

Online Insurance Consumer Targeting and Lifetime Value Evaluation - A Mathematics and Data Mining Approach

Yuanya Li^{1,2}, Gail Cook³ and Oliver Wreford³

¹*River and Harbor Department, Nanjing Hydraulic Research Institute, Nanjing, 210024,*

²*Key Laboratory of Port, waterway & sediment engineering, Ministry of Communications,*

³*InsWeb Corp, 11290 Pyrites Way, Suite 200, Gold River, CA95670,*

^{1,2}*China*

³*USA*

1. Introduction

InsWeb Corporation provides an online insurance marketplace for consumer and insurance companies. The provided insurance products for shopping include automobile, term life and homeowners insurance, as well as annuities. The company business can be classified as two categories: lead purchase and policy purchase. In lead purchase consumer comes to InsWeb insurance online marketplace, compares quotes based on his/her personal information and products provided by insurance carriers, and submits lead to carriers. After that as a marketplace provider InsWeb doesn't involve in following purchase process. It is the issue between consumer and insurance carriers. In this case InsWeb involving the purchase process time is in minutes. In policy purchase category, after consumer submitting lead, the agent of InsWeb will contact the consumer and make the effort to sell the policy to the consumer. After the consumer purchase the policy, InsWeb has the opportunities to renew the policy in annual or half year period each time. For policy purchase consumer, InsWeb has the chance to involve the purchase process from year to decades. Therefore InsWeb tries to improve service for consumer policy purchase and convert more insurance lead to policy to generate more revenue. Since the revenue generating from a policy purchase is much higher than lead purchase from a lead shopping session, InsWeb has to find out the space to improve the lead close rate and retention rate.

For the quarter ended March 31, 2001, InsWeb "records more than 3.5 million unique user sessions and approximately 660,000 completed shopping sessions during the first quarter"[1]. And "InsWeb's insurance agency sold nearly 3,000 new policies during the quarter"[1]. Now "the Company's agency will expand to five additional heavily populated states with existing and new insurance companies. Further expansions are planned for the remainder of the year"[1]. To convert more shopping session into the policies sold by InsWeb insurance agency, adopt the expansion of InsWeb Corporation business, satisfy the immediate auto policy insurance purchase requirement of the internet consumers, provide high quality service to consumers and increase the policy sold number of The InsWeb's agency, the data warehouse group try to find out the internet consumer behavior and direct InsWeb agency to targeted consumers at proper time and in proper way.

Fig. 1. The start internet page of online Auto insurance

2. The consumer information and its internet behavior collection

As an example, Figure 1 shows a lead shopping session flow chart consisted with different web pages named as Start, Drivers, Vehicles, Coverages, Profile and Companies. The work flow is that online consumers put in information of start, drivers, cars, coverage. The third party databases will provide consumer credit history and auto insurance history record through internet simultaneously with consumer personal information put in. With the data, the auto insurance evaluation agents of insurance companies will provide the quotes for the consumer. All the quotes from different insurance companies will be listed on profile page. When a consumer can choose a quote, a lead is generated. The lead has two ways to go: an insurance agent of Insweb or an insurance company. If the lead is sent to an insurance company, a lead shopping session is completed. If the lead is sent to an insurance agent of Insweb, the agent will contact the consumer with phone in proper time, confirm some information and purchase purpose, then an auto insurance shopping session is completed. All the information a consumer puts in and third parties provide will be stored in Insweb Data Warehouse. To record consumer behavior, InsWeb Data Warehouse invented the event log technology and got a patent in 1999, by which consumer behavior on website like turning back page, jumping from one page to another page and timestamp for each step will be recorded in Data Warehouse. Each year more than ten millions of internet consumer information have been accumulated into Insweb Data Warehouse. Those data include the

detail of consumer personal information, such as age, gender, marriage, marriage history, job type, number of children, owner of house, house type, car owner, and car information like maker, car model so on.. All the static data of consumers and cars with dynamic data of consumer web behavior provide the base for the statistics analysis in data mining.

3. Tables, columns and random variables

The tables used in this analysis are Agency_leads, Agency_policy, Auto_Consumers, Auto_Drivers, Auto_quotes and other tables with Auto prefix in data warehouse. In the tables, Agency_leads table records daily the information of consumers who submitted the leads and the leads select the quotes from carriers which InsWeb E_agency covers. All the consumers in Agency_Leads are the potential InsWeb E_agency policy buyer.

InsWeb agency will contact the consumer in the table and hope to sell the auto insurance policy to them. Agency_policy table records the information of consumers who purchased the auto insurance policy from InsWeb agency. Of course only part of consumers record in Agency_Leads would buy auto policy from InsWeb agency, otherwise this analysis isn't necessary. Therefore, the number of consumer in Agency_Policy is always less than the number of consumer in Agency_Leads.

Auto_Consumers, Auto_Drivers, Auto_Quots and other tables with auto as prefix contain all the information of consumer who got to InsWeb web page and put in some personal information such as age, gender, birthday, old auto insurance policy expiration day, new auto insurance effective day, so on.....By joining Agency_policy table with Auto_Consumers, Auto_Drivers, Auto_Quots, and other tables with Auto as prefix, we can get the background information of consumers who purchased Auto policy from InsWeb agency. All the background information will be used to determine the consumer motivation to purchase auto policy.

As the comparison, the Agency_Leads table will be used to join with the same tables as Agency_Policy does to get the same information. The relative information will be used as comparison stadium.

In InsWeb data warehouse, all information is saved into tables. Each table has columns which number varies from two to more than one hundred. Each column records one attribute of consumer. From analysis and mathematics view, each column is a random variable. As a random variable, each column has its value region. Due to the data types of columns are different, from integer, varchar, numerical,, to date, the value regions of random variables are complicate, from simple (YES,NO), a list like Carrier (Travelers, Hartford Agency, Atanta Casualty, CSE, Explorer, FIC, GMAC, Great American, Infinity,...), age (16, 17, 18, 19,...,65) and some theoretical continuous distribute value region like old_policy_expiration_days $(-\infty, +\infty)$ (Notes: human being life is very limited, but the data consumer input would be arbitrarily distributed, not always be reasonable).

4. How to evaluate the potential auto policy purchase consumers

Some facts to determining lead closure rate are presented below in the curve figure. Psychologically, the human behavior vary with age, gender, house owner, and so on, therefore, the correlation coefficients are the functions of these facts. For the comparison purpose, the Agency_Lead and Agency_Policy tables will join with other auto tables. The all

background attributes of each auto lead consumer and auto policy consumer will be listed in a big table for analysis purpose. For more than 10,000,000 auto consumer session, the statistic results are listed below:

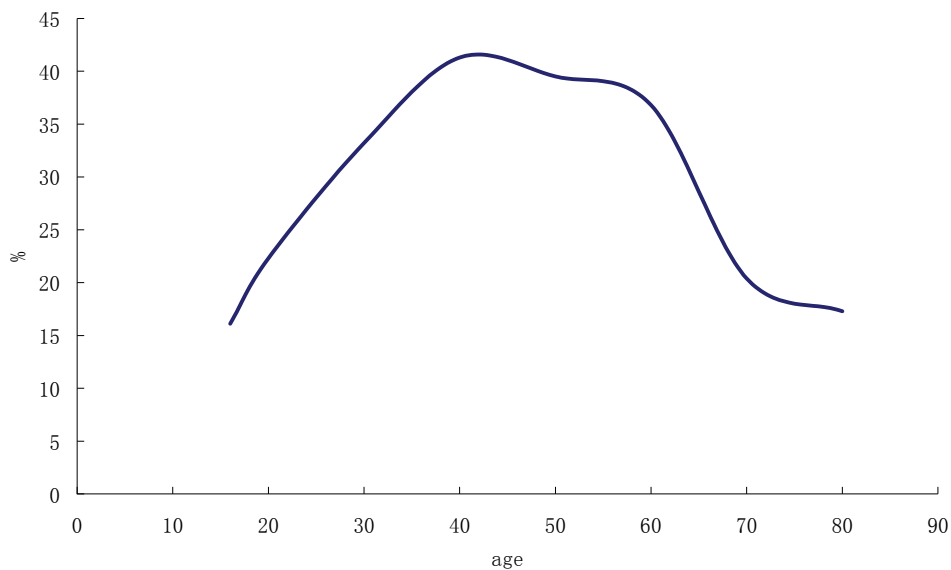


Fig. 2. The variation of consumer lead closure rate with age for female

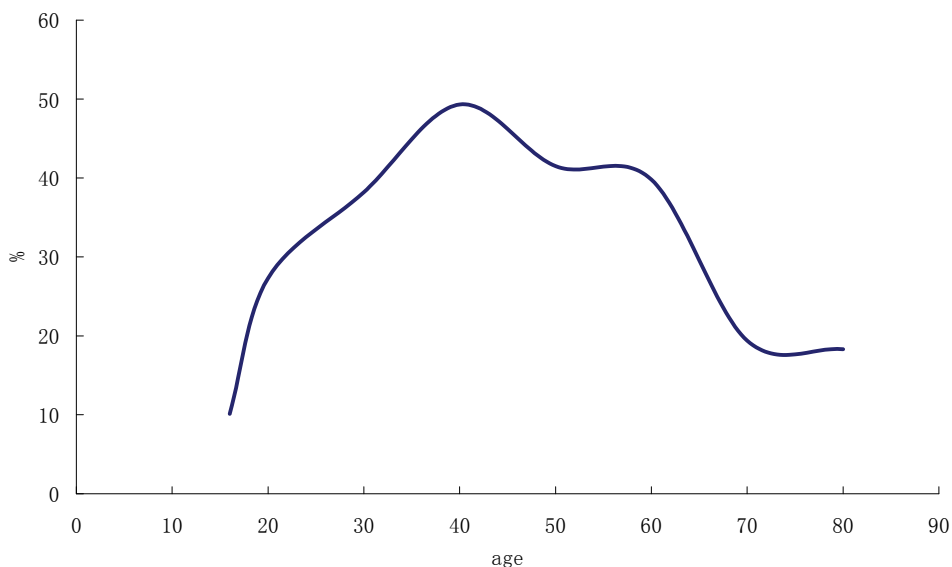


Fig. 3. The variation of consumer lead closure rate with age for male

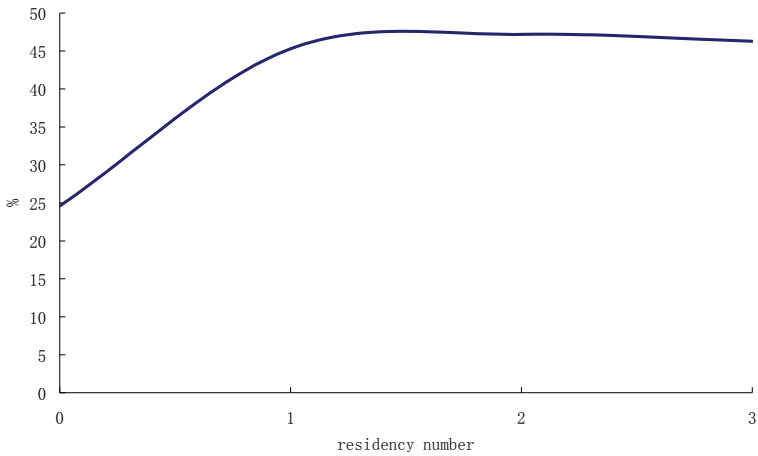


Fig. 4. The variation of consumer lead closure rate with residence_own number

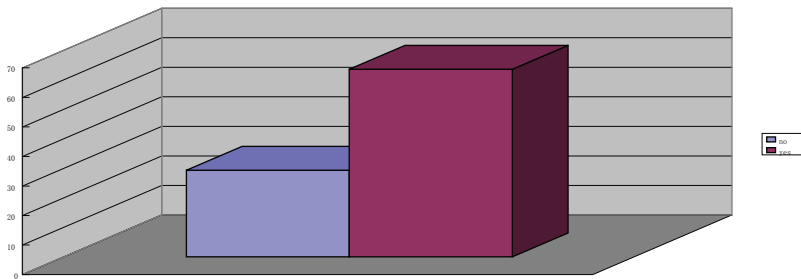


Fig. 5. The variation of consumer lead closure rate of new_policy_effective within3day random variable

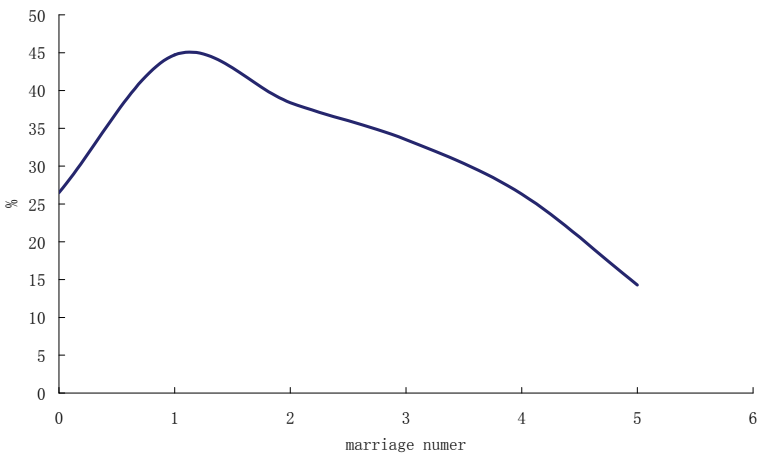


Fig. 6. The variation of consumer lead closure rate of married random variable

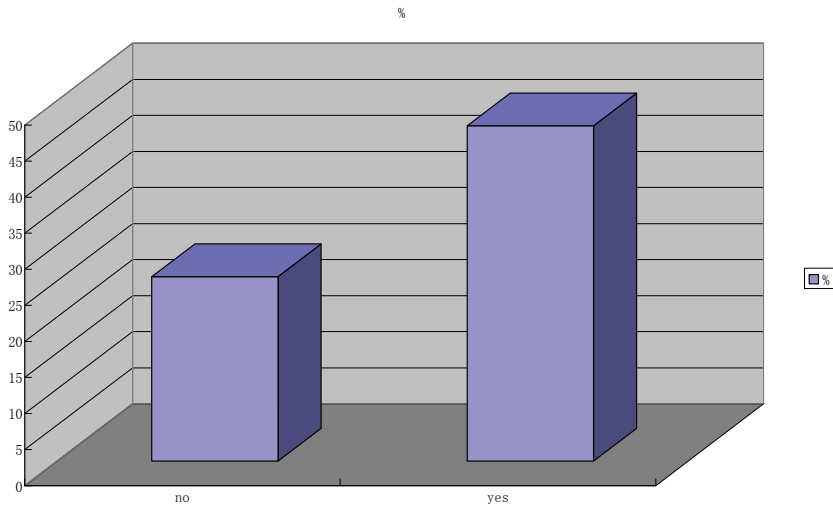


Fig. 7. The variation of consumer lead closure rate of old_policy_expiration_within_3day random variable

Figure 2 to 7 outlines the main fact for lead consumers to determine purchasing the insurance policy. Figure 2 and 3 indicates that the lead closure rate variation with age has the same trend distribution for male and female, but male has higher lead closure rate compared with female. Figure 4 shows that the lead consumers owning residencies has more lead closure rate than that not owning residency. The lead closure rate for marriage status is the person with one marriage. The consumers, completing leads that old auto insurance would be expired and new have to be effective in three day, have more potential to purchase auto insurance policy. The analysis will help auto insurance agency to identify the auto insurance lead consumer with most like hood to purchase the auto insurance policy from the huge insurance lead submitted through internet.

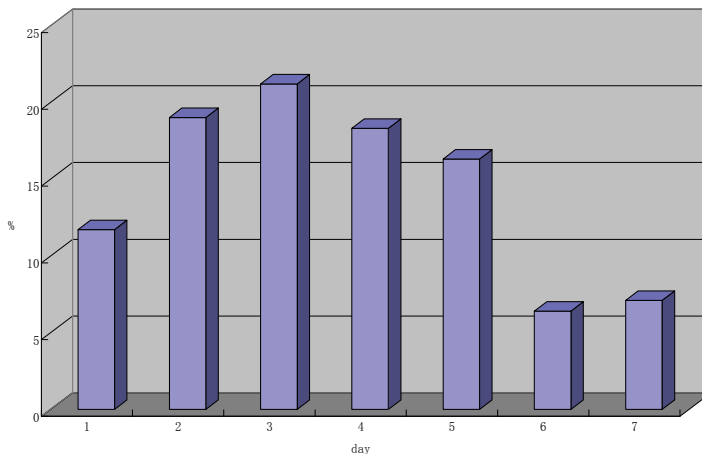


Fig. 8. The consumer internet session distribution in a week

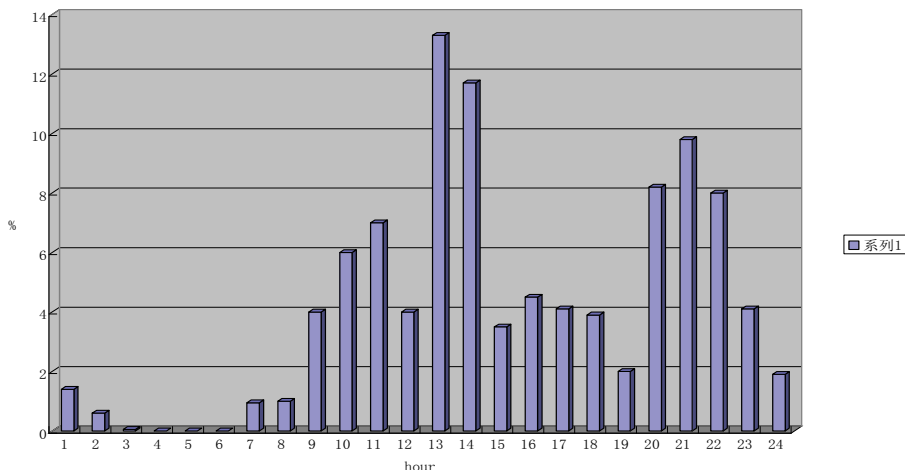


Fig. 9. The auto insurance consumer lead closure rate vary with time in week day

5. The consumer behavior variation on internet with time

Other interesting issue is that consumer session distribution on internet varies with time in week period from Monday to Sunday. When is the good time for effective consumer to submit lead, and the auto insurance policy purchaser to go to the website.

Figure 8 is the auto insurance internet session distribution in a week. From the figure, the consumers internet session is higher in week day and lower in weekend. The Monday is lowest in week day and the Friday is the second lowest in week day. The possible reason for this kind of distribution is that in weekend more people does outdoor activities and has less time going to internet, and in Monday the potential consumers have to pay more attention for the work and spend more time for weekend plan in Friday. Figure 9 is the lead closure rate distribution with a day time which is accounted in hours. The figure shows that there are two peak time for auto insurance consumers to submit the lead, one is the lunch time and other period is from 20:00 to 22:00 in the nigh.

6. Fuzzy theory for determine online auto insurance consumers

Based on the analysis above, a fuzzy formula is obtained to determine the possibility for online auto insurance consumers who would buy the auto insurance policy. The fuzzy function for all auto insurance leads is:

$$P = \sum_1^7 w_i p_i \tag{1}$$

In the formula above, p_i is the value for each random variable for figure 2,3,4,5,6,7,8, and 9. Its value is simulated by y / y_{peak} , the weighting value w_i is determined by data mining and auto insurance agent feeling. In this way, all the online auto insurance session is evaluated automatically and the insurance agents can contact the consumers in the order the fuzzy formula provides. The experience shows that the fuzzy formula works well.

7. The mathematic expression of customer lifetime value (lead)

Each time a customer comes to InsWeb website, put its personal information and wonders in the website, all the personal information and customer behavior will be recorded in InsWeb data warehouse. Different customers have different lifetime value for InsWeb. InsWeb needs to determine each customer commercial from customer personal information and her/his behavior on website by data mining. A customer lifetime value for a lead can be written as

$$S = P * V \quad (2)$$

Where S is the customer lifetime value (Lead); P is the likelihood of a lead becoming a policy and V is the customer lifetime value of the policy.

To make the mathematic formula developing process clearly, this situation can be considered: number of M customers with the same personal information and background come to InsWeb website and submit agent leads. Since the decision to purchase policy from InsWeb or not is random variable for each customer, assume M_0 purchasing policy from InsWeb and $(M - M_0)$ customers don't. After half or one year, M_1 customers renew the policy and $(M_0 - M_1)$ customers don't. When the policies expiration, M_2 customers renew their policies and $(M_1 - M_2)$ customers don't. Assuming the current age of those customers is T_1 and the age those customers give up to purchase auto insurance policy is T_2 , the possible time period for those customers to purchase auto insurance policy from InsWeb is

$$T = T_2 - T_1 \quad (3)$$

In T years the consumers renew their policies N times. So the process for consumers to make decision to renew their policies or not will continue N times. It is a typical N steps random walk problem. Therefore in the final step, there M_n customers will renew their policies and $(M_{n-1} - M_n)$ don't. The revenue InsWeb obtained from those M customers in their lifetime will be

$$S_{total} = C(M_0Q_0 + M_1Q_1 + M_2Q_2 + \dots + M_nQ_n) \quad (4)$$

In which C is the commission rate, and Q_i is the auto insurance quote. Since the income and spending would change for the customers with age, the car type and insurance quote would also change with age. To each customer in the whole purchase and renew process, the revenue InsWeb obtained would be

$$S = \frac{S_{total}}{M} = C \left(\frac{M_0}{M} Q_0 + \frac{M_1}{M} Q_1 + \frac{M_2}{M} Q_2 + \dots + \frac{M_n}{M} Q_n \right) \quad (5)$$

To make equation 4 more meaningful, it can be rewritten as

$$S = C \left(\frac{M_0}{M} Q_0 + \frac{M_0}{M} \frac{M_1}{M_0} Q_1 + \frac{M_0}{M} \frac{M_1}{M_0} \frac{M_2}{M_1} Q_2 + \dots + \frac{M_0}{M} \frac{M_1}{M_0} \frac{M_2}{M_1} \dots \frac{M_n}{M_{n-1}} Q_n \right) \quad (6)$$

let

$$P = \frac{M_0}{M} \tag{7}$$

$$R_1 = \frac{M_1}{M_0} \tag{8}$$

$$R_2 = \frac{M_2}{M_1} \tag{9}$$

.....

$$R_n = \frac{M_n}{M_{n-1}} \tag{10}$$

Obviously P is the probability that customer purchases the auto insurance from InsWeb agent; R_1 is the first year retention rate; R_2 is the second year retention rate and R_n is N year retention rate. Therefore the equation (6) can be written as

$$S = P * C * \left(Q_0 + R_1 Q_1 + R_1 R_2 Q_2 + R_1 R_2 R_3 Q_3 + \dots + R_1 R_2 R_3 \dots R_n Q_n \right) \tag{11}$$

By using the mathematics notation, equation (11) can be written as

$$S = P * C * \sum_{i=0}^n \left[Q_i \prod_{j=0}^i R_j \right] \tag{12}$$

In which R_0 is 1.0. Equation (12) is the general equation to evaluate auto insurance customer Value.

8. The simplified evaluating function for lifetime online insurance consumer policy value

In equation (12), commission rate C is constant which is determined by the deal between InsWeb and auto insurance carriers and can be considered as constant in the analysis process. Q_0 is the quote from carrier based on the consumer personal information. Therefore, the variables needed to be determined by analysis and data mining analysis are P, R_1, R_2, \dots, R_n , and Q_1, Q_2, \dots, Q_n . Due to the short history of data record in InsWeb data warehouse, no enough data to determine R_2, R_3, \dots, R_n , and Q_2, Q_3, \dots, Q_n currently. Some assumptions are taken as below:

$$R_2 = R_3 = R_4 \dots = R_n = R_1 \tag{13}$$

and

$$Q_i = (1+r)^i Q_0 \tag{14}$$

In which r is the quote increasing rate annually. Therefore equation (11) can be rewritten as

$$S = P * C \left(\begin{array}{l} Q_0 + R_1 Q_0 (1+r) + \\ R_1^2 Q_0 (1+r)^2 + \dots + R_1^n Q_0 (1+r)^n \end{array} \right) \quad (15)$$

When $R_1(1+r) = 1.0$, obviously

$$S = N * P * C * Q_0 \quad (16)$$

Multiply both sides of equation (15) with $R_1(1+r)$ and obtain

$$R_1 S(1+r) = P * C * Q_0 \left(\begin{array}{l} R_1(1+r) + R_1^2(1+r)^2 \\ + R_1^3(1+r)^3 + \dots + \\ R_1^{n+1}(1+r)^{n+1} \end{array} \right) \quad (17)$$

Equation (15) minuses equation (16), the result is

$$S - R_1 S(1+r) = P * C * Q_0 \left(1 - R_1^{n+1}(1+r)^{n+1} \right) \quad (18)$$

When $R_1(1+r) \neq 1.0$, finally, the simplified estimating function for insurance policy value is

$$S = \frac{P * C * Q_0 \left(1 - R_1^{n+1}(1+r)^{n+1} \right)}{1 - R_1(1+r)} \quad (19)$$

Equation combining (16) and (19) is the simplified lifetime value evaluation function for InsWeb online insurance policy consumer.

9. Data mining result for evaluating function of auto insurance customer value

In equation (19), r is considered as independent from consumer personal information. It is increased by insurance carriers annually. For example. State Farm and All State just increased the quote for auto insurance 2.7 last year 8.3% this year[1][2]. Based on that,

$$r = 2\% \quad (20)$$

is assumed for long time period for all quotes. For all consumers, the lead close rate P is the function of quote Q_0 and consumer personal information such as prior policy expiration day, residence year, residence month, homeownership, gender, marital status, source, state and others, P its can be written as

$$P = f_1 \left(\begin{matrix} Q_0, age, gender, expiration_day, \\ resident_year, resident_month, \\ marital_status, state, source, \\ homeownership, \dots \end{matrix} \right) \quad (21)$$

In the same way, R_0 is

$$R_0 = f \left(\begin{matrix} Q_2, age, gender, marital_status, \\ state, source, homeowner, \dots \end{matrix} \right) \quad (22)$$

P and R_0 can be determined by decision tree technology of data mining from InsWeb data warehouse data.

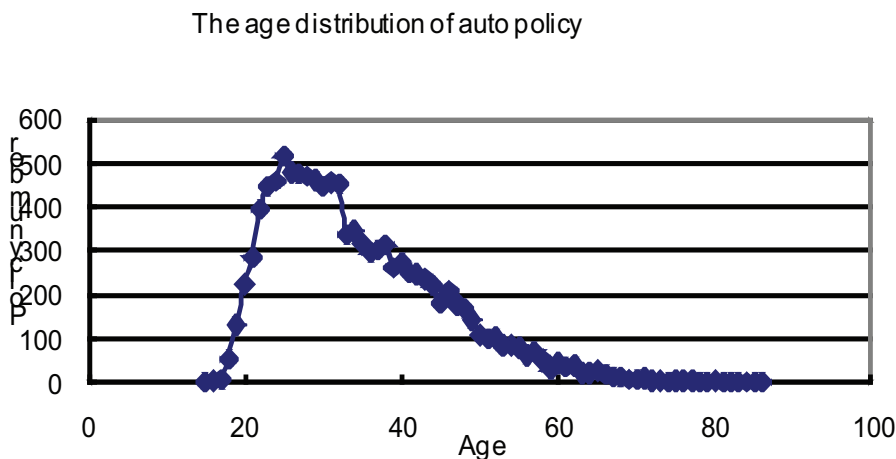


Fig. 10. The age distribution of auto policy

Another variable is N , the total times of retention for a consumer. It is the function of consumer age. To simplify the analysis, all are adjusted to annual based and the total policy renew number is the difference the consumer age to give up driving and current age, so the N can be reasonably written,

$$N = age_s - age \quad (23)$$

In which age_s is the age for consumer to stop driving and purchasing auto insurance age and age is the current age.

The age distribution of agent lead

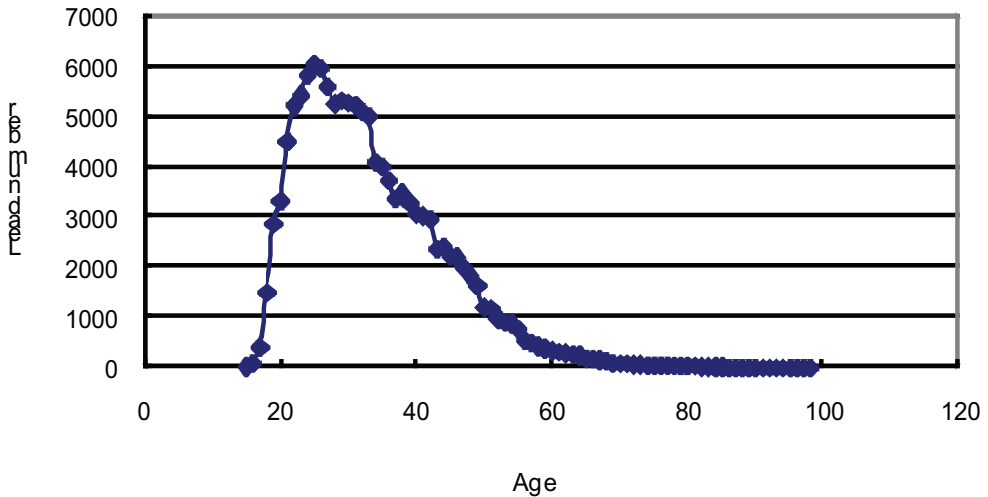


Fig. 11. The age distribution of auto agent leads

From statistic data of www.census.gov[3], the expectation of life for USA citizen is about 71 for male and 78 for female. From Figure 1 and Figure 2, there is deeply decreasing from age 62 to age. Therefore,

$$age_s = 63 \quad (24)$$

For the consumer who age is larger than 63, N is considered as zero. The average age for all policies is 34 year. The average age for population in U.S.A is 36 year [2]. It indicates that young persons trend to purchase online policy. The N can be calculated out as

$$N = 63 - 34 = 29 \quad (25)$$

It is a big task to determine equations (21) and (22). From data mining, the average lead close rate is

$$\bar{P} = 10\% \quad (26)$$

and the average retention rate is

$$\bar{R}_o = 75\% \quad (27)$$

All is based on data of 2001 year.

10. The sensitivity analysis of evaluating function of online insurance policy

From “InsWeb Reports Fourth Quarter and Year End Financial results” for 2001[4], the auto insurance agency revenue for 2001 is \$1634000. That means

$$\sum \left(N_{2001} * \bar{P} * C * \bar{Q}_{2001} \right) = \$1634000 \tag{28}$$

The total revenue generated from those policies is

$$\begin{aligned} S_{total} &= \sum \frac{P * C * Q_0 \left(1 - R_1^{n+1} (1+r)^{n+1} \right)}{1 - R_1 (1+r)} \\ &= \$1634000 \times \frac{\left(1 - (0.75 \times 1.02)^{30} \right)}{1 - 0.75 \times 1.02} \\ &= \$1634000 \times 4.254 = \$6950942 \end{aligned} \tag{29}$$

Equation (28) indicate that with renewing the policies, InsWeb would get additional 3.254 times revenue that is obtained from first policy purchase from InsWeb agent.

To analysis which direction is the best direction to increase InsWeb revenue, the parameters sensitive analysis is made as below:

$$\frac{\partial S_{total}}{\partial P} = \frac{C * Q_0 \left(1 - R_1^{n+1} (1+r)^{n+1} \right)}{1 - R_1 (1+r)} \tag{30}$$

$$\frac{\partial S_{total}}{\partial R_1} = \frac{P * C * Q_0 \left(1 - R_1^n (1+r)^{n+1} \right)}{1 - R_1 (1+r)} \tag{31}$$

$$\frac{\partial S_{total}}{\partial n} = \frac{P * C * Q_0 (n+1) R_1^n (1+r)^n}{1 - R_1 (1+r)} \tag{32}$$

$$\frac{\frac{\partial S_{total}}{\partial P}}{S_{total}} = \frac{1}{P} \tag{33}$$

$$\frac{\frac{\partial S_{total}}{\partial R_1}}{S_{total}} = \frac{1 - R_1^n (1+r)^{n+1}}{1 - R_1^{n+1} (1+r)^{n+1}} \tag{34}$$

$$\frac{\frac{\partial S_{total}}{\partial n}}{S_{total}} = \frac{(n+1) R_1^n (1+r)^{n+1}}{1 - R_1^{n+1} (1+r)^{n+1}} \tag{35}$$

Calculating equation (33), (34) and (35), the corresponding results are shown in tables (1), (2) and (3)

P	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$\frac{1}{P}$	10	6.7	5	4	3.33	2.86	2.5	2.22	2

Table 1. The sensitivity analysis for lead close rate P

The relative gain coefficient for lead close rate

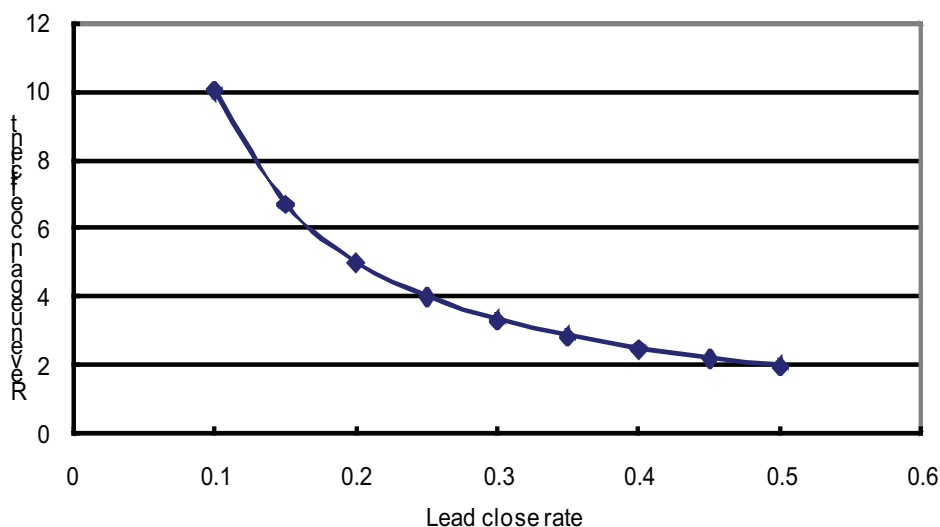


Fig. 12. The relative gain coefficient for lead close rate

R_1	0.75	0.8	0.85	0.9	0.95	1.0
$\frac{1 - R_1^n (1+r)^{n+1}}{1 - R_1^{n+1} (1+r)^{n+1}}$	1.0	1.0	1.0	1.0	1.0	1.0

Table 2. The sensitivity analysis for lead close rate R_1

n	29	30	31	32	33	34	35
$\frac{(n+1)R_1^n (1+r)^n}{1 - R_1^{n+1} (1+r)^{n+1}}$	0.012	0.01	0.0079	0.0062	0.0049	0.0039	0.003

Table 3. The sensitivity analysis for policy renew number N

The revenue gain coefficient for average renewq years

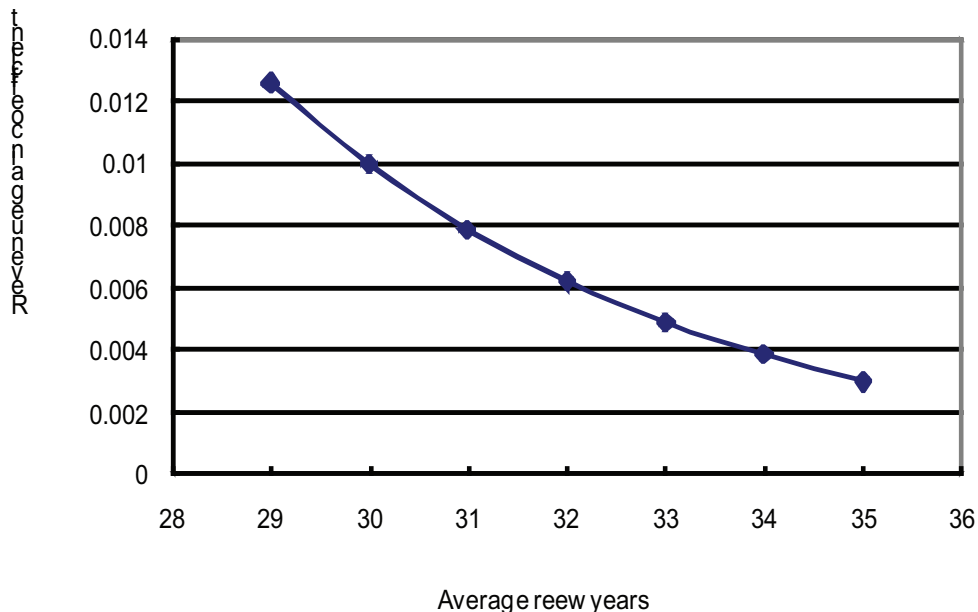


Fig. 13. The revenue gain coefficient for average renew years

Summarizing tables 1,2,3 and figures 2,2,3, the conclusions can be obtained:

1. The lead close rate is the most important parameter to determine the InsWeb online policy revenue. The revenue gain from lead close rate increasing will generate 10 times revenue obtained from retention rate increasing in the same percentage based on currently close rate and retention rate.
2. The relative revenue gain coefficient will decreasing with the close rate increasing, the theoretical minimum value is 1 when lead close rate reach 100%.
3. The retention rate is the second important parameter to affect the InsWeb online insurance policy revenue increasing. It almost keeps constant, 1.0, to the relative online policy revenue gain coefficient.
4. The parameter of average policy renew years is less important for InsWeb online policy revenue on current data mining results.

11. The maximum revenue increasing direction and revenue potential

The online policy revenue for year *i* can be written as

$$T_{2000+i} = \sum_{j=1}^{M_{2001+i}} (C * Q_j * R_j) + \sum_{j=1}^{N_{2001+iii}} (P_j * C * Q_j) \tag{36}$$

By introducing the average values, the equation (35) can be written as

$$\frac{T_{2001+i}}{\bar{R}_{2001+i} + N_{2001+i}} = M_{2001+i-1} * C * \bar{Q}_{2001+i} * \bar{P}_{2001+i} * \bar{R}^{-i-j} \quad (37)$$

In equation (36), \bar{R}_{2001+i} and \bar{P}_{2001+i} will be considered as non-history related; M_{2001+i} is strongly history related, N_{2001+i} and \bar{Q}_{2001+i} can be considered as changing history related trends (this research will be done later).

$$M_{2001+i-1} = \sum_{j=1}^{i-1} N_{2001+j} * \bar{P}_{2001+j} * \bar{R}^{-i-j} \quad (38)$$

Therefore, equation (26) can be written as

$$\begin{aligned} T_{2001+i} &= \sum_{j=1}^{i-1} N_{2001+j} * \bar{P}_{2001+j} * \\ &\bar{R}^{-i-j} * C * \bar{Q}_{2001+i} + N_{2001+i} * \bar{P}_{2001+i} \\ &* C * \bar{Q}_{2001+i} \end{aligned} \quad (39)$$

Moreover, equation (39) can be written as

$$T_{2001+i} = \sum_{j=1}^i N_{2001+j} * \bar{P}_{2001+j} * \bar{R}^{-i-j} * C * \bar{Q}_{2001+i} \quad (40)$$

Considering the nature increasing process of online insurance shopping consumer number and auto insurance quote premium with time going, so

$$N_{2001+j} = N_{2001} * (1 + r_1)^j \quad (41)$$

and

$$\bar{Q}_{2001+j} = \bar{Q}_{2001} * (1 + r_2)^j \quad (42)$$

so equation (40) can be written as

$$\begin{aligned} T_{2001+i} &= \sum_{j=1}^i N_{2001} * (1 + r)^j * \bar{P}^{i-j} \\ &* C * \bar{Q}_{2001} * (1 + r_2)^i \end{aligned} \quad (43)$$

Generally, \bar{P}_{2001+j} can be assumed as constant, so

$$\begin{aligned} T_{201+i} &= N_{2001} * \bar{P}_{2001} * C * \\ &(1 + r_2)^i \sum_{j=1}^i (1 + r_1)^j * \bar{R}^{-i-j} \end{aligned} \quad (44)$$

Since

$$\sum_{j=1}^i (1+r_1)^j * \bar{R}^{i-j} = \frac{(1+r_1) \left(\bar{R}^{i+1} - (1+r_1)^i \right)}{R-1-r_1} \tag{45}$$

Therefore, equation (44) can be written as

$$T_{2001+i} = N_{2001} * \bar{P}_{2001} * \bar{Q}_{2001} * \frac{(1+r_2)^i (1+r_1) \left(\bar{R}^{i+1} - (1+r_1)^i \right)}{R-1-r_1} + T_{2000} * \bar{R}^{i+1} * (1+r_2)^{i+1} \tag{46}$$

Since there was relocation even in the end of year 2000, we have to consider year 2000 online auto policy revenue separately. Now we can predict the InsWeb online agent policy revenue based on analysis above:

year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
T_{2001+i}	1.634	2.354	2.973	3.522	4.024	4.498	4.957	5.412	5.872	6.343

Table 4. The InsWeb online agent policy revenue prediction

Unit: million

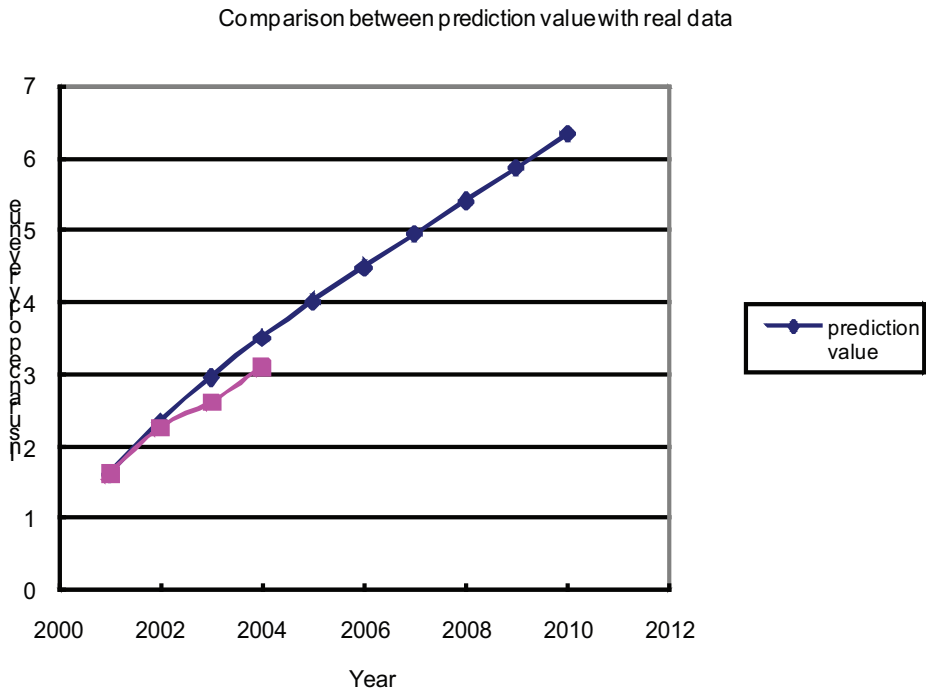


Fig. 14. The InsWeb online agent policy revenue prediction

The result in figure (4) is based on the assumptions of 2001 year carrier-state combination number, auto online auto online agent coverage and lead close rate. The whole research and prediction mentioned above was done in 2001. After 3 years some validation data is available now. The validation data is put on figure (4) for comparison. The reason that real business data is somehow lower than prediction value is that carrier-state combination number has some decrease with time going, but the trend is follow the prediction and the comparison is acceptable.

12. Conclusion

1. The paper presents a formula to score online insurance consumer to help insurance agents.
2. The report derived a formula to estimate lifetime value of online auto insurance consumer based on probability analysis and data mining of consumer information. The formula can be written as

$$S = P * C \sum_{i=0}^n \left[Q_i \prod_{j=0}^i R_j \right] \quad (47)$$

3. After assumption of same retention rate, the simplified formula to evaluate lifetime value of online auto insurance consumer is

$$S = \frac{P * C * Q_0 \left(1 - R_1^{n+1} (1+r)^{n+1} \right)}{1 - R_1 (1+r)} \quad (48)$$

4. According to data analysis, retention rate is 0.75, average possible renew year is 29 and the total revenue generating from a new sold auto online policy is 4.25 time its first commission revenue.
5. This report makes the sensitive analysis to different parameters. The result shows that increasing lead close rate is most priority. It will obtain 7~10 time gain for the same revenue gain from the retention rate increasing.
6. A formula to estimate Insweb auto insurance online policy revenue is developed and 2002 to 2010 revenue prediction is given out based on this formula.
7. The comparison between predicted business revenue and real data from 2002 to 2004 is acceptable. Therefore the model and the data mining are successful.

13. Acknowledgment

The publication of the chapter is supported financially by Nanjing hydraulic Research Institute, China.

14. Reference

- [1] http://more.abcnews.go.com/sections/business/dailynews/auto_insurance_rates_010607.html
- [2] http://more.abcnews.go.com/sections/wnt/dailynews/healthcarecosts_wnt010725.html
- [3] <http://www.census.gov/prod/2002pubs/01statab/vitstat.pdf>
- [4] http://biz.yahoo.com/prnews/020131/sfth063_1.html



Knowledge-Oriented Applications in Data Mining

Edited by Prof. Kimito Funatsu

ISBN 978-953-307-154-1

Hard cover, 442 pages

Publisher InTech

Published online 21, January, 2011

Published in print edition January, 2011

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by 'Data Mining' address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yuanya Li, Gail Cook and Oliver Wreford (2011). Online Insurance Consumer Targeting and Lifetime Value Evaluating - A Mathematics and Data Mining Approach, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1, InTech, Available from:
<http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/online-insurance-consumer-targeting-and-lifetime-value-evaluating-a-mathematics-and-data-mining-app>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.