Text mining social media for competitive analysis

Minería de texto de los medios de comunicación social para el análisis competitivo

Germán Gémar
University of Malaga, Department of Economics and Business Administration, Campus El Ejido s/n. 29013, Malaga, Spain, ggemar@uma.es

José Antonio Jiménez-Quintero
University of Malaga, Department of Economics and Business Administration, 29013, Malaga, Spain, jajq@uma.es

Abstract

Social media are utilised widely. Companies increasingly use social media to communicate and interact with customers. Much information is thereby generated and is available to everybody, including competitors. Firms need to analyse what their customers say and interact with them. Using text mining tools, companies can know where they are in relation to their competitors and control the behaviour of these. Transforming text into data and data into knowledge can be vital to make the right decisions and improving the competitive strategy of companies.

This study used a text mining tool to analyse the primary social media sites, including Twitter, Facebook, LinkedIn, YouTube and others, with a focus on a sample of hotels. The dimensions analysed were sentiments, passion and reach. A dependence was found between several variables obtained through text mining and financial performance. The results indicate that analysis of social media using these techniques can be a method to improve financial performance.

Keywords: Competitive intelligence, social media, text mining, hotel industry, financial performance.

Resumen

Los medios de comunicación social se han implantado de una manera generalizada. En el caso de las empresas, el uso es cada vez más frecuente para comunicarse e interactuar con los clientes. Debido a este hecho, se genera gran cantidad de información, disponible para todos, aunque también para la competencia. Las empresas deben analizar lo que dicen sus clientes e interactuar con ellos, pero también con técnicas de minería de textos pueden saber qué lugar ocupan con respecto a su competencia. También pueden vigilar el comportamiento de sus competidores. Transformando texto en datos y datos en conocimiento se podrían tomar decisiones acertadas para mejorar la estrategia competitiva de la empresa. Este trabajo analiza mediante minería de texto los principales medios de comunicación social como twitter, facebook, linkedin, youtube y otros, en una muestra de hoteles. Se analizan en concreto dimensiones como sentimiento, pasión y alcance. Se encuentra relación entre algunas de las variables obtenidas mediante minería de texto y la rentabilidad financiera. Los resultados indican que el análisis de los medios sociales con estas técnicas debe ser tenido en cuenta para mejorar la rentabilidad financiera.

Palabras clave: Inteligencia competitiva, medios de comunicación social, minería de texto, sector hotelero, rentabilidad financiera.

1. Introduction

Consumers habitually use the Internet to read news and other information, as well as to buy products and services. Currently, many consumers are using platforms to share information: the phenomenon of social media.

Social media can affect a company’s reputation. They can influence sales and even a company’s survival. However, many business executives avoid or ignore such means of communication, because they do not understand what social media and their possible future effects are. Kietzmann, Hermkens, McCarthy and Silvestre (2011) recommend that companies formulate strategies for monitoring and understanding social media.

Social media is increasingly present in the decisions of most business executives today. Therefore, researchers are trying to identify ways to make profitable use of applications such as Twitter, Facebook, YouTube or LinkedIn. Kaplan and Haenlein (2010), for example, focus on social media opportunities.

Social media have become important for communication and exchanges of information, but the material generated in these sources will be lost if not used. Asur and Huberman (2010) have demonstrated how social media content can be used to predict real-world outcomes. In particular, they have shown how feelings extracted from Twitter can be used to improve the predictive power of social media.

In particular, social media are playing an important role as sources of information for travellers. Xiang and Gretzel (2010) researched social networks and found that these constitute a substantial source of information. Their research confirms the growing importance of social media in online tourism.

2. Literature review

Competitive intelligence has been present ever since the first businessperson began operations and was observed by a second businessperson who wanted to know what he was selling and to whom. It has existed since the beginning of civilisation. Sun Tzu and many other ancient warriors realised that it was an essential part of the art of war (Tzu, 2013).

Today, we are in the age of intelligence, an intangible asset to companies seeking information to examine and use to gain a competitive advantage. These companies value the minds of their employees as the most important assets of their organisations (Jaworski & Wee, 1993).

The literature and evolving management practices reveal that the analysis of competitors takes a variety of forms and becomes critical to developing an appropriate context and
Wide perspective. Although it is not an aim of this study to provide a comprehensive review of the term “competitive intelligence”, it is still appropriate to present some of the most representative definitions. In this regard, McGonagle and Vella (2002) argue that this is a process where public resources are used to develop data on competitors and the environment, transforming these data into information that can support business decisions. In addition, Calof and Wright (2008) state that competitive intelligence is an environmental analysis system that integrates the knowledge of all members of an organisation, covering, among other things, marketing and organisational strategy. According to Agarwal (2006), competitive intelligence – defined as focused and coordinated competition in a particular market control – plays a key role in knowledge management and decision making processes of organisations, permitting vision and anticipation of competitors’ activities.

Competitive intelligence is key to strategic management because the environment places significant limits on organisations. The competitiveness of organisations depends on their ability to adopt strategies developed through competitive intelligence activities, adjusting to their changing environment. Therefore, as Nasri (2011) writes, competitive intelligence provides information about competitors and their strategies, objectives, research, strengths, weaknesses and so on, giving companies the opportunity to understand their position relative to competitors.

Gray (2010) indicates that competitive intelligence necessarily deals with:

1. Emerging technologies
2. Changes in consumer preferences
3. The opening of new markets
4. Changes in demographics
5. Changes in legislation

These aspects affect businesses and competition. Moreover, the same author writes that competitive intelligence should not be confused with the concept of “competitor intelligence”, because the latter refers to only one competitor.

As a basic tool in the decision making process of organisations, competitive intelligence is a must for companies, whatever the sector in which they are active. Jaworski, Macinnis and Kohli (2002) establish a framework for competitive intelligence formed of three interdependent phases:

1. Organisation
2. Information search
3. Decision making

In addition, the success of each of the phases depends on the following related factors:

1. Intelligence network
2. The business environment
3. Information environment

In any case, searching for information is key to competitive intelligence. However, as Adidam, Gajre and Kejriwal (2009) assert, there may be significant differences, at the international level, in the development of competitive intelligence, especially in data collection. In particular, the historical and geographical background of each country affects the behaviour of people, when investigating environmental factors.

The specific objectives of competitive intelligence are basically to (1) manage and reduce risk, (2) achieve profitable knowledge, (3) ensure information security and (4) avoid information overload (Myburgh, 2004). This means that managers need to understand their customers and competitors better to build a competitive advantage. Therefore, a vital component in strategic planning and management is to gather a variety of information, which allows companies – by analysing the strengths and weaknesses of competitors – to anticipate market developments, as they can predict what will happen in their competitive environment.

Researches on competitive intelligence are not as common in the literature as compared to other areas of knowledge. However, some contributions include Faust and Gadotti (2011), who studied the presence of competitive intelligence tools in the hotel sector in Brazil (critical success factors, benchmarking, balanced scorecard and so on), and Jiménez, Arroyo and Reis (2008) and Nasri (2011), who focused on the presence of competitive intelligence in Tunisian companies in various sectors. The role of structural capital in the development of competitive intelligence was studied by Zangouinezhad and Moshabaki (2009), in a large sample of both public and private firms in Iran. Heppes and du Toit (2009) did the same in the banking sector in South Africa, and Tarraf and Molz (2006) examined this in small multimedia companies based in Canada.

However, competitive intelligence is linked incrementally with strategic intelligence and with knowledge management (McGonagle & Vella, 1999). From this point of view, intelligence can help in areas of business where managers want a sustainable competitive advantage. Therefore, we have chosen to rank competitive intelligence higher than product innovation and process, to become part of organisational thinking. Different authors who cannot conceive of an intelligence that is not strategic and therefore competitive have already accepted this vision.

Companies today are adopting and implementing systems intelligence to find strategic needs. Thus, the intelligence that organisations often require concerns strategic events such as long-term planning, capital investment and technology issues. This is because competitive intelligence acts as a filter, creating new opportunities and helping managers to detect negative situations. In fact, companies practice better intelligence gathering when they are trying to be innovative.

In addition, competitive intelligence is aligned with technological surveillance, which can convert smart environment information into knowledge and enable
decision making processes to anticipate change. As this structure requires clear processes, we can specifically point to the information reflected in management systems standards, in our case, the UNE 166006 EX (AENOR, 2006). This standard aligns with and supports others, such as ISO 9001:2008, ISO 14001:2004 and UNE 166002:2002.

Subsequently, it is relevant to analyse the logical cycle necessarily followed by a formal competitive intelligence system, in order to give this tool strategic value. The process of obtaining intelligence starts from the most basic unit: the data. Once organised, the data is converted into information; information, when analysed, becomes intelligence.

Before carrying out the intelligence process, members of the organisation must agree about what is data, information, intelligence and knowledge. If managers and intelligence professionals do not pinpoint these basic definitions, the staff will misunderstand their objectives, which can lead to an ineffective intelligence process. Based on this model, professional intelligence usually runs the process or cycle through four phases:

1. Intelligence requirements for making key decisions are identified. To do this, key decision makers are pinpointed within the organisation, along with their particular intelligence needs.
2. Information is collected – from printed, electronic and oral sources – about events in the environment of the company.
3. This information is analysed and synthesised.
4. Intelligence derived from this process is distributed to people who make decisions in the organisation.

Competitive intelligence systems (CIS) are closely related to knowledge management systems (KMS). The latter show an interest in the internal environment of companies, but CIS also seek to understand the competitive challenges of the companies’ environment (see Figure 1).

![Figure 1 - Connections between competitive intelligence and knowledge management](source: Authors)

Competitive intelligence helps companies to:

1. Gain a better understanding of their business environment and industry
2. Learn about corporate and business strategies of competitors
3. Forecast opportunities and threats
4. Anticipate the research and development of competitors’ strategies
5. Validate or deny industry rumours
6. Take effective decisions
7. Act instead of react

Therefore, it reduces the risk inherent in selecting organisations’ competitive strategies.

To gain this advantage, competitive intelligence has to work in parallel with knowledge management. Barclay and Kaye (2000) predict that the management of knowledge and intelligence functions will converge. In turn, this convergence serves to strengthen and optimise relationships with customers. Summarising their findings, Ruizalba and Vallespin (2014) suggest that three dimensions should be considered:

a. The generation of intelligence through data collection in companies’ domestic markets
b. The communication of information on the internal market and diffusion within organisations
c. The companies’ response to intelligence on the internal market.

Until now, companies have needed to work with information obtained from traditional sources. Currently, these traditional sources of information are being complimented with the chaotic world of open social platforms. According to Harrysson, Metayer and Sarrazin (2012), to survive in this new environment, companies will need to develop new skills and the willingness to participate in social conversations, rather than just collect information. Table 1 shows the new analytical tool kit.
Therefore, competitive intelligence is one of the main factors in successfully maintaining or improving the competitiveness of firms. However, the information necessary for competitive intelligence is often limited by the lack of sufficient sources of data about competitors. However, currently, with the emergence of the Web 2.0 concept and the large amount of customer feedback, information about competitors is available and has become a new source from which to mine competitive intelligence. Xu, Liao, Li and Song (2011) found that analysis of this data generated by consumers has become critical.

McGonagle and Vella (2002) researched competitive intelligence and concluded that 90% of the information a company needs to understand its market and competitors and to make key decisions is already public. This suggests that the use of public sources of information for organisations and for monitoring the competition is open to everybody.

According to Lau, Lee and Ho (2005), text mining explores data in text files to establish valuable patterns and rules that indicate trends and significant features about specific topics. The Internet generates a large amount of information. Online text mining represents a rich reservoir of business intelligence that is potentially valuable to companies. Online text mining involves four steps:

- Definition of mining context and concepts
- Data collection
- Dictionary construction
- Data analysis

Companies can use text mining tool to convert online data into competitive intelligence.

3. Methodology

Maintaining or improving the return on equity (ROE) is essential for any business, including hotel companies (Gémar & Jimenez, 2013). Occasionally, the best way to achieve this goal is internationalisation, taking into account cultural distance to choose which entry mode to use in the destination country (Gémar, 2014). However, sometimes competitive intelligence can use simple tools to improve ROE.

This study aims to investigate whether competitive intelligence through text mining can effectively help to improve the ROE of companies, specifically, in the hotel industry of Malaga, Spain. The data were obtained from SABI (the Iberian Balance Sheet Analysis System). SABI is a database with information on all Spanish and Portuguese companies, which facilitates searches by different criteria (company name, VAT code, location, activity, financial data, market data, cartographic location and so on). This study was done on hotels with at least a one million euro operation revenue. The sample consisted of 83 hotels.

SABI was examined for each hotel’s age, operating revenue, number of employees, economic profitability, financial profitability, overall liquidity, debt collection period (days) phase, credit period (days) and employee costs/operating revenue (%). Most variables were ruled irrelevant or highly correlated. The main variable under study was costseor.

In parallel, a study of text mining of social media was made using the tool provided by socialmention.com (2014). Social Mention returns indicators related to the social reputation of brands.

This allowed us to track and measure easily what people are saying about people, companies, new products or any topic across the web’s social media landscape in real time. Social Mention monitors 100+ social media properties directly, including Twitter, Facebook, LinkedIn, YouTube and Google. This tool measures social media’s influence using four parameters: strength, sentiment, passion and reach.
“Strength” is the likelihood that a brand is being discussed in social media. A very simple calculation is used: Phrase mentions within the last 24 hours are divided by the total possible mentions. This dimension is not studied in this paper. “Sentiment” is the ratio of mentions that are generally positive to those that are generally negative.

“Passion” is a measure of the likelihood that individuals talking about a brand will do so repeatedly. For example, if a company has a small group of extremely passionate advocates who talk about its products or brand all the time, the company receives a higher passion score. Conversely, if a different author writes every mention, the company gets a lower score.

“Reach” is a measure of the range of influence. It is the number of unique authors referencing a brand, divided by the total number of mentions.

Other interesting data that this tool throws out for companies are:

- “Feelings” (positive, neutral and negative): great for companies to see entries, give and click signs and excellent especially for checking negative feelings
- “Top keywords”: very useful to check which keywords have more impact on searches and references to a brand
- “Top users”: the users mentioning a company the most, which is very useful to work with users and encourage loyalty
- “Top hashtags”: main hashtags associated with brand mentions
- “Sources”: the main sources of brand mentions

In addition, the data includes a list of all references in which companies are included.

This study’s aim was to analyse if the ROE depends on the three dimensions described above: sentiment, passion and reach. The analysis was done with the variable: employee costs/operating revenue (%). A descriptive analysis of the data and a correlation study of the variables were performed using the Statistical Package for Social Sciences version 15.0 (SPSS 15.0) for Windows. A correlation analysis between variables was made, and a significant correlation at the 0.05 level (bilateral) between sentiment and passion variables was obtained.

For the econometric analysis, the multiple linear regression model was used to predict the value of the dependent variable, and coefficients of the independent variables that best predict the value of the dependent variable were estimated.

4. Results

Pearson Correlation measures the strength and direction of the linear relationship between the two variables. From the scatterplot of the variables sentiment and passion below, we can see that the correlation is positive (see Table 2).

Table 2 - Correlations

<table>
<thead>
<tr>
<th></th>
<th>Sentiment</th>
<th>Passion</th>
<th>Reach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment</td>
<td>Pearson correlation 1</td>
<td>0.254(*)</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed) 0.020</td>
<td>0.829</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N 83</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Passion</td>
<td>Pearson correlation 0.254(*)</td>
<td>1</td>
<td>-0.193</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed) 0.020</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N 83</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>Reach</td>
<td>Pearson correlation -0.024</td>
<td>-0.193</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed) 0.829</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N 83</td>
<td>83</td>
<td>83</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).

Source: Authors.

The f-ratio in the ANOVA table (see Table 3) tests whether the overall regression model is a good fit for the data. The table shows that the independent variables predict the dependent variable in a statistically significant way: $F(3.79) = 3.361, p < 0.05$ (i.e. the regression model is a good fit for the data).

Table 3 - ANOVA (b)

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Regression</td>
<td>147,479,267</td>
<td>3</td>
<td>49,159.756</td>
<td>3.361</td>
</tr>
<tr>
<td></td>
<td>Residual</td>
<td>1,155,505.318</td>
<td>79</td>
<td>14,626.650</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1,302,984,585</td>
<td>82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Predictors: (constant), costs, reach and passion
(b) Dependent variable: ROE

Source: Authors.

Unstandardised coefficients indicate how much the dependent variable varies with an independent variable, when all other independent variables are held constant. Consider the effect of passion in this study. The unstandardised coefficient, B1, for passion is equal to -1.583 (see Table 4). This means that for each 1% increase in passion, there is a decrease in ROE of 1.583%. This negative relationship shows that a different person must write every mention to obtain an improved ROE.

We can test for the statistical significance of each of the independent variables. This tests whether the unstandardised (or standardised) coefficients are equal to 0 (zero) in the population. If $p < 0.10$, you can conclude that the coefficients are different at a statistically significant level to 0 (zero). The t-value and corresponding p-value are located in the “t” and “Sig.” columns, respectively, as shown in Table 4.
If \( p < 0.10 \), we can conclude that the passion and costseor coefficients are different to 0 (zero) at a statistically significant level. ROE depends significantly on the passion and costseor variables.

The assumptions of the linear regression model were checked. (1) For linearity, partial regression graphs were generated. They demonstrated that, for each independent variable, the relationship is linear if the effects of other independent variables are eliminated. (2) For independence, the Durbin-Watson statistic was applied, resulting in a value of 2.155. We can assume independence between residues when Durbin-Watson values fall between 1.5 and 2.5. (3) For homoscedasticity or equal variances, we used a forecast dispersion diagram established by standardised residuals but not unequal variances. (4) Normality was checked by finding graphic standardised residuals. Finally, (5) non-collinearity was confirmed, as indices of conditioning were not identified above 30, with variance ratios above 90%.

5. Conclusions

The employee costs/operating revenue (%) ratio is the variable that most affects all variables analysed. This ratio depends on companies’ management; therefore, we conclude that good management is critical. The monitoring of this ratio is essential for hotels to improve their financial performance.

However, there are other ways to improve financial performance. Using competitive intelligence – with free text mining tools that analyse social media information on companies and their competition – can produce conclusions that help companies to make decisions that will improve their competitive advantage.

In social media, companies should look for passion. Passion is a measure of the likelihood that individuals talking about a brand will do so repeatedly. If a company has a small group of extremely passionate advocates who talk about its products or brand all the time, the company gets a higher passion score. Conversely, if a different author writes every mention, the company receives a lower score.

However, a lower score is important in obtaining an improved ROE. Conversely, different authors must write every mention to result in an improved ROE.

Companies and individuals increasingly use social media as sources of intelligence. For businesses, this use is increasing as they communicate and interact with customers. Thus, much information is generated and is available to everybody, including to competitors. Companies should analyse what their customers say and interact with them online. Using text mining tools, companies have plentiful information available. Transforming text into data and data into knowledge is important to make the right decisions and to improving companies’ competitive strategy.

This paper discusses how, through competitive intelligence – specifically, text mining in major social media – companies can obtain valuable information. The analysis of social media using competitive intelligence tools (text mining or others) is important as a new way to improve the financial performance of companies. For hotel companies, analysis of social media data on companies and their competitors and customers gives much relevant information that can be used to improve hotels’ competitive position and financial performance.

This paper contributes to the field by using text mining to perform competitive analysis on companies’ generated data on socialmention.com. Social Mention monitors 100+ social media properties directly, including Twitter, Facebook, LinkedIn, YouTube and Google.

References


Table 4 - Coefficients (a)

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardised coefficients</th>
<th>Standardised coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. error</td>
<td>Beta</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>197.133</td>
<td>65.788</td>
<td>-0.194</td>
</tr>
<tr>
<td></td>
<td>Passion</td>
<td>-1.583</td>
<td>0.885</td>
<td>-0.194</td>
</tr>
<tr>
<td></td>
<td>Reach</td>
<td>-1.240</td>
<td>1.444</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>costseor</td>
<td>-2.375</td>
<td>0.876</td>
<td>-0.288</td>
</tr>
</tbody>
</table>

(a) Dependent variable: ROE

Source: Authors.


**Article history:**

Received: 13 May 2014
Accepted: 14 November 2014