STRICTNESS OF FOUR-FOLD TABLE GENERALIZED QUANTIFIERS IN FUZZY LOGIC

Jiří Ivánek
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Prague
e-mail: ivanek@utia.cas.cz

ABSTRACT: In the paper, generalized quantifiers defined on four-fold tables corresponding to pairs of attributes are investigated from the fuzzy logic point of view. Using the notion of strictness, the method of construction of the logically nearest Σ-double implicational and Σ-equivalence quantifiers to a given implicational quantifier is described.

Keywords: Fuzzy Logic, Observational Quantifiers, Data Mining

1 Introduction

The notion of generalized observational quantifiers was introduced in the framework of logical theory of the GUHA method of mechanized hypothesis formation [5], [6]. It should be stressed that this method is one of the earliest methods of data mining. The method was during years developed and various procedures were implemented e.g. in the systems PC-GUHA [7], Knowledge Explorer [3], and 4FT-Miner [15]. Further investigations of its mathematical and logical foundations can be found e.g. in [9], [14].

In the paper, several classes of the most widely used four-fold table quantifiers with truth values in the unit interval are investigated. Such type of quantifications of rules derived from databases is used in various methods of knowledge discovery in databases (see e.g. [16]). On the other hand, there is a connection between four-fold table quantifiers and measures of resemblance or similarity applied on Boolean vectors [2].

In Section 2, basic notions and classes of quantifiers are defined, and the notion of their strictness is introduced. In Section 3, this notion is used for description of the method which provides a logically strong one-to-one correspondence between classes of implicational and so called Σ-double implicational quantifiers. An analogical construction is used in Section 4 to introduce similar correspondence between classes of Σ-double implicational and Σ-equivalence quantifiers.

2 Four-fold table generalized quantifiers

Assume having a data file and consider two Boolean (binary, dichotomic) attributes φ and ψ. A four-fold table $< a,b,c,d >$ corresponding to these attributes is composed from numbers of objects in data satisfying four different Boolean combinations of attributes:

	ψ	$\neg\psi$
φ	a	b
$\neg\varphi$	c	d

a - satisfying φ and ψ,
b - satisfying φ and $\neg\psi$,
c - satisfying $\neg\varphi$ and ψ,
d - satisfying $\neg\varphi$ and $\neg\psi$.

To avoid degenerated situations, we shall assume, that all marginals of the four-fold table are non-zero:
$a+b>0$, $c+d>0$, $a+c>0$, $b+d>0$.

Various relations between φ and ψ can be measured in given data by different four-fold table generalized quantifiers $\sim(a,b,c,d)$ which will be understood here as functions with values in the interval $[0,1]$.

Definition 1
Four-fold table generalized quantifier $\sim$ *is a* $[0,1]$*-valued function defined for all four-fold tables* $< a,b,c,d >$.

We shall write $\sim(a,b)$ if the value of the quantifier $\sim$ depends only on a,b; $\sim(a,b,c)$ if the value of the quantifier $\sim$ depends only on a,b,c; $\sim(a,b,c,d)$ if the value of the quantifier $\sim$ depends on all a,b,c,d. For brevity, we shall call in this paper "four-fold table generalized quantifiers" simply "quantifiers".

The most general class of quantifiers originally introduced in two-valued logic in [5] and called there **associational** is reflecting the following property: If the four-fold table $< a,b,c,d >$ represents the behaviour of the derived attributes φ and ψ in given data, then numbers a,d are supporting correlation of φ and ψ but numbers b,c are against. This property can be formulated in fuzzy logic approach by: The higher is a,d and the smaller are b,c, the better or at least not worse is truth-value of association of φ and ψ in given data.

The most common examples of such assocional quantifiers are following ones:

Example 1
Quantifier $\Rightarrow_{\oslash}$ of basic implication (or confidence of association rules, see [1]):

$\Rightarrow_{\oslash}(a,b) = \frac{a}{a+b}$.

Quantifier $\Leftrightarrow_{\oslash}$ of basic double implication (Jaccard 1900, [2]):

$\Leftrightarrow_{\oslash}(a,b,c) = \frac{a}{a+b+c}$.

Quantifier $\equiv_{\oslash}$ of basic equivalence (Kendall, Sokal-Michener 1958, [2]):

$\equiv_{\oslash}(a,b,c,d) = \frac{a+d}{a+b+c+d}$.

Properties of basic quantifiers are in the core of definitions of several useful classes of quantifiers (introduced originally in two-valued logic in [5], [6]) which can be naturally given in fuzzy logic as follows:

Definition 2
Let a,b,c,d,a',b',c',d' mean frequencies from arbitrary pairs of four-fold tables $< a,b,c,d >$ and $< a',b',c',d' >$.

(1) A quantifier $\sim(a,b)$ is **implicational** *if always $a' \geq a,\ b' \leq b$ implies $\sim(a',b') \geq \sim(a,b)$.*

(2) A quantifier $\sim(a,b,c)$ is **Σ-double implicational** *if always $a' \geq a,\ b'+c' \leq b+c$ implies $\sim(a',b',c') \geq \sim(a,b,c)$.*

(3) A quantifier $\sim(a,b,c,d)$ is **Σ-equivalence** *if always $a'+d' \geq a+d,\ b'+c' \leq b+c$ implies $\sim(a',b',c',d') \geq \sim(a,b,c,d)$.*

Let us prove the following auxiliary propositions connected to these classes of quantifiers.

Lemma 1
A quantifier $\Leftrightarrow^*$ is Σ-double implicational iff the following conditions hold:

(i) if $b'+c' = b+c$ then
$\Leftrightarrow^*(a,b',c') = \Leftrightarrow^*(a,b,c)$,
(ii) the quantifier $\Rightarrow^*$ defined by
$\Rightarrow^*(a,b) = \Leftrightarrow^*(a,b,0)$
is implicational.

PROOF: For Σ-double implicational quantifiers, (i), (ii) are clearly true. Let $\Leftrightarrow^*$ is a quantifier satisfying (i), (ii), and $a' \geq a, b'+c' \leq b+c$. Then

$\Leftrightarrow^*(a',b',c') = \Leftrightarrow^*(a',b'+c',0) = \Rightarrow^*(a',b'+c') \geq \Rightarrow^*(a,b+c) = \Leftrightarrow^*(a,b+c,0) = \Leftrightarrow^*(a,b,c)$. □

Lemma 2
A quantifier $\equiv^*$ is Σ-equivalence iff the following conditions hold:

(i) if $a'+d' = a+d$, $b'+c' = b+c$ then
$\equiv^*(a',b',c',d') = \equiv^*(a,b,c,d)$,
(ii) the quantifier $\Leftrightarrow^*$ defined by
$\Leftrightarrow^*(a,b,c) = \equiv^*(a,b,c,0)$
is Σ-double implicational.

PROOF: For Σ-equivalence quantifiers, (i), (ii) are clearly true. Let $\equiv^*$ is a quantifier satisfying (i), (ii), and $a'+d' \geq a+d, b'+c' \leq b+c$. Then

$\equiv^*(a',b',c',d') = \equiv^*(a'+d',b',c',0) = \Leftrightarrow^*(a'+d',b',c') \geq \Leftrightarrow^*(a+d,b,c) = \equiv^*(a+d,b,c,0) = \equiv^*(a,b,c,d)$. □

We shall use the notion of **strictness** to state relations between different quantifiers:

Definition 3
A quantifier $\sim_1$ is **less strict** *than a quantifier $\sim_2$*
(or $\sim_2$ is **more strict** *than $\sim_1$)*
if for all four-fold tables $< a,b,c,d >$
$\sim_1(a,b,c,d) \geq \sim_2(a,b,c,d)$.

From the fuzzy logic point of view, it means that in all models (data) the formula $\varphi \sim_1 \psi$ is at least so true as the formula $\varphi \sim_2 \psi$, i.e. the deduction rule $\frac{\varphi \sim_2 \psi}{\varphi \sim_1 \psi}$ is correct.

Example 2
$\Leftrightarrow_{\oslash}$ is more strict than $\Rightarrow_{\oslash}$, and less strict than $\equiv_{\oslash}$.

3 Least strict Σ-double implicational quantifier corresponding to given implicational quantifier

Let $\Rightarrow^*$ be an implicational quantifier. There is a natural task to construct some Σ-double implicational quantifier $\Leftrightarrow^*$ such that from formula $\varphi \Leftrightarrow^* \psi$ logically follow both implications $\varphi \Rightarrow^* \psi$, $\psi \Rightarrow^* \varphi$, i.e. deduction rules $\frac{\varphi \Leftrightarrow^* \psi}{\varphi \Rightarrow^* \psi}$, $\frac{\varphi \Leftrightarrow^* \psi}{\psi \Rightarrow^* \varphi}$ are correct. Such a quantifier $\Leftrightarrow^*$ should be as less strict as possible to be near to $\Rightarrow^*$.

Following two theorems show how to construct the logically nearest Σ-double implicational quantifier from a given implicational quantifier and vice versa.

Theorem 1
Let $\Rightarrow^*$ be an implicational quantifier and $\Leftrightarrow^*$ be the quantifier constructed from $\Rightarrow^*$ for all four-fold tables $< a,b,c,d >$ by the formula

$\Leftrightarrow^*(a,b,c) = \Rightarrow^*(a,b+c)$.

Then $\Leftrightarrow^*$ is the Σ-double implicational quantifier which is the least strict from the class of all Σ-double implicational quantifiers $\sim$ more strict than $\Rightarrow^*$.

Remark. Let us mention that this means the following:

(1) deduction rules $\frac{\varphi \Leftrightarrow^* \psi}{\varphi \Rightarrow^* \psi}$, $\frac{\varphi \Leftrightarrow^* \psi}{\psi \Rightarrow^* \varphi}$ are correct;

(2) if $\sim$ is a Σ-double implicational quantifier such that deduction rules $\frac{\varphi \sim \psi}{\varphi \Rightarrow^* \psi}$, $\frac{\varphi \sim \psi}{\psi \Rightarrow^* \varphi}$ are correct,

then $\sim$ is more strict than $\Leftrightarrow^*$, i.e. also $\frac{\varphi \sim \psi}{\varphi \Leftrightarrow^* \psi}$ is correct.

PROOF: Since $\Rightarrow^*$ is an implicational quantifier, $\Leftrightarrow^*$ is a Σ-double implicational quantifier; moreover,

$\Leftrightarrow^*(a,b,c) = \Rightarrow^*(a,b+c) \leq \Rightarrow^*(a,b)$ for all four-fold tables $< a,b,c,d >$

so $\Leftrightarrow^*$ is more strict than $\Rightarrow^*$.

Let $\sim$ is a Σ-double implicational quantifier more strict than $\Rightarrow^*$. Then we obtain

$\sim (a,b,c) = \sim (a,b+c,0) \leq \Rightarrow^* (a,b+c) = \Leftrightarrow^* (a,b,c)$ for all four-fold tables $< a,b,c,d >$

which means that $\sim$ is more strict than $\Leftrightarrow^*$. $\square$

Example 3
For the basic implication $\Rightarrow_{\oslash} (a,b) = \frac{a}{a+b}$*, the basic double implication* $\Leftrightarrow_{\oslash} (a,b,c) = \frac{a}{a+b+c}$ *is the least strict* Σ*-double implicational quantifier satisfying deduction rules* $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\Rightarrow_{\oslash}\psi}$, $\frac{\varphi\Leftrightarrow^*\psi}{\psi\Rightarrow_{\oslash}\varphi}$.

Theorem 2
Let $\Leftrightarrow^*$ *be a* Σ*-double implicational quantifier and* $\Rightarrow^*$ *be the quantifier constructed from* $\Leftrightarrow^*$ *for all four-fold tables* $< a,b,c,d >$ *by the formula*

$\Rightarrow^* (a,b) = \Leftrightarrow^* (a,b,0)$.

Then $\Rightarrow^*$ *is the implicational quantifier which is the most strict from the class of all implicational quantifiers* $\sim$ *less strict than* $\Leftrightarrow^*$.

Remark. Let us mention that this means the following:

(1) deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\Rightarrow^*\psi}$, $\frac{\varphi\Leftrightarrow^*\psi}{\psi\Rightarrow^*\varphi}$ are correct;

(2) if $\sim$ is an implicational quantifier such that deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\sim\psi}$, $\frac{\varphi\Leftrightarrow^*\psi}{\psi\sim\varphi}$ are correct,

then $\sim$ is less strict than $\Rightarrow^*$, i.e. also $\frac{\varphi\Rightarrow^*\psi}{\varphi\sim\psi}$ is correct.

PROOF: Since $\Leftrightarrow^*$ is a Σ-double implicational quantifier, $\Rightarrow^*$ is an implicational quantifier; moreover, using 1

$\Leftrightarrow^* (a,b,c) = \Leftrightarrow^* (a,b+c,0) \leq \Leftrightarrow^* (a,b,0) = \Rightarrow^* (a,b)$ for all four-fold tables $< a,b,c,d >$

so $\Rightarrow^*$ is less strict than $\Leftrightarrow^*$.

Let $\sim$ is an implicational quantifier less strict than $\Leftrightarrow^*$.

Then we obtain

$\sim (a,b) \geq \Leftrightarrow^* (a,b,0) = \Rightarrow^* (a,b)$ for all four-fold tables $< a,b,c,d >$

which means that $\sim$ is less strict than $\Rightarrow^*$. $\square$

4 Most strict Σ-equivalence quantifier corresponding to given Σ-double implicational quantifier

This section will be a clear analogy with the previous one:

Let $\Leftrightarrow^*$ be an Σ-double implicational quantifier. There is a natural task to construct some Σ-equivalence $\equiv^*$ such that the formula $\varphi \equiv^* \psi$ logically follows both from the formula $\varphi \Leftrightarrow^* \psi$, and from the formula $\neg\varphi \Leftrightarrow^* \neg\psi$, i.e. deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\equiv^*\psi}$, $\frac{\neg\varphi\Leftrightarrow^*\neg\psi}{\varphi\equiv^*\psi}$ are correct. Such a quantifier $\equiv^*$ should be as strict as possible to be near to $\Leftrightarrow^*$.

Following theorems show how to construct the logically nearest Σ-equivalence quantifier from a given Σ-double implicational quantifier and vice versa. The proofs of these theorems are very similar to the above ones.

Theorem 3
Let $\Leftrightarrow^*$ *be a* Σ*-double implicational quantifier and* $\equiv^*$ *be the quantifier constructed from* $\Leftrightarrow^*$ *for all four-fold tables* $< a,b,c,d >$ *by the formula*

$\equiv^* (a,b,c,d) = \Leftrightarrow^* (a+d,b,c)$.

Then $\equiv^*$ *is the* Σ*-equivalence which is the most strict from the class of all* Σ*-equivalences* $\sim$ *less strict than* $\Leftrightarrow^*$.

Remark. Let us mention that this means the following:

(1) deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\equiv^*\psi}$, $\frac{\neg\varphi\Leftrightarrow^*\neg\psi}{\varphi\equiv^*\psi}$ are correct;

(2) if $\sim$ is a Σ-equivalence such that deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\sim\psi}$, $\frac{\neg\varphi\Leftrightarrow^*\neg\psi}{\varphi\sim\psi}$ are correct, then $\sim$ is less strict than $\equiv^*$, i.e. also $\frac{\varphi\equiv^*\psi}{\varphi\sim\psi}$ is correct.

PROOF: Since $\Leftrightarrow^*$ is a Σ-double implicational quantifier, $\equiv^*$ is a Σ-equivalence; moreover

$\equiv^* (a,b,c,d) = \Leftrightarrow^* (a+d,b,c) \geq \Leftrightarrow^* (a,b,c)$ for all four-fold tables $< a,b,c,d >$

so $\equiv^*$ is less strict than $\Leftrightarrow^*$.

Let $\sim$ is a Σ-equivalence less strict than $\Leftrightarrow^*$. Then we obtain

$\sim (a,b,c,d) = \sim (a+d,b,c,0) \geq \Leftrightarrow^* (a+d,b,c) = \equiv^* (a,b,c,d)$ for all four-fold tables $< a,b,c,d >$

which means that $\sim$ is less strict than $\equiv^*$. $\square$

Example 4
For the basic double implication $\Leftrightarrow_{\oslash} (a,b,c) = \frac{a}{a+b+c}$*, the basic equivalence* $\equiv_{\oslash} (a,b,c,d) = \frac{a+d}{a+b+c+d}$*, is the most strict* Σ*-equivalence satisfying deduction rules* $\frac{\varphi\Leftrightarrow_{\oslash}\psi}{\varphi\equiv^*\psi}$, $\frac{\neg\varphi\Leftrightarrow_{\oslash}\neg\psi}{\varphi\equiv^*\psi}$.

Theorem 4
Let $\equiv^*$ *be an* Σ*-equivalence quantifier and* $\Leftrightarrow^*$ *be the quantifier constructed from* $\equiv^*$ *for all four-fold tables* $< a,b,c,d >$ *by the formula*

$\Leftrightarrow^* (a,b,c) = \equiv^* (a,b,c,0)$.

Then $\Leftrightarrow^*$ *is the* Σ*-double implicational quantifier which is the least strict from the class of all* Σ*-double implicational quantifiers* $\sim$ *more strict than* $\equiv^*$.

Remark. Let us mention that this means the following:

(1) deduction rules $\frac{\varphi\Leftrightarrow^*\psi}{\varphi\equiv^*\psi}$, $\frac{\neg\varphi\Leftrightarrow^*\neg\psi}{\varphi\equiv^*\psi}$ are correct;

(2) if $\sim$ is a Σ-double implicational quantifier such that deduction rules $\frac{\varphi\sim\psi}{\varphi\equiv^*\psi}$, $\frac{\neg\varphi\sim\neg\psi}{\varphi\equiv^*\psi}$ are correct,

then $\sim$ is more strict than $\Leftrightarrow^*$, i.e. also $\frac{\varphi\sim\psi}{\varphi\equiv^*\psi}$ is correct.

PROOF: Since $\equiv^*$ is a Σ-equivalence, $\Leftrightarrow^*$ is a Σ-double implicational quantifier; moreover,

$\equiv^* (a,b,c,d) = \equiv^* (a+d,b,c,0) \geq \equiv^* (a,b,c,0) = \Leftrightarrow^* (a,b,c)$

for all four-fold tables $< a,b,c,d >$ so $\Leftrightarrow^*$ is more strict than $\equiv^*$.

Let $\sim$ is a Σ-double implicational quantifier more strict than $\equiv^*$. Then we obtain

$\sim (a,b,c) \leq \equiv^* (a,b,c,0) = \Leftrightarrow^* (a,b,c)$ for all four-fold tables $< a,b,c,d >$

which means that $\sim$ is more strict than $\Leftrightarrow^*$. $\square$

5 Conclusions

The theorems proved in the paper show that implicational, Σ-double implicational, and Σ-equivalence quantifiers compose logically affiliated triads $\Rightarrow^*$, $\Leftrightarrow^*$, $\equiv^*$, where

$\Rightarrow^*$ is some implicational quantifier,

$\Leftrightarrow^*$ is the least strict Σ-double implicational quantifier corresponding to $\Rightarrow^*$, and

$\equiv^*$ is the most strict Σ-equivalence quantifier corresponding to $\Leftrightarrow^*$.

Following deduction rules are correct for such triads of quantifiers:

$$\frac{\varphi \Leftrightarrow^* \psi}{\varphi \Rightarrow^* \psi}, \frac{\varphi \Leftrightarrow^* \psi}{\psi \Rightarrow^* \varphi}, \frac{\varphi \Leftrightarrow^* \psi}{\varphi \equiv^* \psi}, \frac{\neg\varphi \Leftrightarrow^* \neg\psi}{\varphi \equiv^* \psi}.$$

The best known example is the triad of basic quantifiers $\Rightarrow_{\oslash}$, $\Leftrightarrow_{\oslash}$, $\equiv_{\oslash}$:

$\Rightarrow_{\oslash} (a,b) = \frac{a}{a+b}$,

$\Leftrightarrow_{\oslash} (a,b,c) = \frac{a}{a+b+c}$,

$\equiv_{\oslash} (a,b,c,d) = \frac{a+d}{a+b+c+d}$.

Let us stress that to each given implicational quantifier such triad can be constructed. This can naturally extend the methodological approach used to the particular quantifier's definition for covering all three types of relations (implication, double implication, equivalence).

There are some other research tasks concerning various classes of four-fold table generalized quantifiers and their possible relations to fuzzy logic. Several results on this topic are included in [12], [13]. A general research idea is to represent useful four-fold table generalized quantifiers by fuzzy formulae and apply fuzzy logic methodology for dealing with them.

This research has been supported by grant 1M0572 of the Ministry of Education, Youth and Sports of the Czech Republic.

REFERENCES

[1] Aggraval, R. et al. (1996): Fast Discovery of Association Rules. In: *Advances in Knowledge Discovery and Data Mining.* (Fayyad, V.M. et al., eds.) AAAI Press / MIT Press, 307-328.

[2] Batagelj, V., Bren, M. (1995): Comparing Resemblance Measures. *J. of Classification 12*, 73-90.

[3] Berka, P., Ivánek, J. (1994): Automated Knowledge Acquisition for PROSPECTOR-like Expert Systems. In: *Machine Learning. ECML-94 Catania* (Bergadano and Raedt, eds.). Springer, 339-342.

[4] Fodor, J., Yager, R.R. (2000): Fuzzy Operators and Quantifiers. In: *Fundamentals of Fuzzy Sets*. (Dubois, D. and Prade, H., eds.) The Handbooks of Fuzzy Sets Series, Kluwer Academic Publishers, Dordrecht.

[5] Hájek, P., Havránek, T. (1978): *Mechanising Hypothesis Formation – Mathematical Foundations for a General Theory*. Springer-Verlag, Berlin, 396p.

[6] Hájek, P., Havránek, T., Chytil M. (1983): *Metoda GUHA*. Academia, Praha 1983, 314 p. (in Czech)

[7] Hájek, P., Sochorová, A., Zvárová, J. (1995): GUHA for Personal Computers. *Computational Statistics & Data Analysis 19*, 149-153.

[8] Hájek, P. (1998): *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht.

[9] Hájek, P., Holeňa, M. (1998): Formal Logics of Discovery and Hypothesis Formation by Machine. In: *Discovery Science* (Arikawa, S. and Motoda, H., eds.), Springer-Verlag, Berlin, 291-302

[10] Ivánek, J., Stejskal, B. (1988): Automatic Acquisition of Knowledge Base from Data without Expert: ESOD (Expert System from Observational Data). In: *COMPSTAT'88 Copenhagen*. Physica-Verlag, Heidelberg, 175-180.

[11] Ivánek, J. (1999): On the Correspondence between Classes of Implicational and Equivalence Quantifiers. In: *Principles of Data Mining and Knowledge Discovery. Proc. PKDD99 Prague* (Zytkow, J. and Rauch, J., eds.), Springer-Verlag, Berlin, 116-124

[12] Ivánek, J. (2004): Using Fuzzy Logic Operators for Construction of Data Mining Quantifiers. *Neural Network World 5/04*, 403-410.

[13] Ivánek, J. (2005): Construction of Implicational Quantifiers from Fuzzy Implications. *Fuzzy Sets and Systems* (to appear).

[14] Rauch, J. (2005): Logic of Association Rules *Applied Intelligence 22 (1)*, 9-28

[15] Rauch, J., Šimůnek, M. (2005): An Alternative Approach to Mining Association Rules. In: *Data Mining: Foundations, Methods, and Applications* (Lin T.Y., Ohsuga S., Liau C.J. and Tsumoto S., eds.). Springer Verlag (to appear)

[16] Zembowicz, R., Zytkow, J. (1996): From Contingency Tables to Various Forms of Knowledge in Databases. In: *Advances in Knowledge Discovery and Data Mining* (Fayyad, U.M. et al.,eds.). AAAI Press / MIT Press, 329-349.