MinEDec: a decision support model that combines text mining with competitive intelligence

Yue Dai, Tuomo Kakkonen, Erkki Sutinen
School of Computing
University of Eastern Finland
Joensuu, Finland
{yvedai, tkakkone, sutinen}@cs.joensuu.fi

Abstract

In order to monitor and analyse the competitive environment of businesses, we propose to integrate the Five Forces framework and a SWOT analysis in a decision support system model called MinEDec (Mining Environment for Decisions), which is supported by text-mining technologies.

1. Introduction

Changes of environment redefine the way in which business enterprises compete and make decisions. Leaders in modern enterprises deal with decisions that require the integration, analysis, and summarization of both internal and external information from a variety of sources. Information technologies, such as enterprise-wide systems, data mining, and *text mining* (TM), can help capture and integrate transactions from a variety of perspectives to support decisions [1, 2].

Competitive intelligence (CI) is a process of monitoring the competitive environment by pulling together data and information from a very large and strategic perspective, to predict or forecast what is going to happen in the competitive environment of an enterprise [3]. Consequently, more and more enterprises are using CI analysis methods, such as competitive positioning analysis, benchmarking analysis, *Five Forces analysis* and *SWOT analysis*, in support of their decision-making processes [4].

A *decision support system* (DSS) helps leaders to make decisions that are unique, rapidly changing, and not easily specified in advance [5]. The target of our research is to establish a decision support system that leverages TM technologies, SWOT analysis, and the Five Forces framework to search and analyse unstructured data (e.g. newspapers, customer feedback, reports, and email). By providing the ability of CI analysis, the DSS can seize early warnings of threats and opportunities in the business environment, which are necessary for the proactive strategy of enterprises. In order to create such a system, we need to have a clearly defined model for decision support that functions as the basis of the system design. This paper describes the model we have designed.

In Section 2, the background of the current study is described. Section 3 presents the Five Forces framework, which is extended with SWOT analysis. Section 4 outlines the *Mining Environment for Decisions* (MinEDec) model applied to SWOT analysis and the Five Forces framework. The paper concludes in Section 5 with final remarks.

2. Background

2.1. The Five Forces framework and MinerVa

It is necessary for leaders in any enterprise to understand the competitive forces in their industry since these will determine the likely successes or failures of particular enterprises within it [1]. The Five Forces framework is a CI analysis model, which was developed for such business environment analyses (Fig. 1) [6].

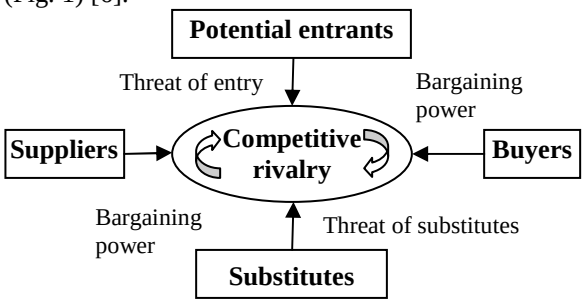

Figure 1. The Five Forces framework [6]

As illustrated in Fig. 1, *rivals*, *potential entrants*, *substitute products*, *suppliers*, and *buyers* are the five basic parties in a competitive environment. The threat of entry means that new entrants will add capacity to the industry and increase the demand and prices, resulting in lower industry profitability. The threat of substitutes describes the risk of market displacement from existing or potential substitutes. The bargaining power of suppliers refers to the ability of suppliers to influence the cost, availability, and quality of input materials. The bargaining power of buyers means the buyers can influence the industry structure, such as prices and quality expectations [1, 4, 6]. Understanding and focusing on these five subjects affect both the profit potential and the prospects for achieving competitive advantage.

MinerVA is a decision support model that we introduced in [7]. MinerVA is based on the Five Forces framework and it integrates three advanced TM technologies – opinion mining, event change detection, and patent trend change mining to monitor the external business environment. The MinEDec model presented in this work is based on MinerVA.

2.2. SWOT analysis

SWOT analysis is a CI method for planning a future-oriented strategy [1, 8]. A SWOT analysis is used to evaluate threats and opportunities from a turbulent environment; it summarizes the strengths and weaknesses of a company to address the issue that the company is facing or will face, and finds a proper strategic plan. The output of the SWOT analysis, called a *SWOT matrix*, is shown in Table 1.

Table 1. SWOT matrix [9]

Strategy *SW* / *OT*	**Strengths** S1 S2	**Weaknesses** W1 W2
Opportunities O1 O2	SO strategy S1O1, S1O2 S2O1, S2O2	WO strategy W1O1, W1O2 W2O1, W2O2
Threats T1 T2	ST strategy S1T1, S1T2 S2T1, S2T2	WT strategy W1T1, W1T2 W2T1, W2T2

As illustrated in Table 1, the decision makers need to search the environmental changes and recognize the factors in the four categories. The strengths and weaknesses are about the internal environment of the company, compared with the competitors of the company, and generally reflect the company's technology, equipment, personnel, products, markets, management structure, and so on. The opportunities and threats refer to the external environment factors, which are favourable or unfavourable to the enterprise. Favourable factors include high technology and a good relationship between buyers; adverse factors include trade policy changes, unexpected events, market changes, and the emergence of competitors.

Four types of strategies can be defined from the SWOT matrix. The SO strategy (positive strategy) is to use strength points in order to make use of the opportunities; the WO strategy (differentiation) is to diminish the weak factors by grasping the opportunities; the ST strategy (gradual) is to use a strength to reduce the threat factors; and the WT strategy is also called the negative or withdrawal strategy, which uses defensive approaches to cover the weaknesses and to avoid the threats [9, 10].

2.3. Text-mining technologies

TM refers to the process of deriving high-quality information from texts [11]. It is based on the theoretical foundation of the computational linguistics on one hand and mathematical statistics and data analysis on the other. The technologies used in TM include: *information retrieval* (IR), *information extraction* (IE), topic tracking, summarization, categorization, concept linkage, [1] information visualization, and question answering [11, 12].

The main advantages of using TM technologies for CI are the ability to process large amounts of textual data quickly and the objectivity and customizability of the process. The TM process typically includes the following steps [13]:

1) preprocessing of the data to the required format for further analysis (data processing);
2) extraction of important concepts and terms through initial text analysis (concept extraction);
3) using a narrative analysis to identify patterns and co-occurrences of identified concepts (narrative analysis);
4) developing an automated solution (automatic categorization); and
5) building a taxonomy of concepts.

3. Integrating SWOT and Five Forces analysis

3.1. From data to knowledge

Data are strings of symbols, facts, measurements, and statistics. Data with context equals information, information with meaning can be intelligence, and intelligence with experience generates knowledge.

[1] Concept linkage tools connect related documents by identifying their shared concepts, helping users find information they perhaps would not have found through traditional search methods.

Intelligence and/or knowledge are the basis for making decisions, and they must be in a useful format to meet the strategic needs of enterprises. In order to transfer unstructured data into strategic knowledge, we need to combine CI analysis and IT technology during the transformation process (Fig. 2) [14, 15].

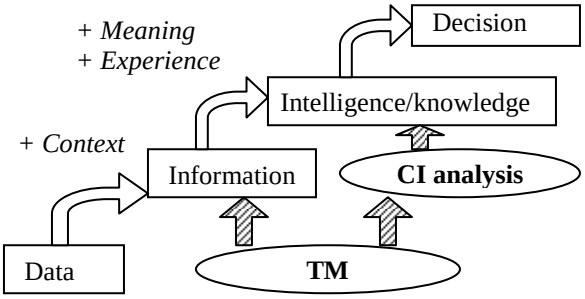

Figure 2. From data to decision supporting by TM and CI

Fig. 2 explains the purpose of our research. As illustrated in Fig. 3, the technologies of TM, such as IR and IE, can be used to search and summarize unstructured data by adding context, and TM can also support CI analysis models. *Enterprise Miner*, which was developed by SAS, uses clustering to mine document sets in order to perform competitor analysis [3]. CI analysis models can extract intelligence/knowledge by adding meaning and experience to information. From TM and CI, leaders can obtain intelligence from countless unstructured data, so leaders can make decisions more easily with the help of TM and CI analysis models.

3.2. Using text mining to support SWOT analysis and the Five Forces framework

As the background in Section 2 showed, TM technologies have the potential to support CI and decision making. However, what is lacking is an integrated framework that can provide the objectives to analyse and summarize from the perspective of CI the huge amount of available textual data.

The Five Forces framework is accepted by most researchers in CI, and it provides clear objectives and a systematic approach to identify and analyse the relevant trends and events and the influence of each of the five factors, not only within themselves but across the other forces. SWOT analysis can help greatly in understanding the business environment and assessing future courses of action as well as generating strategic options [1, 4]. The use of the Five Forces framework provides several advantages. First, it guides information collection, because it helps to identify a set of analytical subjects and make leaders aware of what data and information to look for. Second, it makes other CI analysis models more efficient and focused, such as SWOT analysis [4]. Moreover, the five objectives defined the framework, and they are the most important components of the business environment, so leaders must keep monitoring the situation and their trends. In order to integrate effectively the two CI analysis models – Five Forces and SWOT – into one unified decision support model, we need to consider carefully the factors of each model. We integrate all the factors in Table 2 and the entire factors are the keywords of IR and IE in MinEDec [16, 17].

As illustrated in Table 2, there are five factors for evaluating internal strengths and weaknesses: technology, price, equipment, service, and attitude. The leader can obtain the result about one factor for one objective; he can also use all the factors to evaluate the situation of each objective. For example, by tracking the customers' behaviours, mobile communication companies need to provide real-time news to their customers [16]. The factors of external opportunity and threat are more general than internal factors, and they can indicate trends that are very important for a proactive strategy. Political shifts will give a chance to potential entrants, which would be a threat to one's own company. Economic shifts can increase the power of buyers and suppliers, which is an opportunity. Technological shifts may be conducted by rivals, and are also a threat. Social shifts mean the consumers' attention, for example, by finding out that the 3G service has became a hot topic in Taiwan, telecommunication companies should launch a 3G market, which is an opportunity [16].

Table 2. The major factors used for both SWOT analysis and the Five Forces framework

Example *SWOT* / *Objectives*	**Strengths**		**Weaknesses**			**Opportunities**		**Threats**	
	Technology	Price	Equipment	Service	Attitude	Political Shifts	Economic Shifts	Technological Shifts	Social Shifts
Rivals	[16], [17]							[16]	
Buyers					[16]				[16]
Suppliers									
Substitutes									
Potential entrants									

4. MinEDec decision support model based on TM technologies

As analysed in the previous section, the functions of our model are clear and well established. They all play an important role in supporting strategic decision making. Hence, we propose a model that analyses the business environment combining the Five Forces framework with SWOT analysis by using TM technologies (Fig. 3).

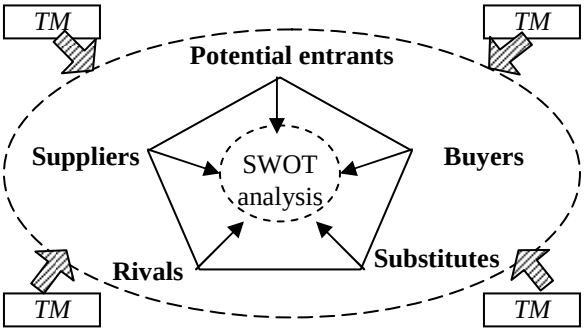

Figure 3. Graphical framework of MinEDec

Figure 3 outlines a graphical framework of our decision support model – MinEDec. In order to meet the strategic requirements of decision making, leaders need to be effectively guided through the collection and analysis of information. Consequently they are supported by IT technology, such as TM.

For rivals, TM technologies can be used for tracking the extent of competitive rivalry in the industry, the rivals' products, strategic drift (such as developing an international market in a specific country), and the development of technology and the trends of the whole industry. Furthermore, we can establish a profile for each rival and then use SWOT analysis to show the differences between one's own company and the rival. A detailed SWOT can, for example, be a comparison of own products vs. the rival's products, or own service vs. the rival's service. A general SWOT model can be summarized from a detailed SWOT and used to generate reports about the whole situation of one's own company, including the product, price, service, etc. The decision between the uses of a general or detailed SWOT analysis depends on the strategic needs of the leaders.

For buyers, TM can be used to detect the attitude of our ultimate customers and immediate customers, and track the behaviours of buyers from these factors: buyer volume tracking from internal data, brand identity, buyers' concentration, price sensitivity tracking from the Internet, the threat of backward integration tracking the buyers' activities from newspapers, etc. For substitute products, we can find out developments in related industries and buyer inclination to substitute in order to be aware of the threat in time. In order to monitor suppliers, TM can be applied to discover the supplier concentration, the impact of materials on cost or differentiation from opinion mining on the Internet, the presence of substitute material tracking on the Internet, and the threat of forward integration tracking the suppliers' activities from newspapers. For potential entrants, the tasks of TM are monitoring the changes in a material's demand and increased price for input [3] [4].

What we want to achieve is to give leaders sufficient support to make a decision. We provide a detailed and general SWOT model for each objective. For example, when leaders want to know the customers' general information, the system could provide the present and potential customers' profile of the company; when managers want to know the buyers' attitude, they can obtain SWOT intelligence from summarizing the buyers' attitude toward their own products, rivals' products, and substitute products. Furthermore, considering the fact that the factors that influence each objective are interrelated, the system can summarize the whole competitive environment from all the factors. These factors can also be used in the SWOT matrix; the system can give suggestions about choosing a strategy (SO, WO, ST, or WT) by combining different factors.

Fig. 4 outlines the architecture of our DSS based on the MinEDec. Decision makers first need to set up the aim of the mining work. Then TM can proceed to analyse the preprocessed data, extract information, and provide useful intelligence reports about the business environment. There are three major processes in the model: text data collecting, TM, and intelligence analysis. We will outline them in the subsequent sections.

4.1. Text data collecting

Data collection is an important foundation of a DSS. The data sources must be diverse, for example, we can use Yahoo, some financial news sites, relevant industry association sites, RSS feeds, and rivals' press releases as the external data sources. For internal text information, the data sources are e-mail, reports, and notes, which are generated from other information systems and workflow. The system utilizes intelligent search engines to collect information automatically. Because we have already defined the five objectives, some specific data sources will be paid more attention and made sure they are kept up to date in time.

4.2. Text mining

During the process of TM, there are three major components: preprocessing, text/data warehouse, and information processing. Because the formats of text are varied – txt, doc, pdf, html, rtf, xml, etc. – we need to identify and convert them to a unified format that can be processed by the system. Furthermore, the words compose sentences, and sentences compose the content of the text according to syntax rules, so we need to preprocess the text as a preparation process for TM. During preprocessing, we can apply natural language processing methods such as stemming, stop words removal, and part-of-speech tagging to decompose text into some meaningful language segments with tags that can be easily extracted. We also use feature extraction and IE to extract information through filtering redundant information. Feature extraction is to extract keywords, which are explained in Section 3 (such as technology, products, attitude, etc.) or other descriptions of these keywords. IE is to extract information about pre-specified types of events, entities, or relationships [18]. In our system, we focus on the events and relationships of the five objectives. The results of the preparation are organized into the text/data warehouse.

After the features and information have been stored in the warehouse, clustering, classifying, association rule analysis, and summarization can be used to process them into meaningful information.

Clustering and classifying are used for information exploitation and retrieval. For example, we can use clustering to analyse emails from buyers to find out common keywords that we have not noticed. Association rule analysis discovers the relationships and changes between two or more data sets [16, 18]. In our system, it can be used to establish the relationship and trends between different objectives, technologies, products, events, etc. Summarization extracts key information and gives a more concise content of the whole text.

We use the domain knowledge base to store ontological information. The design of the company, product, and event (CoProE) ontology arises from the need to have an ontology capable of supporting the domain knowledge needs of our system, and it is based on a standard product and industry type code set. The current version of the ontology enables the characterization of business entities and their key employees, and the classifications in a standardized way of one's own products as well as the modelling of business events and the relations between companies. The CoProE can support the domain knowledge needs of our system [19].

4.3. Intelligence analysis

By using TM, managers already gain intelligence about specific objectives. However, we still give further support to them. These additional intelligence/knowledge products are dependent on the IE technology, such as named entity recognition (NER), template relation (TR), and the scenario template (ST).

NER refers to the recognition of proper nouns, time, and number, etc. TR deals with finding out the relationships between the entities. ST forms a description of the whole event or relationship by connecting entities together [3, 9]. Considering that the expression forms of the same factors are different under different situations, we can cluster and label the keywords, then combine related labels to generate factors, and the factors are displayed in terms of strengths, weaknesses, opportunities, and threats by comparing with rivals or other subjects [10]. Furthermore, we can combine factors in a SWOT matrix to give suggestions for planning a strategy.

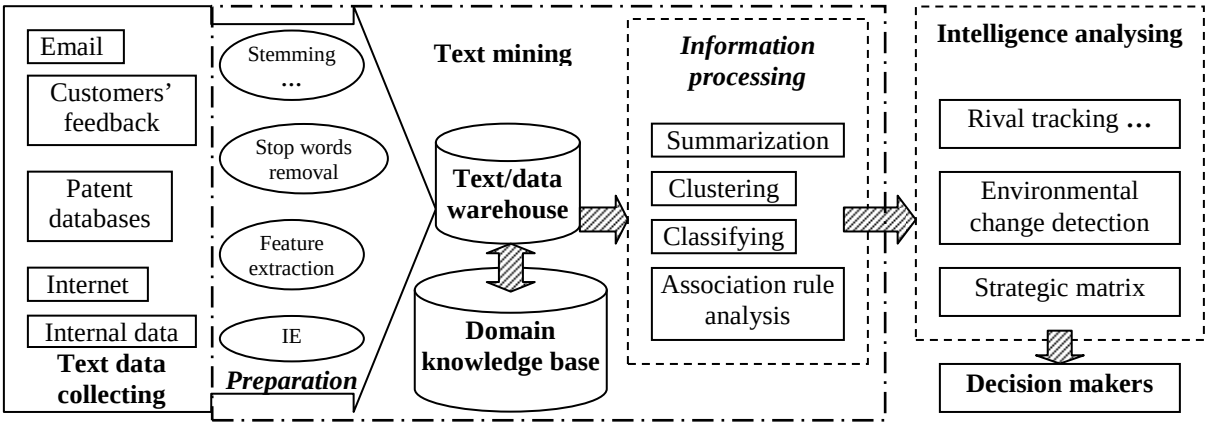

Figure 4. The architecture of a system based on MinEDec, a decision support model

5. Conclusion

In this paper, we have proposed a decision support model that combines two well-known and widely used CI analysis models into a unified model. We have also outlined a DSS that is based on the MinEDec model and applies TM technologies.

First, we explained that the purpose of our MinEDec model is to transform data into useful knowledge. We then described the functions of SWOT analysis and the Five Forces framework in a new model for monitoring the business environment. Based on this, we have proposed a way of integrating SWOT and Five Forces analysis models into a unified decision support model. We find that MinEDec can support decision making better. The capability of the proposed DSS in terms of monitoring the five force parties and deeply analysing the competitive environment is worth our attention.

Future work will focus on developing a working system based on the proposed MinEDec model. After system testing, MinEDec will be evaluated against the proposed functions in real business environments.

6. Acknowledgement

This work was partly supported by grants from the CSC – China Scholarship Council, CIMO – Center for International Mobility, and the "Towards e-leadership: higher profitability through innovative management and leadership systems" project, which is funded by the European Regional Development Fund and Tekes – the Finnish Funding Agency for Technology and Innovation.

References

[1] G. Johnson, K. Scholes, and R. Whittington, *Exploring Corporate Strategy: Text & Cases*, FT Prentice Hall Financial Times, Harlow, U.K., 2008.

[2] V. L. Sauter and D. Free, "Competitive intelligence systems: qualitative DSS for strategic decision making", *The DATA BASE for Advances in Information Systems*, vol. 36, no. 2, 2005, pp. 43–57.

[3] R. Bose, "Competitive intelligence process and tools for intelligence analysis", *Industrial Management and Data Systems*, vol. 108, no. 4, 2008, pp. 510–528.

[4] C. S. Fleisher and B. E. Bensoussan, *Business and competitive analysis*, Pearson Education, Inc., Upper Saddle River, New Jersey, 2007.

[5] K. C. Laudon and J. P. Laudon, *Essentials of Management Information Systems: Managing the Digital Firm*, Prentice Hall, London, U.K., 2004.

[6] M. E. Porter, *On Competition*, Harvard Business Press, Boston, USA, 1998.

[7] Y. Dai, T. Kakkonen, and E. Sutinen, "MinerVA: a decision support model that uses novel text mining technologies", *The 4th International Conference on Management and Service Science*, Wuhan, China, 2010.

[8] D. H. Kim, B. J. Ahn, J. H. Kim, and J. J. Kim, "The strategic approach using SWOT analysis to develop an intelligent program management information system (iPMIS) for urban renewal projects", *Fourth International Conference on Computer Sciences and Convergence Information Technology*, Seoul, Korea, 2009.

[9] H. Y. Shahir, S. Daneshpajouh, and R. Ramsin, "Improvement strategies for agile processes: a SWOT analysis approach", *Sixth International Conference on Software Engineering Research, Management and Applications*, Prague, Czech Republic, 2008.

[10] M. Samejima, Y. Shimizu, M. Akiyoshi, and N. Komoda, "SWOT analysis support tool for verification of business strategy", *International Conference on Computational Cybernetics*, Tallinn, Estonia, 2006.

[11] J. P. M. D. Oliveira, S. Loh, L. K. Wives, "Applying text mining on electronic messages for competitive intelligence", *Proceedings of the 5th International Conference on Electronic Commerce and Web Technologies*, Zaragoza, Spain, 2004.

[12] D. Sullivan, "The need for text mining in business intelligence", Information Management Special Reports, 2004, available at: http://www.information-management.com/specialreports/20040210/8100-1.html.

[13] R. Bose, "Advanced analytics: opportunities and challenges", *Industrial Management & Data Systems*, vol. 109, no. 2, 2009, pp. 155–172.

[14] T. P. Liang, *Decision Support Systems and Intelligent Systems*, Pearson Education International, Upper Saddle River, New Jersey, USA, 2005.

[15] F. Bouthillier and K. Shearer, *Assessing Competitive Intelligence Software: A Guide to Evaluating CI Technology*, Information Today, Inc., Medford, New Jersey, USA, 2003.

[16] D. R. Liu, M. J. Shih, C. J. Liau, and C. H. Lai, "Mining the change of event trends for decision support in environmental scanning", *Expert Systems with Applications*, vol. 36, 2009, pp. 972–984.

[17] M. J. Shih, D. R. Liu, and M. L. Hsu, "Discovering competitive intelligence by mining changes in patent trends", *Expert Systems with Applications*, doi: 10.1016/j.eswa.2009.09.00.

[18] N. Zhou, H. L. Cheng, H. Q. Chen, and S. Xiao, "The framework of text-driven business intelligence", *International Conference of Wireless Communications, Networking and Mobile Computing*, New York, USA, 2007.

[19] T. Kakkonen, "Developing CoProE, a company, product and business event ontology, by reuse", submitted for review, 2010.