Social Media Networking Data Analysis in Life Insurance Underwriting

1 Chiang Ku Fan, Yu Hsuang Lee

1 Department of Risk Management and Insurance, Shih Chien University, Taipei, Taiwan
2 Graduate Institute of Industrial and Business Management, National Taipei University of Technology, Taipei, Taiwan

ABSTRACT

In order to make an appropriate underwriting decision, underwriters will often cast a broad net in discovery requests, seeking as much documentation as possible to search for inconsistencies in the applicant or policyholder's story or indications of potential fraud. Fortunately, the new online social networking technology may help insurance companies to improve their underwriting profits and select prospective policyholders. Neither regulators nor insurers have developed guidelines for the overall use of social data, and scientific studies have not determined what types of social media data are referable. To fill this research gap, this study employs both literatures reviewing and Conjoint Analysis methods to indentify and weight factors underwriters prefer to use in making underwriting decision. The findings may provide information for those who employing social media networking data to make underwriting decision to attain underwriting profits, select prospective policyholders, and provide equity among policyholders.

Keywords: insurance underwriting, social media networking, adverse selection, insurance fraud

1. INTRODUCTION

Insurance companies are charged, on the one hand, with taking policyholders’ premiums to protect the insured from the risk of potential losses; on the other hand, insurance companies are charged with serving as gatekeepers to prevent policyholders from taking too much from the risk pool. Many functions can help insurance companies to take responsibility for this difficult task. One of the most important functions of an insurance company is the underwriting process, including selecting, classifying, and pricing applicants for insurance. The major objective of underwriting is to determine if an applicant is acceptable for the insurance under the conditions indicated. Through underwriting, an insurance company can produce a safe and profitable distribution of business. Insurance scholars, practitioners, and supervisors have a long history of evaluating insurance applicants’ knowledge, skills, and ability directly through a wide variety of sources, including applications, agents’ reports, inspection reports, physical examinations, and attending physicians’ reports [1]. Unfortunately, many of these assessments are at risk of fraudulence and adverse selection. Insurance fraud hurts the insurance companies and everyone else because it adds 10% to the cost of the average policy [2].

To make appropriate underwriting decisions and prevent insurance fraud, insurance companies attempt to collect various sources of data to accurately rate the risk profile of certain classes of policyholders or applicants. Traditionally, underwriters rating loss exposure or those presented with potential adverse selection or moral risk tend to rely on tools for their inspection. The agent is told what types of applicants are acceptable, borderline, or prohibited. Underwriters also require certain information to decide whether to accept or reject an applicant for insurance. The required information includes the application, agent’s report, inspection report, physical inspection, or physical examination [1]. In this context, underwriters will often cast a broad net in discovery requests, seeking as much documentation as possible to search for inconsistencies in the applicant or policyholder’s story or indications of potential fraud. However, these traditional techniques are labor intensive and very expensive [3]. Fortunately, the new online social networking technology may help insurance companies to improve their underwriting profits and select prospective policyholders. Online social networking websites and microblogging services allow users to post and read text-based messages of up to 140 characters, such as “Facebook” and “Twitter”. There are more than 554 million active registered Twitter users and 1.11 billion people using Facebook, according to reports from Twitter statistics and Yahoo Finance in 2012. Almost 72% of all US Internet users are on now Facebook, and 70% of the entire user base is located outside of the US. In other words, Facebook is now used by one in every seven people on earth. Every 20 minutes, more than 2.7 million photos are uploaded, 2.7 million messages are sent, one million links are shared, and 10 million comments are posted on Facebook, based on information provided by “WWW.ONLINESCHOOLS. ORG” in 2011. Because online social networking websites have both high frequency use and wide coverage, employers have arguably been quicker than organizational scientists to realize social media’s assessment potential [4]. Numerous studies have...
examined employers’ social media usage to select job candidates and observe employees [5] [6]. Individuals have often been cautioned to watch what they post or otherwise divulge via social media because employers may base hiring and firing decisions in part on what they find online. Outside of the workplace, many job applicants use social media for personal communication that is unintended for employers [4], often leaving public traces of their social communication in cyberspace through forums such as blogs, tweets, and posts on social networking web sites such as Facebook [7]. In other words, job applicants’ online activity, including Facebook activity, tweets, and online searches, can serve as background for employers making hiring decisions. There is now another group that may also be watching people’s social networking and analyzing the data that they glean from it: insurance companies. Social media data will pay dividends for insurers in areas such as underwriting, claims, and subrogation [8].

Social media networks provide a rich source of data that insurers can use to improve a variety of operational processes [8]. However, insurers face obstacles that may impede the speed-to-market of applying social networking data to underwriting [9]. This is because neither regulators nor insurers have developed guidelines for the overall use of social data, and scientific studies have not determined what types of social media data are referable [9]. To fill this research gap, the first purpose of this study is to identify what underwriting factors underwriters prefer to search for in social media networking. The second purpose of this paper is to explore the types of social media data that may offer the best insights on underwriting factors for insurers to make underwriting decisions. The findings may provide information for those who employing social media networking data to make underwriting decision to attain underwriting profits, select prospective policyholders, and provide equity among policyholders.

2. LITERATURE REVIEW

2.1 Information Provided by Social Popular Networking Sites

Facebook, Twitter, Google +, and LinkedIn will be the most popular social networking sites in the world by 2014 according to research conducted by eMarketer, a company located in New York that provides the most complete view of digital marketing available to the world’s top brands, agencies, and media companies. The following is description of the type of information available from each site.

With 750 million active users on Facebook, it is almost certain that any applicants or policyholders will have a Facebook profile. A profile provides Facebook users with a forum for presenting their experiences, interests, and thoughts to a selected circle of friends or to the public at large. Because it provides a messaging feature that allows direct communication between Facebook users, the information on Facebook can be used to develop a picture of a person’s activities before and after an insurance application [10].

A Twitter posting is a text-based post of up to 140 characters. Tweets are essentially text messages posted in real time for communication or discussion with a tweeter’s followers. Usually, tweets contain links to other sources of information, such as photograph repositories or websites. Moreover, users have direct conversations with other users through tweets directed at individuals using the @ symbol. Searching Twitter may produce information relevant to whether an insured individual suffers from sickness or injuries [3].

Google + is a relatively new player introduced to the social networking field in June 2011. Google + is designed to integrate other Google services related to a user’s Google profile that contain many discussion forums. Google plus also contains new social networking features, including “Circles”, “Hangouts”, “Huddles”, and “Sparks” [11], which may provide a wealth of information to insurance underwriters about a policyholder’s friends, interests, group video chats, and text messages within various circles.

LinkedIn, with 225 million members in more than 200 countries, is business oriented and is the world’s largest professional networking site. LinkedIn users post resume-type information about their current employment, work history, experience, and educational background. The information posted on LinkedIn may help insurance underwriters recognize policyholders’ real working situation, experience, and environment [3].

2.2 The Role of Social Media in Insurance Underwriting

The immediacy of social media data enables insurers to shift underwriting from a static process that relies upon backward-facing data to a dynamic process that relies upon real-time data [8]. In the near future, insurers will be increasingly sensitive to the connection between an insured person’s credit score and his or her potential risk for loss. The relation between the activities in which users engage online and their riskiness as policyholders is becoming an important issue [11]. The use of social media networking continues to grow in absolute numbers and to expand to all age groups, and new approaches are using social media data from online networking sites in operational applications for underwriting. Insurers should consider social networking because of who uses it and what is being posted [12].
As Ha predicted [9], automatically mined data from social networking sites may find their way into the underwriting pricing process. Social media data may become a factor in determining premiums for both personal and business insurance.

2.3 Social Media Data Used as Sources of Evidence in Courts of Law in Claim Cases

Fraud is a significant challenge to the insurance business. The explosion of new Internet-based technology combined with a poor economy has encouraged unscrupulous individuals to find new ways to commit insurance fraud. In this context, insurers and lawyers have found ways to take advantage of online social media to fight fraudulent claims [13]. Scouring Facebook and other social networking pages of policyholders is a common practice on the claims side of the business. Many investigators report that navigating an insured individual’s online social media page is one of the first things they do when looking into potentially fraudulent claims, according to a report from Boston-based research firm Celent in 2011. Online social media is a goldmine for the discovery of insurance fraud, particularly in the litigation process [3]. Chastain stated that social media is obviously an important factor in insurance fraud investigation [14]. There have been many situations in which the public information available through social media has been beneficial in insurance fraud investigations.

Social media network data are used extensively as sources of evidence in claim cases in courts of law. Underwriting will be the next area [9] if key techniques can be developed or enhanced, including reliable authentication methods, improved data extraction tools, and more advanced analysis tools [12]. Insurers have not yet provided guidelines in terms of the overall use of social data, and these data are not yet approved for use in the pricing process [9].

2.4 Important Underwriting Factors That Determine a Life Insurance Premium

The world of underwriting is evolving. Paramedical exams are used more often, and blood tests have become a staple of underwriting. However, the basic factors considered by insurers to make underwriting decision are similar to those in the past [15], according to many previous studies (e.g., [16] [17] [18] [19]). The factors considered in making underwriting decisions include 11 determinants and can be framed as in the following structure (Figure 1).

![Figure 1 Determinants of Underwriting Decision Making](image)

2.5 Useful Social Media Data in Underwriting

As users interact with multiple social networking sites, purchase items online, and communicate with others in public forums, they leave behind data about their preferences, lifestyle, operations, and habits. Another piece of useful information that social media data can provide is the “social graph”, which shows how individuals or companies are...
linked together, providing a picture of who is friends with whom, who follows whom, and people’s friends of friends. In addition to identifying fraud organizations, these graphs can give underwriters further insight into how an individual may perform in terms of risk based on the behavior of those to whom he or she is connected [3] [20]. In general, useful information can be searched by underwriters through social media networking sites, including individuals’ interaction with multiple social networking sites, purchase of items online, communication with others in public forums, and social graph.

3. METHODOLOGY

The purposes of this study are to identify what factors underwriters prefer to search on social media networks and to explore what types of social media data may provide the best insights for insurers to judge underwriting factors. To satisfy the purposes of the research, this study first reviews prior studies to identify the factors considered in underwriting by insurers and the types of social media data typically posted on social media networks. Then, this study employs the conjoint analysis (CA) to identify the weight of each considered factor. To compare the weight of each factor, this study identifies the factors that are searched most frequently by underwriters on social media networks. The concept of conjoint analysis is introduced in this section, as well as the determined formula of the utility with the conjoint analysis. The final part in this section discusses the process of data analysis with conjoint analysis.

Conjoint analysis (CA) has been employed in research for many years. Panda and Panda have described CA as a “what if” experiment in which buyers are presented with different possibilities and asked which product they would buy [21]. In other words, CA is a multivariate technique used specifically to understand how respondents develop preferences for products or services [22]. Sudman and Blair emphasized that CA is not a data analysis process, such as cluster analysis or factor analysis; it can be regarded as a type of “thought experiment,” designed to display how various elements, such as price, brand, and style, can be used to predict customer preferences for a product or service [23]. The basic CA model was computed with the ordinary least squares (OLS) regression parametric mathematic algorithm [24] using dummy variable regression. This basic model can be represented as follows [25] [26].

\[ U(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_{ij} \cdot X_{ij} \]

where

- \( U(X) \) = Overall utility (importance) of an attribute
- \( \alpha_{ij} \) = Overall utility of the j level of the i attribute
- \( i = 1, 2, \ldots, m \)
- \( j = 1, 2, \ldots, n \)
- \( X_{ij} = 1, \) if the jth level of the ith attribute is present
- \( = 0, \) otherwise.

According to the CA basic model, Churchill & Iacobucci presented a six-stage model that is based on the more critical decision points in a conjoint experiment [27].

3.1 Select attributes: The attributes are those insurance companies can do something about and which are important to underwriters. In other words, the insurance company has the technology to make underwriting decision that might be indicated by some important underwriting factors.

3.2 Determine Attribute Levels: The number of levels for each attribute has a direct bearing on the number of stimuli that the respondents will be asked to judge.

3.3 Determine Attribute Combinations: This will determine what the full set of stimuli will look like.

3.4 Select Form of Presentation of Stimuli and Nature of Judgments: Typically, three approaches can be used: a verbal description, a paragraph description, and a pictorial representation. One method for characterizing judgments is to ask respondents to rank the alternatives according to preference or intention to consider. Another method that is gaining popularity among researchers is to use rating scales.

3.5 Decide on Aggregation of Judgments: This step basically involves the decision as to whether the responses from underwriters will be aggregated.

3.6 Select Analysis Technique: The final step is to select the technique that will be used to analyze the data. The choice depends largely on the method that was used to secure the input judgments from the respondents.

4. DECISION MODEL APPLICATION AND RESULTS

The estimation model in this study consists of two phrases. In the first phrase, underwriting factors for underwriters are identified using the literature reviewing. The second phrase, in which the weights of the underwriting factors, also used
as the decision evaluation criterion, are calculated for insurers to make underwriting decision, is evaluated by employing the CA method. The second phrase is described in detail as follows.

For a formal analysis, the different attribute levels have to be dummy-encoded in a binary manner. The lowest attribute level serves as a reference point and gets a binary code of 0 (Helm et al., 2003). For any other attribute level, a binary digit of 1 is given if the level is present, and 0 is given if it is not.

Due to each of the attributes having two levels, the total number of possible combinations is $2^{11} = 2048$ alternatives (stimuli). This is far too many possible combinations to be evaluated by any decision maker. Therefore, we had to construct a design of the inquiry that defined a restricted set of stimuli to be considered and the pairs of these stimuli to be compared.

Starting with a basic orthogonal plan generated by Addelman (1962), 12 stimuli were determined (see Table 5). Using the stimuli of the orthogonal array, a difference design was constructed by a randomized procedure following the principles given by Hausruckinger & Herker (1992) (see Figure 3).

There are 30 life insurance companies in Taiwan in 2014. Seventy underwriting managers of life insurance companies are selected to comprise the group of experts under the condition that each experts has: (a) at least 10 years of professional experience in the life insurance sector, and (b) participated in the decision-making process of underwriting in life insurance companies. However, only 56 qualified underwriting managers agreed to share their opinion and answered the CA questionnaire.

### Table 5. Attribute Level and Orthogonal Plan Card of Cross-Buying Intentions

<table>
<thead>
<tr>
<th>Factors</th>
<th>Attribute Level</th>
<th>Card No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Low</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Smoking</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Occupation_Hobbies</td>
<td>Security</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Physical_Condition</td>
<td>Good</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Health_History</td>
<td>Bad</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Foreign_Travel</td>
<td>Rare</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Additional_Insurance</td>
<td>No</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Financial_Information</td>
<td>Good</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Moral_Hazard</td>
<td>Low</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Morale_Hazard</td>
<td>Low</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The CA questionnaire was developed on the basis of some of the literature, planned with an orthogonal design, and distributed to 56 experts. According to the CA report (see Table 7), the most important factor was financial information (relative importance = 17.992 %), the second most important factor was moral hazard (relative importance = 16.004 %) and the third most important factor was morale hazard (relative importance = 15.010 %).

### Table 7. Relative Importance of Underwriting Factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Variable</th>
<th>Part-Worth Utility</th>
<th>Relative Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Low</td>
<td>0.102</td>
<td>0.03496</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
<td>0.102</td>
<td>0.03496</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>No</td>
<td>0.204</td>
<td>0.06991</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Occupation and Hobbies</td>
<td>Security</td>
<td>0.233</td>
<td>0.07985</td>
</tr>
<tr>
<td></td>
<td>Danger</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Physical Condition</td>
<td>Good</td>
<td>0.131</td>
<td>0.04489</td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Health History</td>
<td>Good</td>
<td>0.161</td>
<td>0.05517</td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Foreign Travel</td>
<td>Rare</td>
<td>0.263</td>
<td>0.09013</td>
</tr>
<tr>
<td></td>
<td>Usually</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Additional Insurance</td>
<td>No</td>
<td>0.292</td>
<td>0.10007</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Financial Information</td>
<td>Good</td>
<td>0.525</td>
<td>0.17992</td>
</tr>
<tr>
<td></td>
<td>Bad</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Moral Hazard</td>
<td>Low</td>
<td>0.467</td>
<td>0.16004</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>
5. CONCLUSIONS AND RECOMMENDATIONS

According to the decision model application and results, this study has conclusions as follows:

5.1 Because most of the physical factors, such as age, gender, the occupation, and the health history, are declarations and required to fill in the application form. Therefore, this kind of physical factor is not necessary to be searched by underwriters on the social media networking sites. Moreover, through studying a body examination report, underwriters can independently identify insured’s physical condition and then makes the underwriting decision.

5.2 The non-physical factors, such as financial information, moral hazard, and morale hazard, are not required items to fill in the application form, but very important for underwriters to make underwriting decision accordingly. In order to improve underwriting profit, underwriters hope to search more information related to non-physical factors on the social media networking sites.

5.3 To compare all the underwriting factors, Financial Information, Moral Hazard, and Morale Hazard are the most three useful factors that underwriters want to search on the social media networking sites. On the other hand, age, gender, and physical condition are the factors that seldom need be identified by underwriters through social media networking searching.

5.4 As a result of the growing amount of information that is posted to social media networking sites, underwriting professionals, and the experts they engage, have discovered that social media can be a useful investigative tool for conducting research and uncovering relevant information on underwriting. Data derived from social media sites can serve to provide further confirmation of the information filled in an insurance application form, thereby assisting underwriting professionals to develop a proper social media underwriting guideline. The impact and influence of social media on underwriting handling, fraud preventing, and adverse selection avoiding cannot be ignored.

REFERENCES


<table>
<thead>
<tr>
<th>Morale Hazard</th>
<th>Low</th>
<th>0.438</th>
<th>0.15010</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Total Utility</td>
<td></td>
<td>2.918</td>
<td></td>
</tr>
</tbody>
</table>


