attributes. These were retrieved from the database, using again the plotted values. The points retrieved for each attribute were the minimum value, the inflection point value and the maximum value. These points represent the domain of any fuzzified attribute.

Relations. There are four different types of relations in our model, dynamism, profitability, economiccontribution and financial-health. Each relation reflects the perspectives of managers, stockholders, government and banks. Managers are mainly interested in the companies dynamism and productivity. Stockholders are mainly interested in the profitability of companies. The government is intererested in the economic contribution of companies to the national economy. Finally, banks are interested in the financial health of companies.

The four relations and their composing attributes are: Dynamism (sales growth, net profits growth, assets turnover, productivity); Profitability (return on investment, return on equity, profit margins, sales profitability); Economic-contribution (gross added value, gross added value/net sales); Financial-health (indebtedness, solvency, cash flow, financial autonomy, general liquidity).

The membership value for any semantic relations is obtained by:

$$\mu_{relationX}(Company_i) = \left(\frac{1}{n} \sum_{j=1}^{n} attribute_{ij}\right)$$

which is an average of all associated attributes of a relation. Many other aggregation operators could have been used to calculate the relations membership values (for an overview on operators, see [14]). It should be pointed out that, in this paper, a relation is viewed as an aggregation of its composing elements and this is the reason for using arithmetic operators.

2.1.2. Fuzzy natural language elements.

The fuzzy natural language module is composed of a temporary knowledge base which contains the definitions required by the fuzzy translation rules:

modifiers, quantifiers, qualifiers and rules. The modifiers, quantifiers and qualifiers were constructed with triangular and/or trapezoidal functions because they act as «filters» for the attributes or propositions [15] [8].

Modifiers. They are adverbs that modify the fuzzy attribute, such as *not*, *much*, *very*. The modifiers available in were constructed with triangular and trapezoidal functions. Our model also accepts the modifier *not*, which is represented by $(1-\mu_M(F))$.

Quantifiers. These are linguistic expressions that limit the number of cases to be queried, such as *all*, *most*, *aproximately-half*. Like the modifiers, quantifiers were defined as triangular and trapezoidal functions.

Qualifiers. These are adverbs that linguistically qualify a proposition to determine its degree of truth, probability or possibility. For example, the queries «is it *very true* that IBM is *productive*?» or «what is the *possibility* that IBM is *profitable*?» clearly show that some measure of qualification of the proposition is being asked. In our model, instead of the linguistic hedges proposed by Zadeh [19] we used a simple real line to express the degree of truth, i.e. the membership value of the sentence is considered as the value of truth.

2. 2. Fuzzy engine

This module is composed of three submodules: a parser, the PRUF fuzzy logic and a de-fuzzification submodule.

2.2.1. Parser.

A natural language parser was built to translate a query into the syntactic structure accepted in our model and then to validate its semantic consistency. A first validation is performed by checking the existence in the database of company names, modifiers, qualifiers, quantifiers and attributes. 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*?" (The incoherence is due to the fact that either subject is the same and the attributes are different, or the subjects are different and the attribute is the same, in order to being comparable). Another semantic evaluation on type II rules is to

check if the connector is *and* because then it requires that the subject be the same.

Thus, the parser interprets the structure of the sentence in order to recognize the type of the question to be handled. For each type of question we have designed a grammar. A grammar defines the set of rules that can be used to build up a sentence. Due to space restrictions no further details are given.

2.2.2. PRUF Rules.

The rules of our model are based, as mentioned, in PRUF (Possibilistic Relational Universal Fuzzy) [3], which is a representation language that allows the translation of natural language sentences, with their embedded imprecision, and the use of modifiers or quantifiers like *very*, *few* and *many*. PRUF proposes four translation rules, modification (type I), composition (type II), quantification (type III), qualification (type IV).

Type I: Modification («X is m F»)

This rule expresses that a simple fuzzy proposition P, P = X is mF, where X is the subject, F is the fuzzy set and *m* is a modifier, can be expressed by a possibility distribution [14], corresponding to the modified rule translation equation,

 $\mathbf{P}^{+} = \boldsymbol{\mu}_{\mathrm{m}}(\boldsymbol{\mu}_{\mathrm{F}}(\mathbf{x}))$

where μ_m yields the membership value of $\mu_F(x)$ in a modifier function *m*.

An illustrative example can be «Is IBM *very productive*?», where F is the attribute *productive* and *very* is the modifier.

The syntactic structure of the rule for our model is:

<company_name> is <modifier><attribute>

Type II: Composition

This rule comprises three different types of compositions:

- 1) conjunctive/disjunctive composition «X is F_1 and/or Y is F_2 »
- comparative composition «X is more/less F than Y»,

3) conditional composition «If X is F_1 then Y is F_2 ». **Rule 1).** It corresponds to two modification propositions «X is m_1F_1 » and «Y is m_2F_2 », connected by operator *and* or *or*. The operators used are *min* and *max* for *and* and *or* respectively. Thus,

X is m_1F_1 and Y is $m_2F_2 =>$

 $P^{+}=\min (\mu_{m1}(\mu_{F1}(x)), \mu_{m2}(\mu_{F2}(y)))$

X is m_1F_1 or Y is $m_2F_2 =>$

 $P^{+}= \max (\mu_{m1}(\mu_{F1}(x)), \mu_{m2}(\mu_{F2}(y)))$

This rule accepts more than two connected propositions, X and Y can be same subject, and m_1 and m_2 are optional. An example is «Is IBM profitable and Bayer dynamic?».

Rule 2). It has two propositions, «X is m_1F_1 » and «Y is m_2F_2 », but the connector is a comparison between them, to measure if X is *more* or *less* than Y. Thus,

X is more mF then Y =>

if
$$\mu_{m1}(\mu_{F1}(x)) > \mu_{m2}(\mu_{F2}(y)) = 1$$
, else =0

X is less mF then Y =>

if
$$\mu_{m1}(\mu_{F1}(x)) < \mu_{m2}(\mu_{F2}(y)) = 1$$
, else =0

As above, m is optional. An example could be, «Is IBM less profitable than Bayer?».

Rule 3. Not used here.

The accepted syntactic structures for Type II rules, contained in our model are:

- a) <company_name> is <modifier> <attribute> {and/or} <company_name_2> is <modifier> <attribute>,
- b) <*company_name> is* {*more/less*} <*attribute> then* <*company_name_2>*.

Type III: Quantification («qX are mF»)

This rule corresponds to the proposition, P = qX are F, where X is a group of subjects, F is a fuzzy attribute and q is a quantifier. For this rule we need the notion of relative cardinality, **rcard**, of a fuzzy set [14]. Thus, the translation rule equation is:

 $P^+ = \mu_q(\mathbf{rcard}_F(\mathbf{x}))$

where $\mathbf{rcard}_{F}(x)$ is the averaged proportion of elements x in F and μ_{q} is the membership value of **rcard** in the fuzzy quantifier function.

The syntactic structure accepted by our model is:

<quantifier> <companies/region/control> are <modifier><attribute>

We should point out that other interesting approches to linguistically quantified propositions have been proposed in the literature, such as [9] [16]. However, since in this paper we follow Zadeh's proposal they will not be discussed further.

Type IV: Qualification («X is mF is true/false») This rule measures the truthfulness or falsity of a modified proposition. Thus the translation rule is:

 $P^+ = \mu_{truth}(\mu_m(\mu_F(x)))$

where μ_{truth} is the membership value in the truth function (or false function) of the modified rule. An example could be «Is it true that IBM is dynamic?».

The syntactic structure accepted in our model for this type of rule is:

<company_name> is <modifier><attribute> is {true/probable}

In summary, modifiers act as filters for attributes. For example, 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 modifier 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.

2.2.3. De-fuzzification.

The results obtained from the system have two distinct forms, quantitative and qualitative. First, as mentioned, it uses the translation rules calculus, presented in Section 2.2.2, to obtain a quantified result for the query. The quantified answer is then matched with a linguistic value corresponding to the quantitative evaluation. For example, if the value for a modification rule lies in the interval [0.021 0.2] the qualitative answer will be "very small".

We define seven intervals empirically. Many other intervals could have been determined, but the seven seem to cover rather well the nature of the results obtained from our database.

2.3. The database

The analysis of the information needed to build our database was modeled using an entity-relationship diagram, which is a subset of UML [17]. For reasons of space and considering it is a simple database we will not present more details here.

2.4. Interface

The human-oriented interface is built to minimize the possibility of introducing errors since the user can select with a click modifiers, qualifiers, attributes, indicators, select a region or a company from private to public ones. Examples of each of the four types of query are available to help non-expert users formulate their queries. The query system interface is depicted in Figure 2.

🖻 Queries		×
Ask your question in	dicating its type:	
Has Telecel a high p	rofitability?	
Query type		
Modification	The company is/has (not) (modifier) attribute?	
C Composition) Ex:ls TMN verv dvnamic?	
C Quantification		
C Qualification		
Modifiers	Company —	
Quantifiers		
Attributes	Private	
		-
Control	Answer Redo quer	ry Exit
Region	<u> </u>	

Fig. 2. Intelligent queries interface

All the pull-down menus contain a list of existing modifiers, quantifiers, attributes, indicators, control (countries), regions, public companies and private companies. The user does not need to know which modifier, quantifier, attributes and so forth exist in the system. He/she can formulate the query by just selecting from those menus. Further, there is an help box (the one that says «ex: is TMN very dynamic») that after selecting the desired type of query depicts an example of how the query should be made. In summary, our interface is clearly intended for expert and non-expert users.

In the next section we present various types of queries and results obtained to show the capabilities of our model.

3. Illustrative examples

Modification query: Has Dan Cake high return on equity?

Answer: high return on equity is high (85%) because, a) value and membership 18.6/0.82

b) Filtering attribute with modifier, $\mu_{high}(0.82)=0.85$

Composition query: Has Dan Cake more return on equity than Bayer?

Answer: the condition is false(0%) because,

- a) value and membership: 11.7/0.8225
- b) value and membership: 18.6/0.8678

c) Comparative composition: If 0.8225 > 0.8678 μ =1 else μ =0

Quantification query: Most companies on central Portugal have sales_profitability?

Answer: the set has very high sales_profitability (100%) because,

a) rcard_{sales_profit}=0.90 (rel. cardinality of all companies with sales_profit.)

```
b) \mu_{most}(0.90)=1
```

Qualification query: Is it true that IBM is productive?

Answer: the question has low truth (21%) because:

a) value and membership: 11.3/0.21

b) $\mu_{true}(0.21) = 0.21$

This small set of questions illustrates the behavior of our querying model, both at the query and answer levels. It clearly displays how a humanoriented query model can help a non-expert user to ask natural language questions and obtain not only raw data, but also real information about the companies.

4. Conclusions

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

It should also be pointed out that since the syntactic structure of queries is context-dependent the query model needs to be adapted for use with other databases. However, the generality of the proposed approach indicates that the adaptation is not difficult.

We believe this type of human oriented interfaces can be very useful for companies that wish to provide a really user-friendly service to the community. More human-oriented query models should be developed to improve their capabilities and allow easy and fast access to non-expert users.

References:

[1] Bosc, P. and J. Kacprzyk, eds. "*Fuzziness in Database Management Systems*". Studies in Fuzziness, Physica- Verlag, 1995, .

[2] Zadeh, L.A., The Role of Fuzzy Logic in the Management of Uncertainly in Expert Systems. *Fuzzy Sets and Systems*, **11**, 1983, p. 199-228.

[3] Zadeh, L.A., PRUF-a Meaning Representation Language for Natural Langages. *Intern. J. of Man-Machine Studies*, **10**, 1978, p. 395-460.

[4] Kacprzyk, J. and S. Zadrozny, "FQuery for Access: Fuzzy Querying for Windows-Based DBMS", in *Fuzziness in Database Management Systems*, P. Bosc and J. Kacprzyk, Editors, , Physica-Verlag. 1995.

[5] Takahashi, Y., "A Fuzzy Query Language for Relational Databases", in *Fuzziness in Database Management Systems*, P. Bosc and J. Kacprzyk, Editors, Physica-Verlag. 1995.

[6] Takahashi, Y., A Fuzzy Query Language for Relational Databases. *IEEE Transactions on Systems, Man and Cybernetics*, **21**(6), 1991, p. 1576-1579. 7. Zadeh, L.A., A Computational Approach to Fuzzy Quantifiers in Natural Languages. *Computational Mathematics Applications*, **9**, 1983, p. 149-184.

[8] Kacprzyk, J. and A. Ziolkowsi, Database Queries with Fuzzy Linguistic Quantifiers. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-16**(3), 1986, p. 474-479.

[9] Kacprzyk, J. and R. Yager, Linguistic Quantifiers and Belief Qualification in Fuzzy Multicriteria and Multistage Decision Making. *Control and Cybernetics*, . **13**(3), 1984, p. 154-173.

[10] Buckles, B.P. and F.E. Petry, "Query Languages for Fuzzy Databases", in *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory*, J. Kacprzyk and R. Yager, Editors, Verlag, TUV Rheiland. 1985, p. 241-251.

[11] Baldwin, J.F., P. Baldwin, and S. Brown, "A Natural language Interface for FRIL", in *Management Decision Support Systems using Fuzzy Sets and Possibility Theory*, J. Kacprzyk and R. Yager, Editors, , Verlag TUV. 1985. p. 270-279.

[12] EXAME, As Melhores e Maiores. *EXAME* magazine, . 1(Oct/Nov), 1996, .

[13] Castelhano, C.I.S. and S.R. Barros, "Query's Inteligentes", Project report for BSc degree, Computer Science Department, Faculty of Sciences and Technology, New University of Lisbon, Portugal. 1998, .

[14] Klir, G.J. and T.A. Folger, "*Fuzzy Sets, Uncertainty, and Information*": Prentice-Hall International Editions, 1988.

[15] Ribeiro, R.A., "Application of Support Logic Theory to Fuzzy Multiple Attribute Decision Problems", University of Bristol, UK: PhD Thesis. 1993, .

[16] Yager, R.R., Quantifiers in The Formulation of Multiple Objective Decision Functions. *Information Sciences*, **31**, 1983, p. 107-139.

[17] Rational, , www.rational.com/UML. 1997, .