ANALYSIS OF INSURANCE UNDERWRITING USING SOCIAL MEDIA NETWORKING DATA

Chiang Ku Fan
Department of Risk Management and Insurance
Shih Chen University, Taipei
TAIWAN

ABSTRACT

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. 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. Fortunately, the new online social networking technology may help insurance companies to improve their underwriting profits and select prospective policyholders. However, insurers face obstacles that may impede the speed-to-market of applying social networking data to underwriting. 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 medial data are referable. 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.

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

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 inspections, physical examinations, and attending physicians’ reports (Rejda and McNamara, 2014). 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 (Nance-Nash, 2011).
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 (Rejarda, 2013). 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 (Cowan, 2011). 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 (Stoughton and Thompson, 2013). Numerous studies have examined employers’ social media usage to select job candidates and observe employees (Levinson, 2011; Holding, 2011). 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 (Stoughton and Thompson, 2013), 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 (Melidizadeh, 2010). 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 (Kenealy, 2013).

Social media networks provide a rich source of data that insurers can use to improve a variety of operational processes (Kenealy, 2013). However, insurers face obstacles that may impede the speed-to-market of applying social networking data to underwriting (Ha, 2011). 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 (Ha, 2011). 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.

LITERATURE REVIEW

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 (Ramasastrty, 2012).

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 (Cowan, 2011).

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 + also contains new social networking features, including “Circles”, “Hangouts”, “Huddles”, and “Sparks” (Merlinos and Associates, 2011), 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 (Cowan, 2011).

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 (Kenealy, 2013). 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 (Merlinos and Associates, 2011). 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 (Beattie and Fitzgerald, 2011).

As Ha (2011) predicted, 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.

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 (Griffin, 2011).

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 (Cowan, 2011). Chastain (2011) stated that social media is obviously an important factor in insurance fraud investigation. 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 (Ha, 2011) if key techniques can be developed or enhanced, including reliable authentication methods, improved data extraction tools, and more advanced analysis tools (Beattie and Fitzgerald, 2011). 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 (Ha, 2011).

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 (Kaltenbach, 1995), according to many previous studies (e.g., Aniskovich, 1998; James, 2001; Velazquez, 2002; Gersten, 2010). The factors considered in making underwriting decisions include 11 determinants and can be framed as in the following structure (Figure 1).
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 (Grisdela, 2011; Ha, 2011). 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.

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 analytic hierarchy process (AHP) 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. Additionally, by conducting AHP, this study explores the appropriate type of social media data that can be provided to underwriters in their judgment of underwriting factors (Figure 2).

![Figure 2. Research Procedures](image)

As a decision-making method that decomposes a complex multicriteria decision problem into a hierarchy (Saaty, 1980), AHP is a measurement theory that prioritizes the hierarchy and consistency of judgmental data provided by a group of decision makers. Using pairwise comparisons of alternatives, AHP incorporates the evaluations of all decision makers into a final decision without having to elicit their utility functions on subjective and objective criteria (Saaty, 1990). The steps of AHP are as follows.

**Step 1. Establish a hierarchical structure**

Complex issues can be addressed effectively by using a hierarchical structure given the inability of humans to compare more than seven categories simultaneously. A hierarchy should not contain more than seven elements. Under this limited condition, a rational comparison can be made, and consistency can be ensured (Saaty, 1980). The first hierarchy of a structure is the goal. The final hierarchy involves selecting projects or identifying alternatives, and the middle hierarchy levels appraise certain factors or conditions. The hierarchy structure of this study is shown in Figure 3.
The underwriting factors that underwriters prefer to search on social media networks act as evaluation factors to select the best types of social media data that provide the most insights into underwriting factors for insurers to make underwriting decisions.

**Step 2. Establishment of pairwise comparison matrix**

Based on an element of the upper hierarchy, the evaluation standard, a pairwise comparison is conducted for each element. Although n elements are assumed, n(n-1)/2 elements of the pairwise comparison must be derived. Let $C_1, C_2, \ldots, C_n$ denote the set of elements, whereas $a_{ij}$ represents a quantified judgment on a pair of elements $C_i, C_j$. The relative importance of two elements is rated using a scale with the values 1, 3, 5, 7, and 9, where 1 denotes “equally important”, 3 denotes “slightly more important”, 5 denotes “strongly more important”, 7 represents “demonstrably more important”, and 9 denotes “absolutely more important”. This yields an n-by-n matrix $A$ as follows:

$$A = \begin{bmatrix}
C_1 & C_2 & \cdots & C_n \\
C_1 & 1 & a_{12} & \cdots & a_{1n} \\
C_2 & 1/a_{12} & 1 & \cdots & a_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
C_n & 1/a_{1n} & 1/a_{2n} & \cdots & 1
\end{bmatrix}$$

The results of the comparison of the n elements are inserted into the upper triangle of the pairwise comparison matrix $A$. The lower triangle values are relative positions for the
reciprocal values of the upper triangle. Where \( a_{ij} = 1 \) and \( a_{ji} = 1/a_{ij} \), \( i, j = 1, 2, \ldots, n \), two elements (\( C_i \), \( C_j \)) become one quantization value for an important relative judgment. In matrix \( A \), \( a_{ij} \) can be expressed as a set of numerical weights, \( W_i, W_2, \ldots, W_n \), in which the recorded judgments must be assigned to the \( n \) elements \( C_1, C_2, \ldots, C_n \). If \( A \) is a consistency matrix, relations between weights \( W_i \) and judgments \( a_{ij} \) are simply given by \( W_i/W_j = a_{ij} \) (for \( i, j = 1, 2, \ldots, n \)) and matrix \( A \) as follows:

\[
A = \begin{bmatrix}
C_1 & C_2 & C_n \\
C_1 & 1 & \ldots & w_1 \\
\vdots & \vdots & \ddots & \vdots \\
C_n & w_n & \ldots & 1
\end{bmatrix}
\]

2)

Step 3. Compute the eigenvalue and eigenvector

Matrix \( A \) multiplies the elements’ weight vector \( (x) \) equal to \( nx \), i.e., \((A - nI)x = 0\), where \( x \) is the eigenvalue \( (n) \) of the eigenvector. Given that \( a_{ij} \) denotes the subjective judgment of decision makers, the actual value \( (W_i/W_j) \) has a certain degree of difference. Therefore, \( Ax = nx \) cannot be established. Saaty (1990) suggested that the largest eigenvalue \( \lambda_{\text{max}} \) would be

\[
\lambda_{\text{max}} = \sum_{j=1}^{n} a_{ij} \frac{W_i}{W_j}
\]

If \( A \) is a consistency matrix, \( \lambda_{\text{max}} \) be calculated by

\[
(A - \lambda_{\text{max}} I)X = 0.
\]

Step 4. Perform the consistency test

Saaty (1990) proposed utilizing a consistency Index (\( CI \)) and consistency ratio (\( CR \)) to verify the consistency of the comparison matrix. \( CI \) and \( RI \) are defined as follows:

\[
CI = (\lambda - n) / (n - 1) = 0
\]

\[
CR = CI / RI
\]

where \( RI \) represents the average \( CI \) over numerous random entries of same order reciprocal matrices. If \( CR \leq 0.1 \), the estimate is accepted; otherwise, a comparison matrix is solicited until \( CR \leq 0.1 \).

Step 5. Compute the entire hierarchical weight

After various hierarchies and element weights are estimated, the entire hierarchy weight is computed, ultimately enabling decision makers to select the most appropriate strategy.

Step 6. Calculate the whole level weight to select the best alternatives

In an alternative hierarchy level, there are five types of social media data.
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 and types of social media data, which may provide the best insights on underwriting factors for insurers to make underwriting decision, is evaluated - both by employing the AHP theory. The second phrase is described in detail as follows.

**Step 1: Designate the AHP participants**

There are 30 life insurance companies in Taiwan in 2014. Twenty 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 11 qualified underwriting managers agreed to share their opinion and answered the AHP questionnaire.

**Step 2: Establish a hierarchy structure**

The considered factors in underwriting process that selected from previous literature by this study in the 1st phrase are also evaluation factors for exploring an appropriated type of social media data, which comprise several level, including the goal hierarchy, criteria hierarchy, sub-criteria hierarchy and alternative hierarchy (see Figure 1).

**Step 3: Establish a pairwise comparison matrix**

Based on the opinion of experts to assign weight values, the geometric mean value is used to calculate comprehensive decision-making scores from experts. In doing so, the standard weight values can be established to select the most appropriate type of social media data. For instance, the main criteria are formed as the sample, as shown in Table 1. Formula (1) and (2) are used to calculate the aggregate pairwise comparison matrix.

**Table 1. Aggregation of the Pairwise Comparison Matrix for Criteria of Main Criteria**

<table>
<thead>
<tr>
<th>Level 2 Criteria</th>
<th>Physical Factors</th>
<th>Non-Physical Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Factors</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Non-Physical Factors</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

CI =0.00 ; CR = 0.00 < 0.1

**Step 4: Compute the eigenvalue and eigenvector**

The pairwise comparison matrix of the criteria and sub-criteria is used to obtain each hierarchical factor weight, in which the eigenvector is calculated by formula (3) and (4). Table 2 summarizes those results.

**Step 5: Perform the consistency test**

Based on formula (5) and formula (6), the pairwise comparison matrix of consistency is determined for each hierarchy, as show in Table 1. If the results of the six experts in terms of
consistency ratio and consensus of CR are smaller than “0.1” they conform to principles of consistency.

### Table 2. Weights of the Criteria and Sub-criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Criteria Weight</th>
<th>Sub-Criteria</th>
<th>Sub-Criteria Weight</th>
<th>Weights of Overall Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Factors</td>
<td>0.333</td>
<td>Age</td>
<td>0.090</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gender</td>
<td>0.090</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smoking</td>
<td>0.176</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Occupation and Hobbies</td>
<td>0.199</td>
<td>0.066</td>
</tr>
<tr>
<td>Non-Physical Factors</td>
<td>0.667</td>
<td>Physical Condition</td>
<td>0.106</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Health History</td>
<td>0.138</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foreign Travel</td>
<td>0.201</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Additional Insurance</td>
<td>0.127</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Financial Information</td>
<td>0.373</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Moral Hazard</td>
<td>0.272</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Morale Hazard</td>
<td>0.227</td>
<td>0.152</td>
</tr>
</tbody>
</table>

#### Step 5: Compute the relative weight of each hierarchy

Table 2 summarizes the results for the relative weight of the elements for each level. According to this table, the life insurance company selects an appropriated type of social media data based on the following rank: Physical Factors (0.333) and Non-Physical Factors (0.667). Evaluation results of the sub-criteria are summarized as Table 2.

#### Step 6: Calculate the whole level weight to select the most appropriate type of social medial data

In alternative hierarchy level, there are four types of social media data may provide the best insights on underwriting factors for insurers to make underwriting decisions. The most appropriate type of social media data is selected based on the highest score, in the following order: “Social graph” (0.393), “Purchase items online” (0.298), “Communicate with others in public forums” (0.175), and “Interact with multiple social networking sites” (0.134). Confirming that “Social graph” provides the most appropriate type of social media data based on the opinion of the experts from the viewpoint of underwriting effectiveness, as shown in Table 3.

### Table 3. Life Insurance Company Application of the AHP Model to Select an Appropriate type of Social Media Data to Improve the Effectiveness of Underwriting

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Criteria Weights</th>
<th>Social graph</th>
<th>Purchase items online</th>
<th>Communicate with others in public forums</th>
<th>Interact with multiple social networking sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Factors</td>
<td>0.333</td>
<td>0.396</td>
<td>0.299</td>
<td>0.169</td>
<td>0.136</td>
</tr>
<tr>
<td>Non-Physical Factors</td>
<td>0.667</td>
<td>0.392</td>
<td>0.297</td>
<td>0.178</td>
<td>0.133</td>
</tr>
</tbody>
</table>

Rank 1 2 3 4
CONCLUSIONS AND RECOMMENDATIONS

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

1. Life insurance underwriters prefer non-physical factors to physical factors searched on the social media networking sites. This is 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 identify insured’s physical condition and then makes the underwriting decision.

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.

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.

4. If underwriters want to search the useful information related to Financial Information, Moral Hazard, and Morale Hazard, the social media data type of “social graph” is the best choice. This is because “social graph” shows how individuals are linked together, providing a picture of who is friends with whom, who follows whom, and people’s friends of friends. In other words, social 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.

5. 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


