THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2003 ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

VAGUE CRITERIA IN RELATIONAL DATABASE QUERIES

Cornelia TUDORIE

Department of Computer Science, University "Dunarea de Jos", Galati
Domneasca 111, 800008 Galati, Tel, Fax: 460182
email: Cornelia.Tudorie@ugal.ro

Abstract: The most accessible interfaces querying relational database should allow vague terms included in selection criteria. The gradual property stands for the simplest vague selection criterion, but more categories of vague terms can be identified in a natural expression of a query. All of them can be modelled using fuzzy sets theory. A complete general syntax of a vague criterion of database selection will be concluded. Appropriate transformation or operations are defined in order to compute the degree of satisfaction for the vague criteria containing fuzzy terms.

Keywords: Artificial Intelligence, Database, Flexible Queries, Fuzzy Logic

1. INTRODUCTION

The access to database is possible following one of two ways:

- operating with application programs, when a limited set of predetermined functions are available and
- operating directly on data, using relational command languages.

The second one is unavoidable when an occasional operation, in particular terms, is performing. Two major limitations occur in such a database querying access: the rigid artificial language syntax and the difficulty to realize and express precise criteria to locate the information. This is because the humans think and speak not always in precise terms. So, it is very usefully to provide intelligent interfaces to databases, able to understand natural language queries and more important, able to interpret and evaluate imprecise criteria of queries.

We focus on the possible vagueness of the selection criterion, which involve certain vague terms, currently used in natural language speaking.

The fuzzy sets theory will be used, as the adequate framework to model and to manage vague expressions, or in other words, to evaluate vague queries sent to relational database.

Including vague criteria in a database query may have two advantages:

- the flexibility of the query expression
- the possibility to refine the result, assigning to each tuple the corresponding degree to respect the criteria

A query for example may be:

*Retrieve the **old** students in the department*

and a possible result set:

Name	Age		degree
John	33	.....	1
Paul	31	.....	1
Marie	25	.....	1
Joseph	24	.....	0.66
Isabelle	23	.....	0.33

An ordered list of students by the satisfaction degree is performed, so we can say that Paul is certainly old and the same time he is older than Isabelle is.

This paper was recommended for publication by Severin BUMBARU

The present paper tries to identify the most important categories of vague terms[1], to find examples in Romanian vocabulary, to find the most appropriate models for some of them and also for their aggregation operators, using fuzzy sets theory.
Each section of the paper discusses one of the vague terms categories; finally, a complete general syntax of a vague criterion of database selection is concluded.

2. VAGUE SELECTION CRITERIA

Querying relational databases in classical systems implies a set of algebraic operations, applied to component tables. The essential operation for database consulting is the selection. Selecting data (table rows) from database follows a selection criterion, expressed by a Boolean formula. The query result consists in a structure like a table, containing such database rows for what the formula take the *true* value.

STUDENT

Nume	**. . .**	**Nota**	**Vârsta**	**. . .**
Elena		7	20	
Ioana		6	23	
Maria		8	21	
Paul		9	26	
Vasile		4	22	
Costel		8	24	
Ion		10	20	

Fig.1. A relational table.

For example, the following crisp query (precise, or Boolean) is sent to a database including the table in figure 1

Care sunt studenţii care au nota mai mare decât 8?
(Retrieve the students having mark greater than 8)

The Boolean criterion

nota> 8 (*mark > 8*)

is evaluated and the result is

Nume	**. . .**	**Nota**	**Vârsta**	**. . .**
Maria		8	21	
Costel		8	24	

But humans think and speak not always in precise terms. So, very often the database user needs to ask for information, indicating a vague expression as selection criterion, for example

*Care sunt studenţii care au notă **bună** ?*
(Retrieve *the students having **good** mark)*

This time the criterion *notă **bună*** (***good*** *mark*) is no more Boolean, so the selection must be more refined: for each table row a satisfaction degree is computed, which stands for a measure of its compatibility with the vague criterion.
In conclusion, the **vague criterion** in a database context is an expression including vague terms, standing for rows selection condition in a database query.

The vague criteria may be very simple (like the above one), but it may be also very complex. Firstly, we consider the linguistic complexity, but not the logical one. The linguistic complexity of the criterion is coming from various categories of vague terms with different semantic effects on the selection criterion, that is on the logical complexity.
In the following, a review of several categories of linguistically terms with vague meaning, their fuzzy model and specific operations are presented.

3. GRADUAL PROPERTY

The gradual property stands for the simplest vague selection criterion to a relational database. As the name shows, it can be more or less satisfied by the database objects. The gradual property comes from natural linguistic expressions of an object property, related to a database attribute, and its fuzzy model consists in a fuzzy predicate.

The **fuzzy predicate** is an affirmation that may be more or less true, depending on the variables values. It is an extension of the classical logical predicate, that can be either definite true, or definite false. The truth-value of the vague predicate may be express as a number in [0,1], with standing 1 for definitely true and 0 for definitely false.

Usually, the fuzzy predicate is defined by a fuzzy set or a fuzzy number (figure 2).

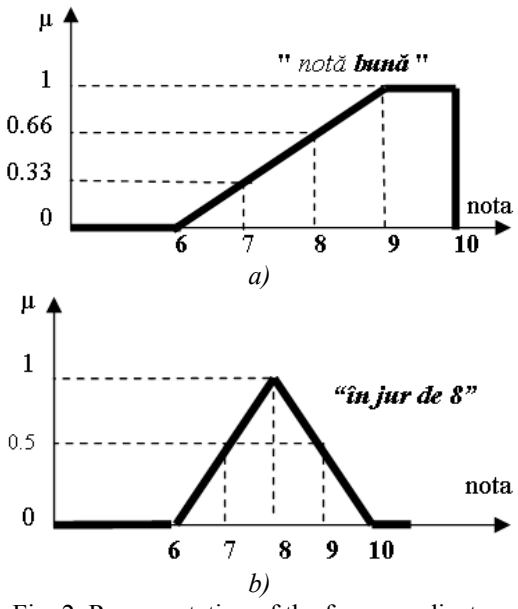


Fig. 2. Representation of the fuzzy predicates:
a) ***good*** mark b) mark ***around 8***

[1] The study is made Romanian language; some examples are translated, but not all.

In a context related to query selection criterion, the fuzzy predicate may be linguistically expressed by a value of a linguistic variable (*linguistic value*), or an *approximate value*.

According to that definitions, the result of the vague query

*Care sunt studenții care au notă **bună** ?*
(*Retrieve the students having **good** mark*)

on the STUDENT table is:

Nume	Nota	μ
Ion	10	1
Paul	9	1
Maria	8	0.66
Costel	8	0.66
Elena	7	0.33

and the result of the vague query

*Care sunt studenții care au nota **în jur de 8** ?*
(*Retrieve the students having mark **around 8***)

Nume	Nota	μ
Maria	8	1
Costel	8	1
Paul	9	0.5
Elena	7	0.5

is:

The satisfaction degree for each tuple is derived from the definition of the fuzzy predicate. This is equal to the value of the membership function corresponding to the attribute value in the current tuple.
For example, if the fuzzy predicate is defined by a fuzzy set F , the satisfaction degree of the criterion for the tuple i is:

$$\mu_i = \mu_F (t_i)$$

where μ_F is the fuzzy set membership function, and t_i is the attribute crisp value for the tuple i .

As shown below, each fuzzy predicate as gradual property, is properly defined on a database attribute.

The set of linguistic forms of fuzzy predicates regarding a database attribute is named **linguistic domain**. In order to evaluate any vague query to a database, it is necessary to define both the crisp domain and the linguistic one, for each usual searching attribute.
For example, [0,10] is the crisp domain, and
{ ***slabă, medie, bună*** } ({ ***bad, middle, good*** }) is the linguistic domain for the
Nota attribute of the STUDENT table.

Other vague terms coming from natural language may be included in the database query and must be taking into account in query evaluation. Some of them are presented in the follow.

4. LINGUISTIC MODIFIERS

Staying close to a gradual property, the **linguistic modifier** affects the intensity of this one. It is generally defined by a transformation function

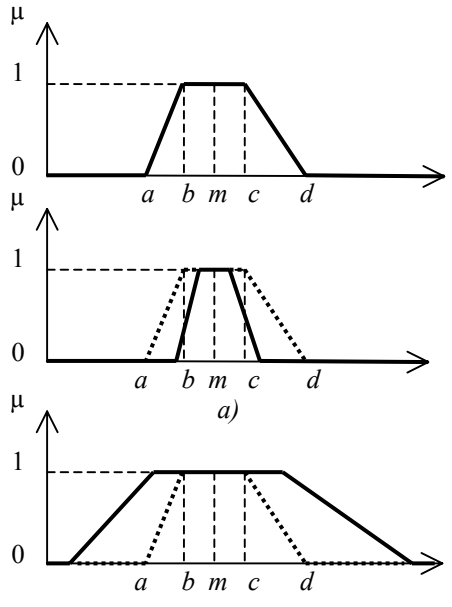
$$T: \mathbb{F}(U) \rightarrow \mathbb{F}(U)$$

that modifies the membership function of the fuzzy predicate to an other one.

There are some categories for the linguistic modifiers, with different semantic effect:

- **concentration and dilatation**

The concentration modifiers provide a new membership function, more limited than the original one. In other words, the predicate becomes more restrictive, the searching criterion more drastic, and the response more selective.
The dilatation modifiers provide a new membership function, larger than the original one. In other words, the predicate becomes more relaxed, the searching criterion more generous, and the response more comprehensive.
The effect of the concentration (a) and dilatation (b) modifiers on the membership function is exemplified in figure 3.



b)

Fig. 3. The effect of the a) concentration modifier
b) dilatation modifier

A set of concentration and dilatation modifiers and definition expressions are proposed in (Marín-Blázquez, Shen, 2001): **MORE**, **VERY**, **EXREMELY** and **LESS**, **GREATLY**, **REMOTELY**.
For example two modifiers, coming from Romanian language, are ***foarte*** and ***relativ*** (fig 4). According these ones, the response for the query:

*Care sunt studenţii care au notă **foarte bună** ?*
(*Retrieve the students having **very good** mark)*

Nume	Nota	μ
Ion	10	1
Paul	9	0.66

is

and for the query

*Care sunt studenţii care au notă **relativ bună** ?*
(*Retrieve the students having **somehow good** mark)*

the response is:

Nume	Nota	μ
Ion	10	1
Paul	9	1
Maria	8	0.86
Costel	8	0.86
Elena	7	0.57
Ioana	6	0.28

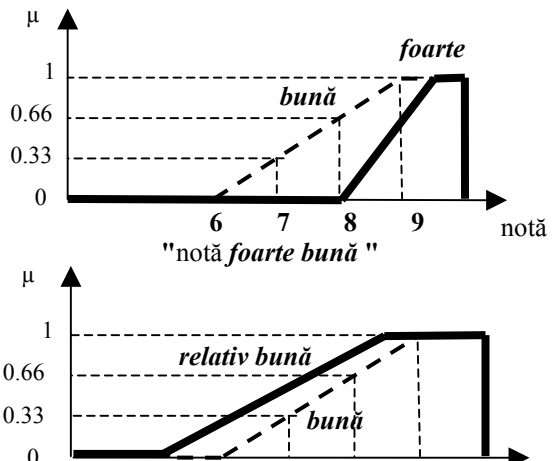


Fig. 4. The membership function for two modified fuzzy predicates

- **detailisation**

The detailisation modifiers take over only a part of the original membership function. It may be express by terms such **LOWER**, **MID**, **UPPER** (fig. 5)

For example, the response for the query:

*Care sunt studenţii care au notă **ceva mai bună** ?*
is

Nume	Nota	μ
Maria	8	1
Costel	8	1
Paul	9	0.5
Elena	7	0.5

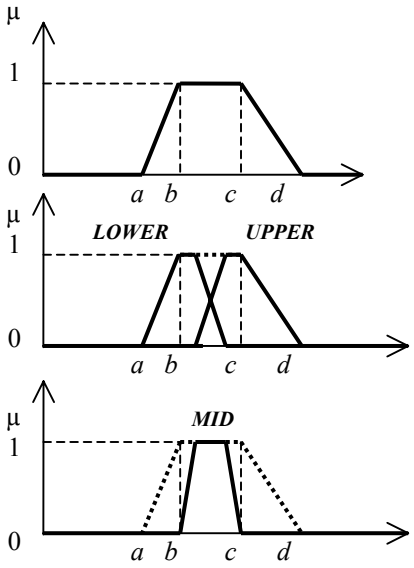
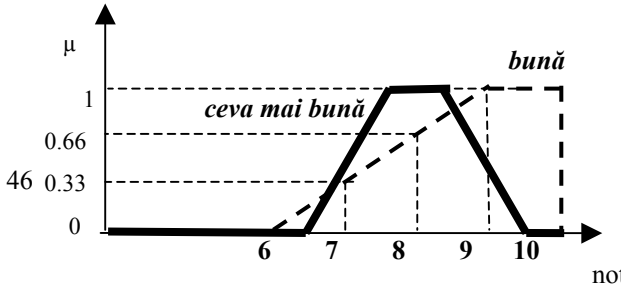


Fig. 5. Membership function modified by detailisation modifiers



"notă ***ceva mai bună*** "

Fig. 6. Membership function of a modified fuzzy predicate

- **covering**

The predicate affected by covering modifiers is defined by a modified fuzzy set, which is a superset for the original one. Depending on the expression of the modifier, the modified fuzzy set may be a *left covering* or a *right covering* for the original fuzzy set. For the Romanian vocabulary, the typical linguistic covering modifiers are ***cel mult*** and ***cel puțin***.
For example, the evaluation of the query:

*Care sunt studenții care au notă **cel puțin medie** ?*

may take in account the definitions in figure 7 and the modified membership function in figure 8.

Fig. 7. Linguistic domain for the **Nota** attribute

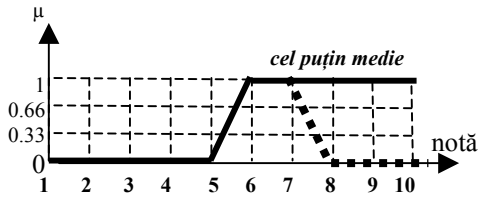


Fig. 8. The membership function for a modified linguistic value

Using linguistic modifiers is a possibility to give a more refined expression to a property, or an intermediate value between the linguistic variable's ones.

5. LINGUISTIC COMPARATORS

The linguistic operators are modeled as fuzzy relations in the fuzzy set theory; thus the result of comparing two vague values is a [0,1] coefficient, indicating the satisfaction degree of the relation between the two operands.

The fuzzy operators are expressed by linguistic terms:

- ***equal***, applicable to two linguistic values (an example for graphical computed result is shown in figure 9)

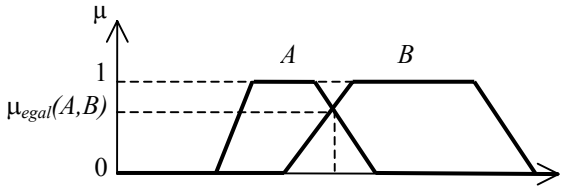


Fig. 9. The truth degree for two fuzzy sets equality

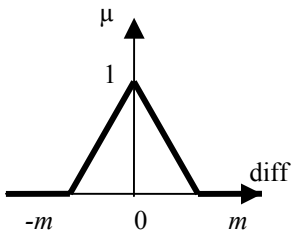


Fig. 10. The definition of the ***approximately equal*** comparator

- ***approximately equal,*** applicable to two crisp values (an example for a definition of the comparator, based on the difference between the two crisp values, is shown in figure 10)

- ***less or equal, greater or equal,*** applicable to two linguistic values or one linguistic value and other one crisp (two examples for graphical computed results are shown in figure 11)

- ***less, greater***, applicable to two linguistic values, and defined as negation of the ***greater or equal*** and respectively ***less or equal*** comparators.

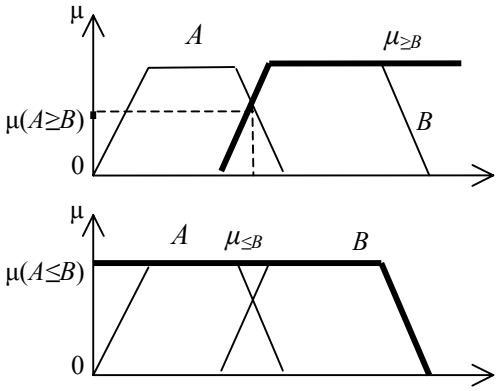


Fig. 11. The truth degree for two fuzzy sets comparation

In the most situations, the last four linguistic operators have the same definition like the covering modifiers.
For example, the evaluation of the query:

Nume	Vârsta	Nota	$\mu_{\widetilde{23}}$	μ_{medie}	$\mu_{bună}$	μ
Costel	24	8	0.8	0	0.66	0.66
Maria	21	8	0.6	0	0.66	0.6
Ion	20	10	0.4	0	1	0.4
Paul	26	9	0.4	0	1	0.4
Elena	20	7	0.4	1	0.33	0.4

*Care sunt studenții care au notă **mai puțin decât bună** ?*

may take in account the definitions in figure 7 and the modified membership function in figure 12.

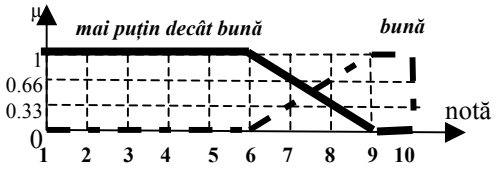


Fig. 12. The comparison between a fuzzy set and a crisp value

6. AGGREGATING VAGUE CRITERIA

In a precise query, the compound criterion is a logical expression containing comparisons and logical operators. In a vague context, the classical logical operators **AND**, **OR**, **NOT**, are extended to fuzzy aggregation connectives. They are able to

- combine different gradual properties if the required conditions are referring some object, or
- follow certain rules to combine satisfaction degrees of different vague criteria.

There are many propositions in the literature for defining aggregation connectives (Yager, 1991).

Usually, the minimum and maximum functions stand for the conjunctive and disjunctive connectives; the complement stands for the negation connective.

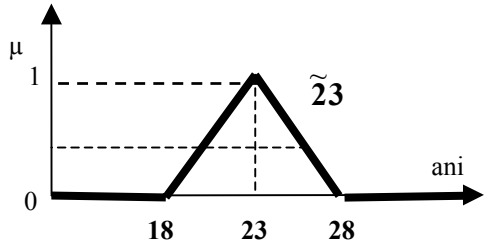


Fig. 13. The definition of a fuzzy predicate by a fuzzy number

For example, if the table in figure 1 and the definitions in figures 7 and 13 are considered, the response to the query

*Care sunt studenții **în jur de 23 ani** care au notă **medie sau bună** ?*

is

7. CONCLUSIONS

The vague criterion including in a database query is defined in the paper. Some categories of elements composing vague selection criteria in database queries are review. The vague character of these elements is coming from the natural expression used in usual human communication and reasoning too. The fuzzy set theory provides the most appropriate framework to model all these vague linguistic terms and to evaluate vague queries.
As a synthesis, a formal grammar for any vague criterion form, is proposed:

CCSV **::=** CSV |
CCSV < *logical precise operator* >CCSV |
CCSV <*vague aggregate operator*> CCSV

CSV **::=** <*attribute*> <*linguistic comparator*> <*crisp value*> |
<*attribute*> <*precise comparator*> <*gradual property*> |
<*attribute*> <*linguistic comparator*> <*gradual property*>

<*gradual property*>**::=** < *linguistic value*> |
< *linguistic modifier*>< *linguistic value*> |
<*approximate value*>

The most difficult task is to find the most adequate model to each linguistic term used in vague query expression. The particularity of such a definitions set is because each national language has a specific vocabulary richer or less rich, including many phrases untranslatable in others.
One time existing a fuzzy knowledge base containing all vague terms definitions, an intelligent system able to interpret and to evaluate vague queries to relational database can be developed.

8. REFERENCES

Bosc, P. and H. Prade (1997). An Introduction to Fuzzy Set and Possibility Theory-based Approaches to the Treatment of Uncertainty and Imprecision in Data Base Management Systems. In *Uncertainty Management in Information Systems: from Needs to Solutions* (A. Motro and P. Smets, (Eds)), pp. 285-324, Kluwer Academic Publishers.

Marín-Blázquez, J.G., Shen, Q. (2001). Linguistic Hedges on Trapezoidal Fuzzy Sets: A Revisit.

Tudorie, C., Segal, C., Bumbaru, S. (2001) Imperfect data in database context. Flexible queries, In *SIMSIS-11 Proceedings*, pp. 189-193, Galati

Tudorie, C. (2003). Cercetări privind aplicarea tehnicilor de inteligență artificială pentru interogarea bazelor de date. Scientific Rapport, University 'Dunărea de Jos', Galați

Yager, R.R. (1991). Connectives and quantifiers in fuzzy sets, In *Fuzzy Sets and Systems*, **40-1**, Elsevier Science, pp 39-75