ON MODELLING OF A LINGUISTIC THESAURUS WITH VAGUE TERMS BY MEANS OF IMPLICATIONS WITH CERTAINTY FACTORS

by Jürgen Herzog,
Fachhochschule Würzburg-Schweinfurt, Germany

Fuzzy Logic and working with certainty factors offer the following model of a linguistic thesaurus, which can be used in expert systems for different fields of knowledge (e.g. in jurisprudence or in any other thesaurus based on natural or technical language).

I. Structure of Implications

Firstly the skeleton of implications will be shown:

1.) There is a term (A) in a natural or a technical language, for instance a legal conception, which is - more or less - implicated by facts (F). These facts are represented in terms of a natural or technical language. The certainty of the implication from F to A (F ---> A) may be doubtful and controversial. But we can prove that F ---> A is true by the transitivity of implications. If there is a term - we name it "connecting link" (C) - which implicates A and which is implicated by F, then the implication F ---> A is true:

If (F ---> C) and (C ---> A) then (F ---> A).

EXAMPLE: In order to stop job reductions, a new - fictitious - law lays down that a company must pay a special tax when it introduces "innovations". There are conflicting opinions if "a new machine which is a substitute for a destroyed older one - (F)" is an innovation (A). When our thesaurus gives us a definition
"A <---> C1 = the procurement or introduction of new equipment, constructions, installations, devices, machines, processes or forms of organization" (- regardless of the concrete possibility of job reduction, because the danger of job reduction is given abstractly in every case -) then the implications F ---> C1 and C1 ---> A are true. Then it is also proved that F ---> A is true.

2.) a) C can be a term which only implicates A, e.g. a special case or instance (Cc).
b) C can also be a definition (Cd); then we have an equivalence between Cd and A: (Cd ---> A) and (A ---> Cd) or (Cd <---> A).

3.) a) The thesaurus can tie up A in generic terms (G). Then A implicates G.
b) The thesaurus can give criteria or properties of A. We will name these criteria or properties P.
The conjunctions of G and these properties or criteria give automatically - more or less suitable - p r o p o s a l s
for definitions in the classical manner of finding the "genus proximum et differentia specifica":

If (G and P) then Cd can be.

In this way new "connecting links", i.e. "definitions", can be found in order to prove that F ---> A is true.

II. Certainty

In practice the relations between F, Cc, Cd, A are not certain in every case.

1.) Often so-called "definitions" - as we know - are not definitions in a logical sense but "explanations" which are not true but only probable. Here Cd is not strictly equivalent to A; the equivalence is only given to a - more or less - certain degree. This means that the implications Cd ---> A and A ---> Cd respectively are probable with a subjective certainty factor.

2.) The same problem is the certainty of F ---> Cd or
of F ---> Cc or - as mentioned before - of F ---> A. Also the certainty of these relationships is mostly determinable only in a subjective way; the result is subjective probability.

3.) But by operating with certainty factors we can improve the certainty of the implication F ---> A , e.g. in this way:
Firstly the certainty factor of F ---> A seems to be 0.6 on a scale from 0.0 to 1.0. However, if the certainty factor of
F ---> Cd is 0.8 and the certainty factor of Cd ---> A is 0.8 then it is proved that the certainty factor of F ---> A is 0.8.
This is based on a generalization of the transitivity of implications. Also if the certainty factor of F ---> Cd is 0.7 and the certainty factor of
Cd ---> A is 0.8 and so on the result can be computed.

> In our EXAMPLE the thesaurus can offer "C2 = the procurement or introduction of new equipment,.... , which leads to job reductions". Then the implication F ---> A is false.
> But we cannot say that one of these two definitions (C1 or C2) is exactly true and the other false. The second definition (C2) has -subjectively - a better certainty factor than the first one (C1).

III. Fuzziness

1.) The terms or linguistic expressions F, Cd, Cc, A, G, P are often vague. Here is the point of attachment to Fuzzy Logic with values like "nearly true", "half true", "scarcely true" or other ones.

2.) The vagueness of these terms can have different causes:
a) A term or a linguistic expression for instance is vague
- when it is uncertain if and how far other terms implicate it
- or when it is uncertain if and how far it implicates other terms.
To avoid errors we must be careful that in these cases the fuzziness of the term or expression mostly corresponds in some way or other with the (un)certainty of implications. In our thesaurus this can concern the relationship between Cd and A or between A and G or A and Cc or P.
b) A term or a linguistic expression is also called vague when we don't know how far it corresponds to the reality that it wants to denote. In our thesaurus this is especially possible with F.
In this case the fuzziness of this term or expression mostly doesn't correspond to certainty factors of implications from this term or expression to other terms and vice versa.
c) Vagueness of terms can also have other causes, e.g. in value judgements or in other terms which are imprecise from the beginning. In our thesaurus this is possible with a l l terms or linguistic expressions.
Here we must investigate in each case if - and if need be how - fuzziness corresponds to certainty factors.

3.) The control of fuzziness and certainty factors must be taken from the "language game" (in the terminology of L. Wittgenstein) which the expert system wants to apply to the thesaurus.

That is why we will not continue our EXAMPLE; we would encounter special problems of legal language games which are not the the problems of this paper; here a general model of reasoning in a thesaurus will be shown.

IV. Result

An expert system can make inferences or conclusions in - and by means of - a thesaurus which is constructed in this way. The conclusions go into all directions which the net of implications gives in such a thesaurus. Fuzzy terms and certainty factors let the system reason in a manner similar to a natural language or a technical language based on a natural language.

Dedication: I dedicate this paper to Prof. Dr. Ernest Czogala, Technical University of Silesia, Gliwice, Poland, as an expression of my appreciation of his valuable information about very important aspects of Fuzzy Logic.