International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 16 (2015) pp 36414-36419

PREDICTIVE ANALYSIS OF CUSTOMER PERSPECTIVE ON INSURANCE POLICIES AND MARKET TREND ANALYSIS BASED ON DATAMINING AND LEARNED ALGORITHM.

AKSHATHA PRABHU[1]
Department of Computer Science
Amrita Vishwa Vidyapeetham, Mysuru campus
akshathaprabhu06@gmail.com

MANISHANKAR.S[2]
Department of Computer Science
Amrita Vishwa Vidyapeetham, Mysuru campus
manishankar1988@gmail.com

Abstract

Customer choice identification is the base for any product based company. Market analysis many times does not provide an exact image when product is one like Life insurance.Life insurance company basically has the belief of covering the risk of life of the policy holder. As product is an immaterial one identifying the customer is a hard task , more over prediction will be based on several criteria that changes with market trends , these factors generate a need for decision making system with a self evolving intelligence to identify proper choices from the customers before launching the policy in market. Major aim is to develop a self evolved machine learned business data intelligence system for an Insurance industry .As there are lot of companies ,customers and policies there is a requirement of a cloud based solution to enhance the system capabilities. System helps to portray the market trend analysis model based on customer suggestions and feedback accumulated from the previous product launch.

Keywords - policy prediction , market trend analysis , hidden markov model , customer preference profiling

Introduction

An insurance industry is one which has an abrupt variations based on the market. Numerous policies being launched every year that satisfies the requirements of customers. The customer in an insurance industry always chooses the package suitable to his/her needs. Various insurance policies catering to customer needs are health, life , vehicle , travel etc. Any policy includes a legal agreement between insurer and insured. There will be a term based premium that is cumulated with the coverage amount. Considering customer perspective insurance companies provide benefits such as Risk cover - during unforeseen events , long term investment , tax and retirement benefits , income through dividends and bonus. Customer enjoys the benefits of the covered amount even after the policy period.

The customer invests for the policy based on following metrices such as - The insurance plan perspective of each customer varies largely based on individual's financial background and the cause for undertaking the policy. Policies have their respective contract details, terms and conditions. The policy coverage includes coverage period, premium amount, frequency of premium amount , coverage benefits and return. The coverage is decided entirely on the life style ,investment needs and debts. The premium amount depends on age , policy cover , health status.

Policies are related to multiple prime factors which makes them customer friendly. Identifying these features forms the crux of the market analysis trend. The process identifying the customer's perspective on the policy , company , agent or the middleman involves business intelligence and analytics. There is always a formula connecting the insurer and the insured. Deriving these parameters of this equation is through data analytics tools. Data mining plays a crucial role in the prediction task of the percentage of customer's interest in the policy. This task actually paves a way for the companies to investigate their market risks when policies have been released into the market. Predicting the status of claim amount during policy period and risk coverage is the key task of business analytics that entirely depends on the policy chosen. The client's main focus lies on these attributes. Major percentage of the success of any policy is this feature. Bifurcating policies based on their types and customer view is also an essential feature. Policy document is the matter of attraction which depicts the policies offers ,terminologies

,liabilities and conditions. so while designing a policy document extreme care is being taken by the company to avoid fake promises .there is so much of legal issues that relates to a policy ,those parameters ensures the service agreement .customer is always behind the imagination that there is some deceit in the word of mouth explanation by the agent and the offer document. Customer many a times is not literate enough to read and understand the terminologies in the policy offer document ,or many times they don't give due care. So there is always a requirement for the company that customer builds the trust by making him realize the policy features. This requires a simple automated system supported with language processing and data mining.

Here the data mining task involves the segmentation and analysis of parameters. The data mining task involves mining of associations, correlations , frequent patterns and clusters required for the data set. The knowledge discovery in data analytics comprises of collecting the data from various sources , cleaning the noisy data, integrating , selecting the features , converting the data into appropriate format , identifying the patterns and represent the processed information.

EXISTING SYSTEM

In the world of insurance presently many data are not properly exposed properly. It is only through two carriers that such an issues arise. The insurance company abstract the data with terminologies unfamiliar to public due to their company policy. The middle men who plays a key role in marketing the policies many a times provide fake details and confuse the customer. They also lapse in providing essential details to the questions arising from the client leading to discarding the policy by the client. Agent always anticipates for a better range in commission amount. Client is not aware of the commission rates and ignorant of the swindle done by the agent.

The client has no other option than depending on the agent or the policy document. Many a times he is uninformed of the calculation of premium amount , coverage time , scope of coverage , claim and benefits. The client is put in a painful and dissatisfied position. Disadvantage of existing system is in the lack of identifying essential parameters and survey by the client. An human intervention has always been a cause for dilemma. There is no automated system supporting the requirements of the client in understanding the policy documents. Clients are accustomed to common terminologies and instructions. New clients would rather find it difficult to take up the policy. Constant guidance is essential for the customer to choose the desired policy and to actuate the policy requirement.

PROPOSED SYSTEM

The customer's profile is very crucial in forming the system that helps him understand the policy documents. All the elementary details are considered when the client takes a policy. The system is a combination of data analytics and learning process involved through suggestions and feedback that arise from customer as well as the related policy document. The system contains various attribute such as analysing a customer profile , then using algorithms to analyse customers' need exactly portraying them in a market trade map which plots various factors connecting the customer needs and the various market factors. The system has a customer helping module which tries to give a better understanding of the policy document. By converting the required features which he/she needs to have an idea while investing on the policy. It gives suggestions to the customer on what the best he can get according to the requirement. He is supported with a conversion of policy perspectives into the natural language which the client prefers. The process is related to preferential selection of quality attributes based on suggested result.

The system collects huge amount of data from the existing customer based review and suggestion portal which is built on top of real world situations. If the client registers in the system for the first time then the system asks for detailed features of the client and if on the other hand has already been registered then it simply allows the client to enter the system. In the situation when the client registers to the system for the initial time , he/she undergoes through a list of questionnaires. Along with the elementary details of name, age , address it is supplemented with key aspects such as qualification , frequency of policy usage , knowledge level in policies and the highlighting aspect that the client is searching for in the policy. If the qualification field of client is basic then the system has to undergo a rigid process in providing details of the policy. For a client who registers for the first time then high level is set in providing knowledge base for the client. Key terms that the client seeks out in the policy document is also

specified in the system. The language processing technique adopted provides synonyms to the words that are difficult to understand by the client. The discriminative methods make use of posteriori probabilities based on observations. With policy type being the index key , the user preferences are mapped into a table. The columns of the table are associated with parameters and relations are established based on the suggestions provided by the user on previous parameters. Each user has a set of keywords in white list specifying positive preferences and in black list specifying negative preferences. All these would be stored when the client is entering the highlighting aspects of the policy document in the system.

Each time when a policy is added into the users list, policy will be searched for the keywords and language processing algorithm is fed with the lists. The result of this language process model is taken as an input for knowledge mining task which helps to portray a model of user policy preferences. This procedure can be base model and learning model for companies in order to predict the success of a policy in market in coherence with the customer perspective chart.

SYSTEM ARCHITECTURE

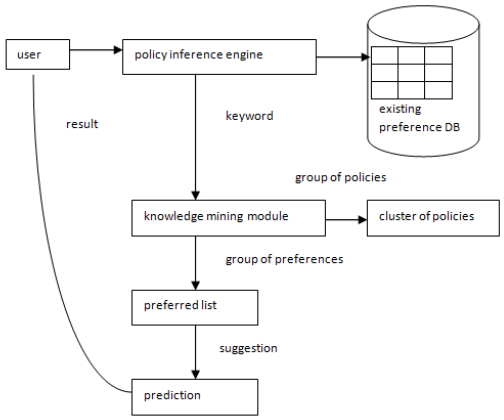


Fig 1 : Architecture of customer and policy relation management system

The diagram provides a pictorial representation of the user in direct relation with the policies that are accepeted or discarded.User is supported with a system which helps to choose the best deal available for him/her in the policy market.The user provides the essential keywords according to his/her preferences initally while registering into the system.These are the guidance factors for the inference engine to map the preferred keywords and policy keywords into appropriate vector space. The preferred keywords are categorised as black list and white list where black list is the least preferred keywords and white list is the most preferred keywords.The matched data from entire preferrential list is processed fed into knowledge mining module.Knowledge mining module encompasses data characterization definition fed into it.It includes the definiton of target class and discrimination categorises the policies based on the type.Preferred list is generated from a group of preferences provides a list of suggestions on the list of policies that is preferred by the user based on the key features input by him/her.These propositions are finally used as prediction to choose among the list of policies by the client.Appropriate policy is chosen by the prediction module.Client gets a set of policies that initiates in easy decision making capability.

METHOOLOGY

ALGORITHM

State based algorithm for generating policy preferences

The processing is defined in terms of a function defined as F(S0,S,Fi,T,G,Lo,L1,N).

S0 is the initial state
S is the set of all states
Fi gives the set of final states
T is the transition set from one state to the other
G is the grammar defined.
Lo language containing user preference keywords
L1 language containing policy list words
N -number of states required for reaching final state

State Based Algorithm

1. Start
2. Fmax is initialised to maximum limit
3. do
4. process initialize the user state to S0 ,where user selects the preferred language to be operated on L0
5. L0={ x€L0,/ The keyword is provided }
6. x={can be white list keywords or blacklist keywords}
7. language L0 has to be operated by the specific transitions T in a grammar G
8. G is a context free grammar which works on the languages which has capacity to accept string of keywords
9. policy material is fed in to language L1 with the help of grammar by moving a transition T if the keyword rule for grammar keyword G is matched.

10. G has a set of rules with common transitions from L1 and L0 moving to set of Final states Fi.
11. Increment N when there is a walk to the next state.
12. if the keyword doesn't match the value then it enters the trap state .
13. If (trap state) , it denotes the unsuccessful keyword synonym.
14. until(N < Fmax).

The algorithm is based on a finite state machine which is an integral part of inference engine. It actually takes the input from the user or the policy with the help of language processing. The policy keywords are generated using Hidden Markov Model. The system is previously fed with the several factors - premium amount , coverage time , scope of coverage , claim and benefits etc. These attributes are represented as transition variables that identifies a state. Each state has a probability distribution over the possible outputs.

Each valid document is mapped in to a state space of markov model ,where there is numerous values present with varying states. The current values are always depend on the previous value obtained through a transition probability. Thus an matrix is obtained which illustrate the transition probabilities the process involved here is to compute parameters and their output sequence ,which is obtained through summation of all possible state sequence.
probability of states are represented as
Y=y(0),y(1),.....y(L-1);
L is the length of sequence .
The probability of states is given by

$$P(Y) = \sum_{x} P(Y|X)P(X)$$

where P(Y) is the probability of the sequence state y depending on present state x.
Hidden node sequences are specified as
X=x(0),x(1).....x(L-1)
computation of the model parameters over the distribution of the variables of predefined model is given p(x(t)|y(1),y(2),....y(L-1) , the process continues with corresponds observation of state at each time..this helps to filter the necessary parameters .
The process above depicts the mathematical model involved in the process of identifying the key parameters of the policy document on the basis of the various sequences of transition state .A state chart sequence diagram is developed on the key word sequence and their possible states are recorded.

These output sequence of keyword obtained are passed as policy document keyword values to be cross checked with user preference list obtained with help of state chart diagram. These policy document keyword values mapping with user preference list are passed on as parameters to the knowledge mining module .this module is meant for the segregation of policy documents in to beneficial list and non beneficial list ,for this governing we are utilising the method of data mining. specifically a learning module based on clustering of policy documents helping for a better prediction task for the customer and company about the success rate of the policy. Here the learning algorithm is based on the distance based approach where the distance from the centroid of cluster is distance calculation based algorithm. initial dataset for the knowledge mining is formed with the data collected from user and inference engine ,collectively taking the apt parameters gives us an option to predict or cluster with a better accuracy

Algorithm -Distance based Clustering

Step 1: start
Step 2: intialize the parameters opted for clustering
Step3: load the dataset D obtained in the data mining module
$D=\{X_1,X_2,X_{3....}X_n\}$
where X1,X2....Xn are the parameter values
Step4: Pre-process the data to avoid null values.
Step 5: Calculate the mean for the dataset D
$\sum X_i = X_1+X_2+....+X_n$
Mean $M=\sum Xi/N$
N is the total number of values present in the data record.
Step6:populate the thresholds necessary for calculating centroid G.
Step 7: points are plotted in to the vector space with a distance calculated from centroids
Step 8:each centroid forms a group around it with in a threshold distance TH_{max}
Step 9: If any attribute lies beyond TH_{max} it is added to the outlier list for further comparisons.
Step 10: new dataset are also explicit plotted and the new clusters are formed
step 11.stop.
.
The clustering algorithm helps in the prediction task. outliers are used in the data set to find the varying predictions in user preference which might be of use in deriving a conclusion of accessing a policy success rate.
The task of prediction is a relative one in which policies best matching to user preferences mapped with cluster and inference engine is passed to the user. The prediction task is an instance based one which involves considering each policy document and specific instance of user preferences.

IMPLEMENTATION DETAILS

The system is implemented with help of web application. There is a login form which has a usual login process. User name and password are provided. If the user is new to the system then he/she has to be registered for the application else

the user can directly enter the system. At the time of clients profile building, user preferences are taken into account with the help of a questionnaire. The user is asked to specify the needs accurately so that his preferences can be correctly taken in to account. User fills the profile , his policy wish list and keywords.

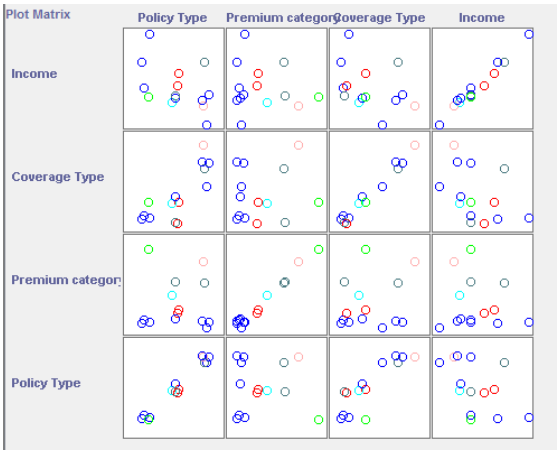
His preferences are fed in to the inference engine for the processing of the formation of black list and white list .policy agents as well as company employees are also taken as users of system. They are responsible for adding the policy details and other information necessary for launching the policy in to market. This is a part of market trend analysis that is essential in predicting the success of the policy.

The application is also responsible for generating policy user preference chart. Each policy is taken into consideration and if user preferences are tallied with the keywords of the policy, the policy is passed to the inference engine to be processed and then recommended to customer.

The user preferences are taken in a form which has multiple fields to support the inference engine in the process of creating a table that stores all these attributes that are input through the windows application. Application maps the preference list with policy list and generates the mapped table. This map table is the resultant data essential for the knowledge mining module ,this application has very important role because it pre-processes the dataset and eliminates those preferences which are not matching in both list. Thus achieving an efficient keyword list.

For the reference of the paper and for generating test cases we have considered a dataset of LIC life insurance policies in karnataka. The parameters considered as user preferences in knowledge mining are

NA-name of the insured person
AG-age of the insured person
PT - Policy Type
NP - Number of policies
PC - Premium Category
PA-Premium Amount
CT-Coverage Type
CP-Coverage Period
CB- Cash back
RG-Region
FS-Family status
IN-Income
ORG-Organization
NOM-Nominee
NB-Number of members
GEN-Gender
MEDI-Medical History



CONCLUSION AND FUTURE SCOPE

The system proposed provides an efficient methodology to choose policy based on user preferences. It supports the company as a decision support system analysing the market trend analysis through acceptance and rejection of policies. Here business intelligence and data mining with a learned approach is being associated. Hidden markov model assists in providing an effective clusterisation of the data sets. Here the wide variety of customer needs are portrayed in a parametric model and studied for an inference engine and knowledge mining. The system can be modeled in future to fetch policy preference in natural language in order to help customer improve his investment towards policies.

REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A. Data Mining: A performance Perspective. IEEE Trans. Knowledge and Data Engineering, vol, 5,6, 1993a, pp. 914-925.

[2] Abdhesh Gupta, Anwiti Jain ,Life Insurance Recommender System Based on Association Rule Mining and Dual Clustering Method for Solving Cold-Start Problem , IJARCSSE , Volume 3, Issue 1olicies.0, October 2013

[3] A.B. Devale, Dr. R. V. Kulkarni Applications of Data Mining Techniques in Life Insurance. International Journal of Data Mining and Knowledge Management Process Vol.2, No.4 July 2012. 1.

[4] Maria N. Moreno, Saddys Segrera, Vivian F Lopez, Maria Dolores Munoz and Angel Luis Sanchez, Mining Semantic Data for Solving First-rater and Cold-start Problems in Recommender system ACM IDEAS 11 2011, September 21-23.

[5] Dr. Arnika Srivastava, Dr. Sarika Tripathi, Dr. Amit Kumar,indian life insurance industry – the changing trend.

[6] Lee, CH., Kim, Y.H., Rhee, P.K. 2001. Web personalization expert with combining collaborative filtering and association rule Mining Technique. Expert System with Applications 21. 131-137.

[7] Sarwar, B., Karypis, G., Konstan, J., Riedl, J. 2001. Item-based Collaborative Filtering Recommendation Algorithm. Proceedings of the tenth International World Wide Web Conference, 285-295.

[8] Resnick, P., Lacovou, N., Suchack, M., Bergstrom, P. and Riedi, J. 1994. Grouplens: An open architecture for collaborative filtering of netnews. Proc of ACM Conference on Computer Supported Cooperative Work, 175-186.

[9] Schafer, J.B., Konstant, J.A. and Riedl, J. 2001. E-Commerce Recommendation Applications. Data Mining and knowledge Discovery, 5, 115-153.

[10] Ran Qu ,Zhenting Qu,Notice of Retraction Interaction between Life Insurance Fund Investment and Financial Market in China: General Investment Strategies for Life Insurance Companies.

[11] Yanfang Liu, "Investment and surpervior of life insurance," Beijing, Chinese Light Industry Press, 2001: 177-183.

[12] Shaoyi Meng, "International Comparation of Insurace Fund Appilication," Beijing, Chinese Financial Press, 2005: 10-13.

[13] Hao Pang, "Econometrics," Chengdu, South Western University of Finance and Economics Press, 2002: 198-205.

[14] Granger. Testing for causality: a personal view point[J]. Journal of Economic Dynamics and Control, 1980, 2: 329-352.

Recieved MS degree in Computer Science and Technology from the University of Mysore, pursuing her research under Bharathiyar University. She is currently working as Lecturer in Department of Computer Science, Amrita Vishwa Vidyapeetham, Mysore Campus. Has an excellent teaching profile and proficiency in guiding students towards research oriented projects. Area of specialization includes Data analytics, BigData ,ERP solutions, Network Security.

Mr Manishankar.S recieved M.Tech degree from SRM University , currently pursuing his doctral degree under Bharatiyar University , a passionate researcher in the field of Big data and Analytics , Currently designated with Assistant Professor at Amrita Vishwa Vidyapeetham , Mysore has a collection of publications and International Conferences to add to his academia. He has steered Postgraduate and Undergraduate students in project thesis and journal publications. His major area of research contributions are Cloud Computing, Distributed Systems and Network Security. His profile is a blend of industry and academia.