Applied Mathematics & Information Sciences
An International Journal

The influence of the expansion of the smart mobile phones on the insurance industry

Pavel Pesout[1] *and Ondrej Matustik*[2]

[1] Faculty of Statistics and Informatics, University of Economics, Prague, Czech Republic
[2] Faculty of Statistics and Informatics, University of Economics, Prague, Czech Republic

Received: Jul 8, 2011; Revised Oct. 4, 2011; Accepted Oct. 6, 2011
Published online: 1 Sep. 2012

Abstract: Development in the area of smart mobile phones is still accelerating. The expansion of smart mobile phones will bring new challenges and tasks to almost all industry areas. One of them will be the insurance industry. The clients will ask for new insurance products according to their actual personal needs and surroundings that could be effect by using smart mobile phones. In this article we try to predict such business opportunities in the new kind of insurance products development and its application. We focus on both the business and the risk model because these changes will cause a massive movement in the way of premium calculation.

Keywords: Smart mobile phones, insurance industry, net premium calculation principles

1. Introduction

Advancement in the field of electronic contact with customers and number of contact possibilities is quite rapid nowadays. The entrance of smart phones like Apple iPhone, Android or MS Mobile 7 devices introduced in past several days brings the new opportunity for the companies how to be in contact with their clients and how to sell them additional products and services. This development brings an opportunity to almost all industry areas which are using internet and web pages to communicate with their customers. The usage of smart phones generally means that the communication between the client and company will be more personal, more tailored to the specific actual client's needs, faster and more flexible.

This brings the new challenges for the insurance companies, as well. A new generation of insurance products will appear to cover the risk of some real-world situations which the client has not covered yet but the client wants to do now face-to-face the potential danger. If we look for a typical product of this new generation of insurance, this can be short term insurance (just for a couple of hours) in the case of some unexpected event, e.g. the insurance of a potential broken leg if there is an ice outside on the path. This kind of insurance would not been effected so far, but it could be now by using smart phones and their applications.

The aim of this article is also situated to the changes in the communication, business and risk shifts and some additional questions in the insurance industry caused by enabling of such new kind of personalized insurance products with different computational model than current ones that are based on the central limit theorem, law of big numbers and statistics valid on the big number of clients connected with big number of observations.

2. Communication and information transfer

As it has been mentioned, we can expect several changes in the insurance industry in the close future. First of all, let us focus a little bit on the possible way of changes in the communication with the clients. With rapid development of smart phones market and number of smart phones accessing to the internet it is more or less obvious that smart phones are one of the main channels for the communication with the clients for servicing them and selling them new products. What is more frequently being seen is the customization of this channel to the company special needs. We can generally distinguish two main possibilities

* Corresponding author: e-mail: pavel.pesout@ness.com

of customization of the communication between insurance company and the clients:

– Some companies are selecting the strategy that the clients will use some build-in internet browser to visit their internet pages. Therefore, they are preparing the mobile version of their internet pages with lower screen resolution and simplifier design and more accessible for the smart phones clients.
– Other companies are developing special application direct for smart phones. This enables the clients to reach needed functionality and bigger comfort more easily and on the other hand it enables to the providing company the access to the client's data saved in the phones.

These approaches provide servicing the clients in a more comfortable way. However, the services themselves are unmodified. But what if there can be whole new kind of insurance products, which were previously not affordable because of the technological and information transfer limitations? We believe that there is a space for new products in the insurance industry, tailored more specifically to the actual client's needs and surroundings. These products should fully benefit from the technology opportunities and data provided by the clients and their smart phones.

We imagine some specific real-world situations that are time limited and are based on current conditions and mind of the client. The client often cannot insure itself in advance because the important conditions would change or the client did not know about the activity until now. These kinds of situations may be e.g. sport activities (parachuting, sky diving) or more often just ordinary activities of the everyday life (driving to work). Is it possible to insure this kind of the client's activity just a minute before it happens because of the client's actual decision? We strongly believe it is. What is needed is just a smart phone. By using it the client will apply for the insurance on the company's web pages or with some special company's application. But what is more, the company itself may collect additional data important for the insurance risk rate directly with the aid of the smart phone.

Basic information for almost every insurable activity that has been proposed is the client location and its conditions (such as weather, traffic, geological lay etc.) However, people localization by using mobile phones (and especially smart phones) is becoming a usual thing and there are a lot of theoretical studies concerning on this topic. For example in [1] the authors are dealing with the idea of monitoring the traffic density in the Bangalore (city in India) based on the Nericell system and smart phones included. There are also some practical results in this research field, users of smart phones can use a service called Google Latitude, see [2], for sending information about their location and observing the locations of their friend. And Baker in [3] describes the real application of mobile localization system in the San Francisco city where the company Sense Network is locating all its system's participants. This helps to build a map where we can distinguish what the users are doing through the day and divide them into different groups. We can see that obtaining the location and surroundings condition (with the client's approval) would not be any problem. Moreover, it is possible to obtain other relevant information such as the client's latest activity, but also (e.g. with help of social networks) data about the location of the client's friends and locations of other people in the surroundings.

In fact, it can be said that what we propose is in some sense other (more developed) kind of model that uses the insurance company Progressive from Ohio USA in car insurance nowadays, see [4]. Thanks to the build-in computers in the cars it monitors the data about kilometers run, average speed and usual places of riding and another factors. Its clients can claim an insurance bonus (or can be exposed to the insurance malus) based on this data, when they send them to the insurance company office. Although, this communication is in the off-line mode but our communication and information transfer using smart phones will be in on-line mode. It is very important to do all the data obtaining tasks, communication with the clients and risk modeling (calculating premium) in short time period. When the client wants to effect the insurance and go e.g. parachuting, it is necessary to act very quickly because the client cannot wait for a long time.

3. New products principles

Having defined the possibilities of current communication we may focus on the definition of the new kind of insurance products itself. We assume these main characteristics:

– The products will be more personalized - each client has different needs in a time, so the products will be more suited to the client's specific actual needs. The products can be very time limited and can be valid even just for several minutes or hours.
– The data needed by the insurance company for the products settings will be obtained in two ways - in classical way during the first (or any further) contact with the client (the client will have to effect some general insurance agreement with the company to provide the services included and will afford basic data like sex, age, health status or previous experiences present) and in new technology way as it has been proposed above (when actual client's needs to insure occurs). The products and their prices will more depend on the client's surroundings and with the access to this information via smart phones the potential involvement of this data will growth rapidly.
– The insurance costs will be influenced by provided client's data and it is obvious that some clients would like to provide not so fully correct, true and complete data set. Also the model should calculate with several risks and uncertainties. In order to minimize the risks the insurance company will value the client's information and privacy, which can be very sensitive for some clients. So the clients would be naturally motivated to

provide all the important, actual and correct data by some discount. It is simply. When the client will be not willing to provide all required data, the risk model will be not so specific and it will have to count with some level of uncertainty.

– We assume each client to settle single payment or payment by installments as a form of compensation for overall overhead cost, e.g. handling, policy, data storage or operating (it will be settled by including the new client to the system). Moreover, there will be payments for the insurance - premium - of actual and individual client's needs. These will be settled in different manner. Current practice when clients are paying with some delay via their banking account or on the branches should be changed in the way of immediate payment. The payments must be done via internet credit card or via special methods supported by smart phones in the moment of the client's needs.

4. Actuarial calculation for the product

The new kind of insurance products based on mentioned principles will ask for new statistical approach to actuarial calculations than are current ones, see [5] and [6] for details. Current life insurance models are based on mortality tables, probabilities of life and dying and technical interest rate. Though, they have no influence on the risk calculation for the sake of insured activities that come into being presently. What is more, in the methods of general insurance mathematics with the law of big number we are also not so interested because of required individual ingress to each application.

First of all we have to deal with the data reduction problem, because the company will derive via smart phones and additional sources much more information than is in fact needed. For each client's need it is necessary to identify the set of key features within application data and behavioral attributes that significantly influence the end-result death, the type of an accident or survival without consequences. The features will be different according to the type of activity, of course. Let the general contract with the K-th insurable actions include insurance claims u_h, $h = 2, \ldots, H$, representing $H-1$ different accident types and u_1 as event without any claim. We assume that for the k-th action there is the training or historical data set available to our analysis formed in $k_n\ x\ k_A$ matrix $\mathbf{Y}_{k,A}$ of k_n applications assigned to the clusters $S_{k,1}, \ldots, S_{k,H}$ representing u_h knowledge, and of the k_A variables representing all possible attributes of the client and its surroundings that may be obtained. Our aim is to choose only the attributes that account for maximal amount of data total variance and play the most important role in specifying the result of the action. The dimension of a large number of interrelated variables, while keeping as much information as possible, may be reduced by using the principal component analysis, see [7]. We transform

$$\mathbf{Z}_k = \mathbf{C}_k \mathbf{Y}_{k,A}, \tag{1}$$

where $\mathbf{C}_k = (C_{k,1}, \ldots, C_{k,k_A})^T$ is the matrix of eigenvectors of the covariance matrix of the original data set arranged as columns and order by eigenvalues, highest to lowest. The first principal component $Z_{k,1}$ accounts for maximal amount of total variance and each following component accounts for maximal amount of variance that was not accounted for by the preceding components, and is uncorrelated with all of the preceding components. This is why only the first few components are retained. In order to determine the number of important components we set k_B so that

$$\frac{\sum_{g=1}^{k_B} Z_{k,g}}{\sum_{g=1}^{k_A} var(Y_{k,A,g})} > 1 - \epsilon. \tag{2}$$

We eliminate the last ($k_A - k_B$) components. To determine the importance of each variable, we use the absolute values of first k_B of each component's coefficient and we choose only first k_D of them. Using principal component analysis we obtain the new $k_n\ x\ k_D$ data set matrix $\mathbf{Y}_{k,D}$. Now let us find the most important variables to the results u_h, thus to the partitioning into $S_{k,h}$ groups. It can be achieved by maximizing the generalized Rayleigh quotient, see [8],

$$F_{k,D}(\mathbf{a}_{k,D}) = \frac{\mathbf{a}_{k,D}^T \mathbf{B}_{k,D} \mathbf{a}_{k,D}}{\mathbf{a}_{k,D}^T \mathbf{E}_{k,D} \mathbf{a}_{k,D}}, \tag{3}$$

where $\mathbf{B}_{k,D}$ is the $\mathbf{Y}_{k,D}$ between classes scatter matrix and $\mathbf{E}_{k,D}$ is the $\mathbf{Y}_{k,D}$ within classes scatter matrix. The problem reduces to finding the weights which discriminate well among groups according to various criteria (e.g. orthonormality). Hence, for $k_q = min(H-1, k_D)$, we find the eigenvectors $\mathbf{a}_{k,D,1}, \ldots, \mathbf{a}_{k,D,k_q}$, of $\mathbf{B}_{k,D}(\mathbf{E}_{k,D})^{-1}$, which elements are called the canonical coefficients. According to a goodness-of-fit-parameter, Wilk's lambda, and a F-test of comparing the means, we choose only those with the biggest eigenvalues. The most important variables are marked by the absolute values of the elements of correlation vector between original variables and the canonical variables as well as the canonical coefficients. By finding the features which optimally separate our clusters we obtain the new $k_n\ x\ k_M$ data set matrix $\mathbf{Y}_{k,M}$. From the whole large set of possible attributes that we eventually could collect to calculate the premium, we find those having the greatest influence on the data set variability and the action result. Thus, the problem of obtaining the high level data is considerable simplified because the insurance company may intend only in this way defined attributes.

Now we can focus on the net premium calculating of the new application represented by vector $\mathbf{Y}_{k,M,k_{n+1}} = (Y_{k,M,k_{n+1}1}, \ldots, Y_{k,M,k_{n+1}k_M})^T$ (only most important k_M attributes for the k-th insurable action are involved). Our aim is to find the probabilities of all possible u_h results. Hence, we have to deal with the classification problem, see [9]. We may apply the canonical discriminant analysis in a similar manner as it has been introduced by finding the most important features, but in this case to finding partitioning of new element into existing groups. For the reason we construct the $\mathbf{M}_k$ matrix of eigenvectors

$\mathbf{a}_{k,M,1}, \ldots, \mathbf{a}_{k,M,k_p}$ of the $\mathbf{B}_{k,M}(\mathbf{E}_{k,M})^{-1}$ matrix, where $k_p = min(H-1, k_M)$, $\mathbf{B}_{k,M}$ is the $\mathbf{Y}_{k,M}$ between classes scatter matrix and $\mathbf{E}_{k,M}$ is the $\mathbf{Y}_{k,M}$ within classes scatter matrix. The condition labeling the new individual into $S_{k,h}$ (u_h result) takes the form

$$d^2_{Mh}(\mathbf{M}_k^T \mathbf{Y}_{k,M,k_n+1}, \mathbf{M}_k^T \overline{\mathbf{Y}}_{k,M,h}) = \tag{4}$$

$$\min_{r=1,\ldots,H} d^2_{Mh}(\mathbf{M}_k^T \mathbf{Y}_{k,M,k_n+1}, \mathbf{M}_k^T \overline{\mathbf{Y}}_{k,M,r})$$

with the Mahalanobis distance d_{Mh} and the r-th cluster mean vector $\overline{\mathbf{Y}}_{k,M,r}$.

Another tool widely used for multi-class classification is the functional discriminant analysis. However by using both methods we set for the application $\mathbf{Y}_{k,M,k_n+1}$ of the k-th action the probability vector of possibly results u_h - $\mathbf{w}_{k,k_n+1} = (w_{k,(k_n+1)1}, , w_{k,(k_n+1)H})^T$. We also set the individual net premium as equal to the individual accidental loss through the medium of the results probabilities and the appropriate sum insured level. With regard to administrational and implementation reasons we suppose that for each k-th event there are k_J sum insured levels $e_{k,j} = (e_{k,j1}, \ldots, e_{k,jH})^T$, $j = 1, \ldots, k_J$, belonging to the u_h results. So, more transparently, for each j-th sum insured level we define the net premium as

$$\Psi_{k,j}(\mathbf{w}_{k,k_n+1}) = w_{k,(k_n+1)1} e_{k,j1} + \ldots \tag{5}$$

$$+ w_{k,(k_n+1)H} e_{k,jH}.$$

Nevertheless, it is also possible to create the whole tariff system. We may take advantage of the clustering analysis and its non-probabilistic K-means method, see [10], and for $s = 1, \ldots, k_n$ iterative relocate the objects $\mathbf{Y}_{k,M,s} = (Y_{k,M,s1}, \ldots, Y_{k,M,sk_M})^T$ included in the $\mathbf{Y}_{k,M}$ matrix into the $T_{k,i}$ tariff classes, $i = 1, \ldots, k_T$, by minimizing the within-cluster variance. Because of result knowledge of every measurement from the training data set we can calculate for each $T_{k,i}$ class vector the mean u_h results probabilities $\overline{\mathbf{w}}_{k,i} = (\overline{w}_{k,i1}, \ldots, \overline{w}_{k,iH})^T$, $i = 1, \ldots, k_T$. The new $\mathbf{Y}_{k,M,k_n+1}$ application is being assigned into the closest $T_{k,i}$ class. Therefore, after choosing the sum insured $e_{k,j}$, the net premium is given as

$$\overline{\Psi}_{k,i,j} = \overline{w}_{k,i1} e_{k,j1} + \ldots + \overline{w}_{k,iH} e_{k,jH}. \tag{6}$$

In fact, instead of net premium, the gross one will be written. It will be given as the sum of the net premium, the safety loading which covers the negative risk development, the profit margin and some amount of money covering the communication costs. The other costs will be compensated by the payments specified at the moment of the general contract conclusion.

5. Conclusion

In this article we proposed the set-up of new kind of insurance products as an affordable solution of close cooperation between the technology development on the field of smart mobile phones and increasing abilities of obtaining personal data. We were dealing with the information transfer channels and with the new products basic business and risk principles. Our proposed framework integrates the net premium calculation with some methods of multivariate statistics because current models of actuarial mathematics are insufficient. By using principal component analysis and discriminant analysis we were able to reduce the data dimension while keeping essential features and as much of information as possible. Finally, by solving classification problem we calculated individual and tariff net premium of the application reflecting the actual client's needs.

References

[1] P. Mohan, V. Padmanabhan, R. Ramjee: Nericell: Rich monitoring of road and traffic conditions using mobile smartphones, Proceedings of SenSys ACM (2008), p. 323.

[2] Information on http://www.google.com/latitude.

[3] S. Baker: *The Next Net*, The McGraw Hill Companies/BusinessWeek (2009).

[4] Information on http://www.progressive.com.

[5] C. D. Daykin, T. Pentikdinen and M. Pesonen: *Practical risk theory for actuaries*, Chapman Hall (1994).

[6] M. Baxter, A. Rennie: *Financial calculus*, Cambridge University Press (1996).

[7] I. T. Jolliffe: *Principal component analysis*, Springer-Verlag (1986).

[8] J. A. Rice: *Mathematical Statistics and Data Analysis*, Duxbury Press (2006).

[9] A. J. Izenman: *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer (2008).

[10] J. A. Hartigan: *Clustering Algorithms (Probability Mathematical Statistics)*, John Wiley Sons Inc (1975).

Pavel Pesout obtained his Ph.D. degree at the University of Economics, Prague, in 2010 and since then he is there a researcher/lecturer at the Department of Statistics and Probabilities. His current research interests include functional data clustering techniques.

Ondrej Matustik obtained the Ph.D. at the Department of System Analysis, University of Economics, Prague, in 2011. His current research interests include ubiquitous computing and cloud computing interactions with the real-world applications.