Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 11 (2013) 4 – 10

Procedia Technology

The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)

A Decision Model for Insurance Advisors: A Case Study

Siti Fatimah Abdul Razak*, Shing Chiang Tan, Way-Soong Lim

Multimedia University, Jln Ayer Keroh Lama,Bukit Beruang, Melaka 75450, Malaysia

Abstract

This study presents a decision model using fuzzy inference system (FIS) for insurance advisors to identify and suggest appropriate policies to potential or existing clients which can minimize the subjective prejudice of the insurance advisors. The proposed model consist of four main process- (i) selecting inputs and outputs, (ii) identifying membership functions, (iii) constructing fuzzy rule base and (iv) validating the rule base. The decision model is developed based on five types of insurance policies under the Family Plan. Five attributes are selected which are age, gender, marital status, monthly income and job risk. These inputs are transformed into fuzzy variables using triangular membership functions and then used to construct the fuzzy rule base. Apart from machine learning, an expert is also engaged to verify the model. To validate the model, records of new policy subscriber are applied. Results and findings are also discussed at the end of this paper.

Keywords: fuzzy logic; fuzzy inference system; decision support; insurance

1. Introduction

Insurance advisor is the first contact the insurance company has with a consumer. He or she is responsible to help consumer assess their insurance needs and complete the formalities required to purchase an insurance policy. Insurance companies rely on these agents to obtain less easily quantifiable information about the applicant especially in risk assessment. It is widely acknowledged that agents often employ subjective criteria in evaluating

* Corresponding author.
E-mail address: fatimah.razak@mmu.edu.my

doi:10.1016/j.protcy.2013.12.155

insurance applicants and suggesting suitable policies. Decision to purchase an insurance policy is not trivial matters as there are many risk factors which may influence the decision. Generally, there are three main categories of insurance – general, family or group. Each category includes a list of policies which consumers may choose from. The backgrounds of potential purchasers of the policies are quite different. Additionally, the motivation or likelihood of buying a particular insurance significantly varies from insurance to insurance. It is an important issue for insurance advisors to propose a suitable insurance for their clients.

This study presents a decision model using fuzzy inference system (FIS) for insurance advisors to identify and suggest appropriate polices in the Family Plan based on selected attributes. Below, we briefly introduce the characteristics of the five insurance policies under this category (Table 1). The fuzzy rule base was prepared based on 438 historical data from one of the major industry players in Malaysia. The evaluation models proposed by this study provide tools for insurance consultants to determine insurance purchasing policies for their customers which can reduce the subjective prejudice of the insurance advisors.

Table 1. Insurance policies under the Family Plan

Policy Type	Description
Education	A plan which helps consumer to set aside a small amount each month towards his or her child's education that will grow into a substantial sum as time goes by.
Savings	A plan that participates in profit, which will be distributed back to participants.
Investment	An investment plan which covers death and total permanent disability benefit.
Pilgrim	A plan which provides a mean to accumulate fund to perform hajj or other noble intention.
Women	A plan which provides four different plans to protect women against specific cancer and female illness, maternity benefit option, death and total permanent disability.

2. Fuzzy Inference System (FIS)

2.1. The components

Fuzzy inference system (FIS) refers to a system which uses fuzzy logic and fuzzy set theory to map inputs to output(s) [1]. The basic structure involves five main components – a database, the fuzzifier, the inference engine, the defuzzifier, and a fuzzy rule base (see Fig. 1).

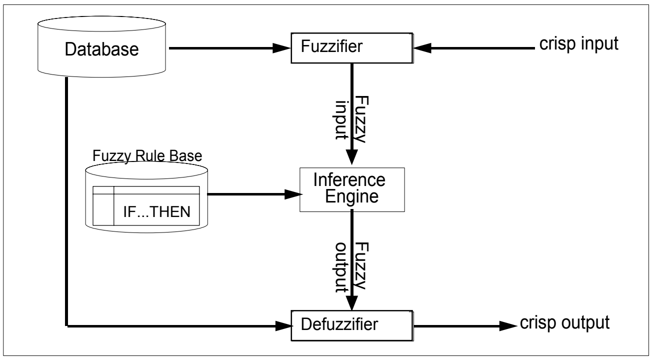

Fig. 1. Fuzzy Inference System (FIS) components

The fuzzifier is responsible to fuzzify crisp inputs from the database into linguistic values to enable association to the input linguistic variables. After the fuzzification process, the inference engine will refer to the fuzzy rule base containing fuzzy IF-THEN rules to derive the linguistic values for the intermediate and output linguistic variables. The two main steps in the inference process are aggregation and composition. Aggregation is the process of computing for the values of the IF (antecedent) part of the rules while composition is the process of computing for the values of the THEN (consequent) part of the rules. During aggregation, each condition in the IF part of a rule is assigned a degree of truth based on the degree of membership of the corresponding linguistic term. From here, either the minimum (MIN) or product (PROD) of the degrees of truth of the conditions is usually computed to clip the degree of truth of the IF part. This is assigned as the degree of truth of the THEN part. Once the output linguistic values are available, the defuzzifier produces the final crisp values from the output linguistic values.

2.2. Fuzzy inference mechanism

There are two common inference methods which are Mamdani and Sugeno. Mamdani method requires the centroid of a two-dimensional shape by integrating across a continuously varying function. It involves four steps – (i) fuzzification of the input variables, (ii) rule evaluation, (iii) aggregation of the rule outputs and (iv) defuzzification. It is widely accepted for capturing expert knowledge since it allows the expertise to be described in more intuitive and human-like manner. It is also useful when only a small number of variables are involved due to its simple structure of 'min-max' operations. Compared to Mamdani, Sugeno method is computationally effective and works well with optimization and adaptive techniques, which makes it very attractive in control problems, particularly for dynamic nonlinear systems. It is very similar to the Mamdani method. The difference is in the rule consequent which is not a fuzzy set but a mathematical function of the input variable. Since the process of proposing a suitable insurance policy for a prospect is more intuitive and requires human judgment, the Mamdani inference method is applied in this study.

2.3. Related work

De Wit [10] was the first who worked on applying fuzzy logic to insurance area. His work was to quantify the fuzziness in insurance underwriting. Since then, the fuzzy logic technologies which include fuzzy set theory, fuzzy numbers, fuzzy arithmetic, fuzzy inference systems, fuzzy clustering, fuzzy programming, fuzzy regression, and soft computing have been applied in insurance-related areas.

Shapiro [3] has discussed the major application areas of insurance within 20 years' time frame (1984-2002). It includes classification, underwriting, projected liabilities, ratemaking and pricing, and asset allocations and investments. However, there are not many work proposed in the literature to develop an evaluation model for selecting insurance policies. In 2007, Huang, Lin and Lin [4] proposed an evaluation model for purchasing life insurance and annuity insurance using analytical hierarchy process (AHP) and fuzzy logic. Four factors are considered as the inputs of the proposed model including age, annual income, educational level and risk preference. They later extended the model for purchasing five types of insurances including life, annuity, health, accident, and investment-oriented insurances. The new hybrid model incorporated the Delphi technique into their proposed evaluation model in 2008 [5].

Lazim and Mohd. Nordin [6] proposed the fuzzy rules-based method to classify the likelihoods of purchasing health insurance based on three risk factors which are age, salary and risk of illness. However, to our best knowledge, there is no work on developing a decision model using fuzzy inference system specifically for insurance advisors based on real historical dataset.

3. Case study and results

In this study, the fuzzy inference system is developed according to the framework illustrated in Fig. 2. The rule base was prepared based on 438 Family Plan subscribers record. There are a few steps involved to prepare the rule base.

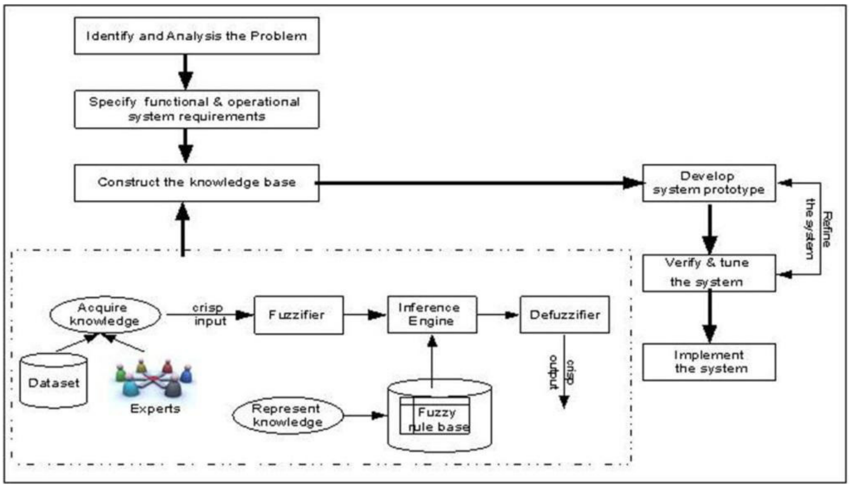

Fig. 2. The framework

3.1. Selecting input and output

The dataset consists of 33 initial variables are pre-processed. Variables are evaluated to identify subset of variables which works best for prediction. The internal 10-fold cross validation technique where the entire dataset is divided into 10 non-overlapping pairs of training and test sets was applied. Each training covers 0.9 (*(n-1)/n*th) of the dataset while the related test set covers the remaining 0.1 (1/*n*th). Following developments of models with the training set, the predictions for the test set are performed. Thus, predictions are made for all variables of the initial dataset, since each of them belongs to one of the test sets. Using Weka, *BestFirst* search method was chosen. The *BestFirst* method may start with the empty set of variables and searches forward (default behavior), or starts with the full set of attributes and searches backward, or starts at any point and searches in both directions (by considering all possible single variable additions and deletions at a given point).

The correlation coefficients values of the proposed variables are compared for two different attribute evaluators which are *CfsSubsetEval* and *WrapperSubsetEval* (Table 2). *CfsSubsetEval* is the default variable evaluator in *Weka*. This method evaluates the worth of a subset of variables by considering the individual predictive ability of each one along with the degree of redundancy between the variables. Subsets of descriptors that are highly correlated with the property/activity values and having low inter-correlation are preferred. *WrapperSubsetEval* evaluates attribute sets by using a learning scheme. Cross validation technique is used to estimate the accuracy of the learning scheme for a set of attributes.

An expert is also involved in the process of identifying the input and output variables. The expert has been in the industry for more than 10 years and owns his own firm. The correlation coefficient values based on proposed variable from the expert is also compared.

Table 2. Comparison of correlation coefficients values

Selected variable	CfsSubset_Eval	WrapperSubset_Eval	Expert
Proposed Variables	Age, Gender, Relationship Marital Status, BasicSum, Term, Policy type	Age, Gender, MonthlyGross, BasicSum, Term, Policy type	Age, Gender, Marital status, Job risk, Monthly income, Policy type
Correlation coefficient	0.9104	0.9728	0.8846

All three evaluator indicates strong positive relationship between the proposed variables. However, considering the expert point of view and experience, the input variable that may influence the likelihoods of purchasing any of the five policy type (Table 1) in this plan is identified. The input variables are *age*, *gender*, *monthly income*, *marital status* and *job risk*, whereas the output is the p*olicy type*.

3.2. Defining membership functions

Fuzzy logic aims to capture the impression of human perception and to express it with appropriate mathematical tools. Since fuzzy logic enables the use of non-numerical attributes, both qualitative and quantitative attributes can be used for this study. The fuzzy sets theory proposed by Zadeh [2], uses a gradually changing value (matching degree μ) to describe the belonging relationship between an element and a set. The matching degree can conceptually be considered as the degree of how a particular value (element) belongs to a fuzzy set. Fuzzy variables are therefore employed to represent the linguistic expressions of human beings due to the property of uncertainty. A membership function is utilized to define how each point in the input space (referred to as the *universe of discourse*) is mapped to a membership value (or degree of membership) between 0 and 1. The most important feature of linguistic variables is that every term of a linguistic variable represents a fuzzy set. The membership function of the fuzzy set is defined over the domain of the corresponding attribute. There are a few different types of membership functions, which are triangular, trapezoidal, Gaussian, bell-shape, sigmoidal and etc. In this study, the triangular membership function which is applied by most researchers in similar research area [1][4][5][6] is applied.

Five inputs were used for the evaluation models including *age*, *gender*, *monthly income*, *marital status* and *job risk*. The five inputs were expressed in fuzzy variables using five triangular membership functions (Table 3). The output is five different policy types with three membership functions referring to its likelihood of purchasing the policy which are less likely [0-0.4], likely [0.3-0.7] and most likely [0.6-1.0].

Table 3. Input variables and its membership functions

Input Variable	Membership functions
Age	young adult : below 25 years old; adult : 25-35 years old; mature adult: more than 35 years old
Gender	Male; Female
Marital status	Single; Married; Others: divorcee, widower
Monthly income	very low: below 1300; low: 1000-2500; moderate: 2000-3500; high: 3000-4500; very high: more than 4000
Job risk	low: teacher, executives, clerk; moderate: salesman, carpenter, businessman; high: policeman, doctor, nurse, factory worker

3.3. Constructing fuzzy rules

A set of linguistic rules are constructed to construct the fuzzy rule base. In a fuzzy inference system, a rule base should be defined based on the characteristics for each variable or feature. A single fuzzy If-Then rule assumes the form of IF x is A_1 THEN y is B_2 where A_1 and B_2 are linguistic variables defined by fuzzy sets on the ranges (i.e. universe of discourse) X and Y respectively. The IF-part of the rule 'x *is* A_1' is called the antecedent or premise and the THEN-part of the rule 'y *is* B_2' is called the consequent. The rules are used to describe the importance of the factors on consumers over the possibility of purchasing insurance policies. The input variables are processed by these rules to generate an appropriate output. The rules may be provided by an expert or via machine learning.

In this study, Guaje v2.0 [7] is used to induce the rules based on the dataset. Two fuzzy algorithms are applied which are Wang and Mendel algorithm (WMA) [8] and fast prototyping algorithm (FPA) [9]. WMA is among the first methods to induce fuzzy rules from a dataset. Since then, the paper discussing the algorithm has been cited by more than 1156 publications listed in ISI Web of Knowledge repository. WMA starts by generating one rule for each data sample in the training set but new rules will compete with existing ones. As a result, it generates complete rules while simultaneously considering all the available variables. To complement rules induced by WMA, Fast Prototyping Algorithm (FPA) induced rules which are at a higher level of abstraction compared to WMA. The rules are more general than the ones produced by WMA. It starts by generating a grid with all possible combinations of input labels (complete rules like the ones generated by WMA) and then, in an iterative process, outputs are defined by removing redundancies and inconsistencies.

3.4. Evaluating fuzzy rules

The rules induced by both WMA and FPA are evaluated based on accuracy and confidence value. Accuracy refers to the percentage of data samples properly classified, see Eq. (1).

$$Accuracy = 1 - (EC + AC(error) + UC)/(1!\,Data\ (TOTAL)\,) \tag{1}$$

UC (*unclassified cases*) refers to number of data samples that do not fire at least one rule with a degree greater than the pre-defined threshold. *AC(error)* (*ambiguity cases)* is the number of remaining cases related to error cases for which the difference between the two highest output confidence levels is smaller than an established threshold. *EC* (*error cases)* reflects the number of remaining cases for which the observed and inferred output classes are different. *Data (TOTAL)* refers to the total number of instances in the dataset.

Confidence refers to the probability that a (randomly selected) class is relevant. *TP* is the number of data samples that are classified as belonging to class, C_i and they actually belong to such class and *FP* is the number of data samples that are classified as belonging to class, C_i but they are actually related to another different class, see Eq. (2).

$$Confidence = TP / (TP+FP) \tag{2}$$

Apart from that, an expert is also engaged to identify significant rules. The expert identified 41 rules which are later entered in the Rule Editor provided in Matlab©2010 Fuzzy Logic Toolbox. All 41 rules identified are also part of the induced rule by machine learning (WMA and FPA, respectively). We set the inference options for Mamdani as MIN-MAX-PROD and applied the centroid method for defuzzification. The rule base was validated by applying the predictive verification technique which requires historic dataset and known results. In this study, past input data from historical dataset are compared with corresponding result for 50 new policy holders. Example of the rules with confidence value of 1.00 is shown in Table 4.

Table 4. Sample rules

No	Rules
1	IF(Age is young adult) and (JobRisk is high) and (Monthly Income is very_low) THEN (SavPol is mostlikely) (1.00)
2	IF(Age is young adult) and (JobRisk is moderate) and (Monthly Income is low) THEN (InvestPol is likely) (SavPol is likely) (WomenPol is mostlikely) (1.00)
3	IF (Age is mature adult) and (JobRisk is low) and (Monthly Income is moderate) THEN (eduPolicy is likely)(pilgrimPolicy is likely)(investPolicy is likely)(savingsPolicy is likely) (1.00)
4	IF (Age is adult) and (Gender is Male) and (JobRisk is high) and (MonthlyIncome is high) THEN (EduPol is likely)(PilgrimPol is likely)(SavPol is mostlikely) (1.00)
5	IF (Age is mature adult) and (JobRisk is low) and (Monthly Income is very_high) then (InvestPol is mostlikely) (1.00)

There are rules which only propose one type of insurance policy like rule number 1 and 5 in Table 5 and there are also rule with more than one proposal, i.e. rule number 2, 3 and 4. However, the likelihood of those policies is not equivalent with one and another. Therefore, insurance advisors may use these rules as a guide for them to propose suitable insurance policies to their fellow prospects. There are also rules with confidence value of less than 1.00 which are not shown in this paper. This is because the confidence value of each rule depends on the dataset which is based on historical data. A summary of the rule base of each origin is described in Table 5. The accuracy and confidence of this rule-base is 0.687 and 0.624 which is higher than WMA-based rule base. FPA-based rule base have higher confidence value of 0.824. However, the number of rules is three times more than the expert-based rules. This could be due to redundant rules and similar rules. Therefore, the rule base proposed by the expert is sufficient even though the number of rules is relatively small. Additionally, the rules are also induced by both algorithms.

Table 5. Comparison of induced rules

Output Variable	WMA	FPA	Expert
Number of rules	97	127	41
Accuracy (%)	0.512	0.445	0.687
Confidence (%)	0.562	0.824	0.624

4. Conclusions

We proposed a decision model using fuzzy rule base to assist insurance advisors in proposing suitable insurance policies under the Family plan based on five factors – age, gender, monthly income, job risk and marital status. An expert with more than 10 years of experience in insurance consultancy was involved in the process of setting up the rule base. Fuzzy logic was applied to transfer the inputs to the fuzzy variables using appropriate membership functions. To validate the model, 50 new policy holder records were used. In future, the accuracy and the completeness of the rule base will be improved to consider persistent, consistent and redundant rules. In addition, the model will be expanded to propose the policy details to guide new and inexperienced insurance advisors.

References

[1] Guillaume S, Designing fuzzy inference systems from data: an interpretability-oriented review, *IEEE Transactions on Fuzzy Systems*; 2001, 9(3): 426-443.
[2] Zadeh, AL. Discussion: probability theory and fuzzy logic are complementary rather than competitive, Technometrics; 1995,37(3):271-276.
[3] Shapiro AF, Fuzzy Logic in Insurance: the first 20 years, Journal of Insurance: Mathematics and Economics; 2004;35:399–424.
[4] Huang CS, Lin YJ, Lin CC. An evaluation model for determining insurance policy using AHP and fuzzy logic: case studies of life and annuity insurances, Proceedings of WSEAS International Conference on Fuzzy Systems; 2007.Vol.8. p. 126-131.
[5] Huang CS, Lin YJ, Lin CC. Determination of insurance policy using a hybrid model of AHP, fuzzy logic and Delphi technique: a case study, WSEAS TRANSACTIONS on COMPUTERS; 2008, 7:(6):660-669.
[6] Lazim Abdullah, Mohd Nordin Abd Rahman. Employee likelihood of purchasing health insurance using fuzzy inference system, International Journal of Computer Science Issues; 2012 vol. 9, issue 1, no. 2, p. 112-116.
[7] Alonso JM., Magdalena L. Generating Understandable and Accurate Fuzzy Rule-Based Systems in a Java Environment, In Lecture Notes in Artificial Intelligence - 9th International Workshop on Fuzzy Logic and Applications, Springer-Verlag, LNAI; 2011;6857:212-219.
[8] Wang LX, Mendel JM. Generating fuzzy rules by learning from examples, IEEE Trans. on Systems, Man and Cybernetics; 1992;22(6):1414–1427.
[9] Delgado M, Gomez-Skarmeta AF, Martin F. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling, IEEE Transactions on Fuzzy Systems; 1992;5(2):223-233.
[10] DeWit, GW, Underwriting and Uncertainty, Insurance: Mathematics and Economics; 1982;1:277–285.
[11] Berenji, HR. Fuzzy Logic Controllers, "*An Introduction to Fuzzy Logic Applications in Intelligent Systems",* Yager RR and Zadeh LA. Boston: Kluwer Academic Publishers; 1992; 69-96.
[12] Chennakesava RA, Fuzzy Logic and Neural Networks: Basic Concepts & Application. New Age International Publishers; ISBN: 8122421822. 2008.
[13] Cox E, O'Hagan M. The Fuzzy Systems Handbook: a Practitioner's Guide to Building, Using, and Maintaining Fuzzy Systems, 2nd ed. CA, USA: Morgan Kaufman; 1998.
[14] Kostas SM, Psarras JE, Askounis DT. GENESYS: an expert system for production scheduling, Industrial Management & Data Systems; 2002;102(6):309 – 317.