A Fuzzy Expert System Approach to Insurance Risk Assessment Using FuzzyCLIPS

Luis A. Carreno
Dr. Yashvant Jani
Togai InfraLogic, Inc.
17000 El Camino Real, Suite 208
Houston, TX 77058
Phone: (713) 480 - 8904, Fax: (713) 480 - 8906

Abstract

A knowledge based system (KBS) that combines fuzzy processing with rule-based expert system is developed to provide an improved decision aid for evaluating risk for life insurance. This expert system application illustrates the use of FuzzyCLIPS tool to build a knowledge based decision support system, capable of possessing fuzzy components to improve user interactions and KBS performance. The design of the fuzzy solution consists of a CLIPS rule-based system combined with fuzzy logic rules. The results employing FuzzyCLIPS are compared with the results obtained from the solution of the problem using traditional numerical equations. This paper briefly describes the problem, proposes a solution, describes the test scenarios, presents the results and conclusions, and provides a sample output of the software product.

Introduction to FuzzyCLIPS

FuzzyCLIPS adds fuzzy processing capability to CLIPS (C Language Integrated Production System) version 5.1. CLIPS was developed by NASA/JSC as a rule-based expert system development tool. FuzzyCLIPS architecture is a separate processing element similar to that used to incorporate object-oriented programming into CLIPS [1,2]. The basic fuzzy constructs and function calls (like definition of a membership function and fuzzy rules) can be written intermixed with usual CLIPS statements providing an extension of rule syntax and user definitions of membership function types. Principal fuzzy constructs define rule bases and membership functions. There are also functions by which a CLIPS [2] program can test the degree of membership of a sensor value, execute a fuzzy rule base that returns defuzzified control values to CLIPS and, optionally, assert facts giving belief values for the possibilities that might be useful in an expert system. In addition, C interface functions support embedded fuzzy applications that can invoke the fuzzy processor directly for speed in embedded control applications. The main features of FuzzyCLIPS are: a) fuzzy reasoning capability is combined with conventional rule based technology, b) the flexibility and portability of CLIPS is retained, and c) development of both stand-alone and embedded systems is possible.

Problem Statement

An insurance company needs to assess the degree of health risk associated with each client based on physical characteristics such as height, weight, age and additional information such as exercise, smoking, drinking, and eating habits. The output risk value serves as the basis for the determination of insurance premiums billed to clients. Generally, insurance premiums have a base rate (perfect health, good habits, 35 years old) and an increment to adjust the premium based on the risk for a particular client. A risk value between 0 and 1 suffices to set a net rate. The equation is

$$\text{Cost} = \text{Base Rate} + ((\text{Risk} / \text{Base Risk}) - 1) * \text{Increment} \quad (1)$$

The relation between decision factors to compute the risk and the rate change need be neither incremental nor linear. Complex interdependence of the factors mean that computer-based decision aids (a software system) are useful to a human agent and that sharp decision boundaries such as those produced by a normal rule based system are sensitive to small uncertainties in the input data. Fuzzy logic [8.9,10,11] provides a basis for accommodating such uncertainty with finesse. It also allows the software system to be defined in human-like terms and aids in the transfer of human knowledge and intuition into a KBS.

The system has two different types of inputs: base and incremental. The base input variables are Age (A), Weight (W), and Height (H). Incremental input variables deal with particular habits and characteristics of prospective clients. Such variables are: exercising (E), dairy products intake (DI), red meat intake (MI), vegetable intake (VI), fat/sweet intake (FSI), smoking (S), and drinking (D) habit. The output of the system is the risk used in equation (1).

The body mass index (BMI) is a measure that indicates if a person is overweight [3]. It is calculated by dividing the Weight in kilograms by the square of the height in meters,
BMI = Weight/(Height)2.

Table I shows the scale used to interpret BMI and the corresponding BMI-risk .

Table I. Risk contribution due to BMI

BMI	Condition	BMI-risk
under 23	Underweight	0.25
23 -25	Ideal	0.0
25 - 30	Overweight	0.75
over 30	Obese	1.0

Traditional Numerical Solution

For the traditional method solution, we treat all of the variables as a number input or a selection from a finite, discrete, closed set of possibilities. Each variable is represented as a lookup table of intervals where the value of the corresponding risk is specified for each interval. For example, Table II provides the contribution to risk due to the Age.

Table II. Risk contribution due to age

Age	Age-risk
0 to 30	0.25
31 to 60	0.5
61 to 90	0.75
> 90	1.0

This table could be used in a rule-based KBS in the following form

```
(age ?age&:( <= ?age 30)
        => (assert (age-risk .25))
```

When each factor has been evaluated to provide an intermediate risk, the total risk can be computed as a weighted combination of these risks due to various factors. In a traditional system, the first step in the solution is to define a mathematical relationship between the inputs and outputs of the system. The objective is to obtain a numerical value that represents the possible risk of a person having medical problems due to his physical characteristics and eating habits. Risk is defined as having a range of [0,1]. The various factors are also assumed to have values in the [0,1] range by mappings similar to those presented above for age and BMI. A risk measure of 1 represents the maximum degree of risk, on the contrary, a measure of 0 or less represents the minimum degree of risk [4].

In addition to age and BMI, data reflecting a person's habits also contribute to the risk assessment. Two approaches are used to handle such data. The normal approach is to attempt to quantify habits in terms of frequency of the participation and amount of time, or activity concerned. The second approach is to estimate the frequency and level of activity into literal categories (or linguistic values). Qualitative values indicating the change in risk due to various habits is shown in Table III [7].

Table III. Relationship between habits and health risk

	Risk Increases	Risk Decreases
Smoking	High	None
Drinking	High	None
Exercising	Low	High
Veg. Intake	Low	High
Meat Intake	High	Low
Dairy Intake	High	Low
Fat/Sweet Intake	High	Low

Fuzzy Logic Solution

In a fuzzy logic based system, an expert defines the rules to describe the characteristics of the risk assessment for each factor [5,6]. The input variables are processed by these rules to generate an appropriate output. A schematic fuzzy decision support system is shown in figure 1. For fuzzy reasoning max-dot inferencing and centroid defuzzification techniques are used.

For this particular example, five different sets of fuzzy rules are defined. The first rulebase computes a risk_1 based on age and BMI. The second rulebase computes a risk_2 based on smoking and drinking habits. The third rulebase computes a risk_3 based on the amount of exercise and intake of vegetables. The last rulebase computes a risk_4 based on the intake of dairy products, red meat, and fat and sweet products. A fifth rulebase relates risks 1-4 to the overall risk to complete the risk assessment. The importance of breaking down the problem into smaller related groups is the fact that the number of rules needed to control the system decreases dramatically. In our example, the number reduced from $4^2 * 3^7$ (34992) rules to a maximum of 686 rules.

When a client provides age, height, and weight, the BMI is computed and an initial measure of risk , risk_1, using Rbase1 is evaluated. This measure serves as the basis for subsequent decisions. If the risk obtained is considered by the system as very high, no further inquiries of the user are necessary. On the other hand, if the risk obtained is considered low, medium, or high, further inquiries into the user's habits are necessary to arrive at a more meaningful result.

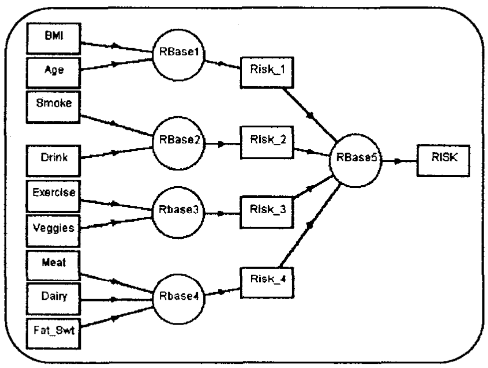

Fig.1 A schematic view of the fuzzy logic risk assessor

The output of the system consists of a crisp value for Risk in the range [0, 1]. The system also produces a truth value associated with each output fuzzy set, i.e., the degree to which each fuzzy set defining risk contributes to the output value of risk.

For the fuzzy logic method, as seen in the following table, we defined the following sets of membership functions associated with each variable.

Table IV. Variables and its membership functions

Variable	Membership Functions
Risk	Very Low, Low, Medium, High, Very High
BMI	Under, Ideal, Over, Obese
Age	Very Low, Low, Med, High
Habits	Low, Medium, High

The universe of discourse for each of the above fuzzy variables is [0,1.2] for each risk (Fig. 2), [0,60] for BMI (Fig. 3), and [0,100] for Age (Fig. 4).

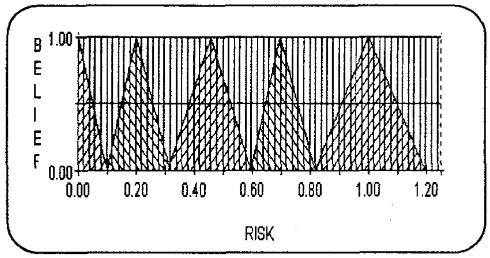

Fig. 2 Risk Membership Functions

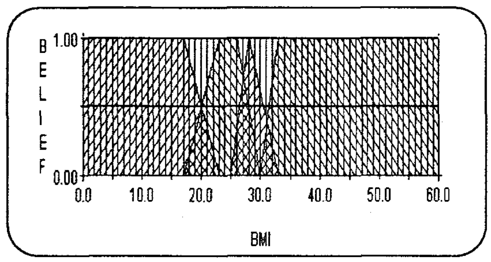

Fig. 3 BMI Membership Functions

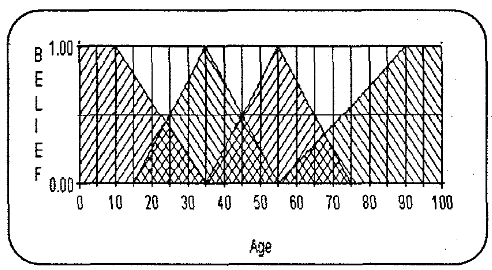

Fig. 4 Age Membership Functions

A sample set of fuzzy logic rules for final value of Risk, based on all the inputs, is shown in Table V.

Table V. Sample set of rules for final value of risk

RULE A		RULE B	
IF	Age is Low	**IF**	Age is High
	& BMI is Ideal		& BMI is Over
	& E is Medium		& E is Low
	& VI is High		& VI is Low
	& DI is Medium		& DI is High
	& MI is Low		& MI is High
	& FSI is Low		& FI is Low
	& S is Low		& S is Medium
	& D is Low		& D is Med
THEN	Risk is Low	**THEN**	Risk is High

As explained earlier, a rulebase with that many inputs is difficult to implement due to the large number of possible combinations of the input variables. Examples of fuzzy rules, using the alternative approach of breaking down the input variables into smaller and related groups, is shown next.

IF Age is High &
BMI is Obese
THEN Risk_1 is Very High

IF S is High &
D is Low
THEN Risk_2 is High

IF E is High &
VI is Medium
THEN Risk_3 is Low

IF MI is Low &
DI is Medium &
FSI is Medium
THEN Risk_4 is Medium

In the application, five rulebases are defined. As explained earlier, each one produces an intermediate risk that is fed into the final rulebase to provide a relative assessment of risk. Such risk is compared with a base risk associated with a base rate, as explained in section 2. The total risk of a particular person is calculated and substituted in Eq. (1) to produce a premium amount.

There are two special cases in the processing of the problem: 1) if the initial risk, based on age and BMI, is greater than 0.8, the risk is considered very high. Therefore, there is no need for further processing of the system. 2) if the initial assessment of BMI is greater than 30, meaning the person is obese, all questions related to the habits of consumption of dairy products, red meat, and fat/sweet products are omitted. Otherwise, the user interface is the same for both methods.

Results and Conclusions

To compare both methods, the traditional vs. the fuzzy logic based, two experiments were designed. For both experiments, a batch program was created to process the two methods and to produce the individual output for each method. In the first experiment, sample data was created and processed by the batch program to obtain the output risk. The sample data consisted of a group of persons with constant weight/height relationship, and constant eating and exercise habits, the only variant was the age of the individuals. The constant characteristics are: 1) BMI Ideal, 2) S is no, 3) D is Low, 4) VI is High, 5) FSI is Low, 6) E is high, 7) DI is Low, and 8) MI is Low.

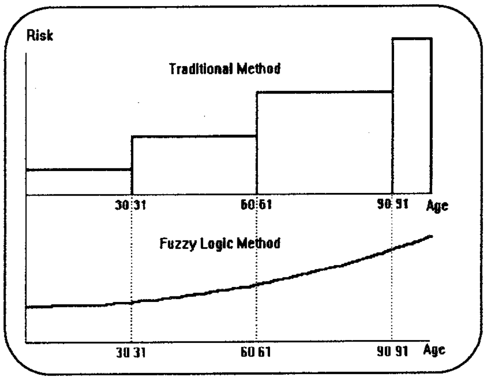

Fig. 5 Risk Vs Age for Constant Characteristics and Habits

The results were as expected. For the traditional method, abrupt changes occurred in the value of risk associated with ages. As observed in figure 5, the risk value jumps at age

30 and then continues to be constant until it reaches the age of 60 where it jumps again. The process is repeated at age 90.

For the fuzzy logic solution, as observed in Fig.5, no sharp differences or jumps are observed at any specific age, i.e., the risk values increase smoothly along the whole domain. The fuzzy system produces more realistic values for different ages, specially for those cases in which the age varies from 30 to 31, 60 to 61, or 90 to 91.

In the second experiment a sample data set was created to produce a population of 1000 subjects with random physical characteristics (age, weight, and height) and random eating and exercise habits. The goal of the second experiment was to observe the behavior of the system in a totally random environment.

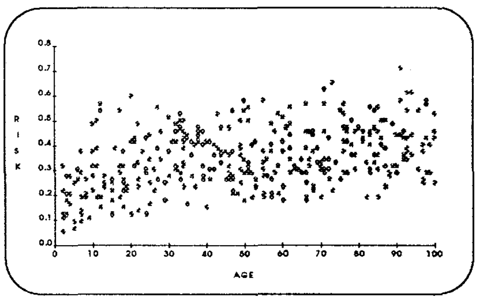

Fig. 6 Risk Obtained From Traditional Method

As observed in figure 6, the behavior of the system in the traditional method is very unpredictable and produces a totally random spectrum of values for the output risk. The most important fact is that there is no correlation found between the age of the subjects and the risk assigned to that particular subject. This result contradicts the common sense reasoning about the relationship between the age and risk. The results obtained using the fuzzy logic method, as shown in figure 7, show a more predictable and smoother behavior. The relationship between the age and risk is maintained, that is, the risk increases with the increase in age.

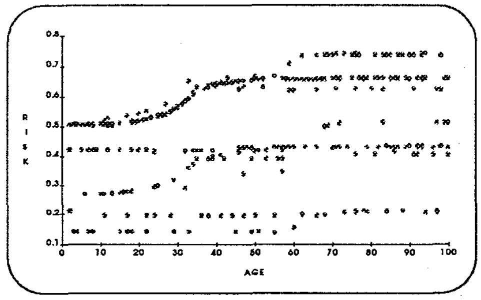

Fig. 7 Risk Obtained From Fuzzy Logic Method

Another important observation in figure 7 is the clearly defined grouping of risk values at different levels, which are related to eating and exercise habits of the population. This result allows a better evaluation of the risk at any time based on the particular habits of the individual.

The insurance risk assessor system, using fuzzy logic principles, provides insurance companies and insurance clients with several advantages over traditional expert systems: a) the expert system behavior can be controlled and modified without major consequences to the existing expert system, b) the results do not rely on mathematical models that could become obsolete, c) the premiums and risks are properly distributed among the population, d) there is appropriate predictability of risk measures and premium amounts, and e) the risk measure is obtained with high degree of certainty.

Some of the advantages for the insurance clients are: 1) a fair distribution of risk and premiums among population, eliminating current over/under payments, 2) a definite predictability of future premiums, eliminating uncertainty about future coverage, and 3) identification of weak areas that can be improved to reduce the risk, therefore reducing future premium amounts.

The insurance risk assessor described in this paper represents only a sample of business applications using fuzzy logic principles. Business systems with a large number of inputs, such as the insurance risk assessor, can be prototyped in a relatively short period of time using the FuzzyCLIPS tool.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.