Knowledge Discovery from the structure of Persuasive Communication

Katsutoshi Yada, and Naohiro Matsumura, *Member, IEEE*

Abstract— It is difficult to carry out quantitative measurements of the persuasive power of business communications (i.e., persuasive skills) and such communications are likely to involve difficult to understand, unseen and unknown knowledge. However, using unstructured recorded communication data based on conversations with business customers, we have been developing explicit knowledge concerning skills necessary for effective communications in the form of an expression framework. The objective of this research is to generate a framework and a process for explicit management knowledge concerning understandable communication skills as opposed to the tacit, hard to understand the negotiating skills related to overdue payment collection personnel and to verify the actual usefulness of this knowledge using the accumulated data in a company. Using this process we have developed, it is possible to discover the special characteristics of the communication content of high success overdue payment collection personnel.

I. INTRODUCTION

IN recent years, in the case of Japan, many companies are facing problems related to the so-called "2007 Problem", a problem based on the massive numbers of workers of a specific generation that will come up for retirement in that year. This will cause a shortage of workers with high skill levels based on long experience and training supplied by these companies. This factor is related to the sustainable competitive advantage and the source of company superiority in the market and constitutes a valuable and irreplaceable asset (Barney, 2000). The skills of these employees have created great value for their companies and include skills that are difficult to explain in words and are a form of what is called "tacit knowledge (Nonaka and Takeuchi, 1995)". Many different research efforts have pointed out the importance of tacit knowledge (Yada, 2004) and this shortage of skilled workers will cause a crisis in the case of many companies. However, these tacit knowledge skills are difficult to pass on and in order to make them the common knowledge of all the members of a given organization, it has become important for companies to convert this tacit knowledge into explicit knowledge. In this paper, we examine one of these cases of such skills, persuasive communication skills for negotiating with customers to persuade them to take a specific course of action.

Based on the fact that a certain Japanese company related to the telecommunications industry has been experiencing stagnant performance in recent years, overdue payments owing to the company have been expanding and payments for products and services have been slowing down. Thus it has become necessary for special operators to negotiate over the telephone with these late-paying customers to collect the payments owed. These operators can be divided into talented operators who continuously have a high overdue payment collection rate and those operators that have a low collection rate. It is possible to surmise that the difference between the performance of the high success rate operators and the low success rate operators is related their levels of negotiating skills. In the next few years, large numbers of these highly skilled operators will retire and it is thought that this will have a very serious effect on the company. In order for the company to rapidly communicate the special skills of these operators to other employees of the company, the company has a strong need to convert these skills into understandable, explicit knowledge. However, although interview-based and questionnaire-based research with a number of these operators and their supervisors was carried out, it was not possible to discover a coherent set of overdue payment collection negotiating skills.

At this point, we decided to recommend that a large-scale operations database of communications data be compiled using the recorded conversations of the operators and the overdue payment customers. From the database containing all of the recorded conversations, we surmised that it might be possible to extract the pattern or rules of negotiation conversation that would have some type of meaningful influence on the rate of collections. However, this type of data based on conversation and communication cannot easily be converted into written text data and is entirely unstructured. The data was massive and, as far as we could ascertain, there existed no quantitative or general use methods that could be used to analyze such data.

The objective of this research is to develop a knowledge management framework and process that can be used for expressing as explicit knowledge the tacit knowledge in the overdue payment collection operator conversations and inherent in the related skills, named SMPC (Skill Mining from Persuasive Communication). The objectives also included verification of the usefulness of the results based on

K. Yada is with the Faculty of Commerce, Kansai University, Suita, 564-8680 Osaka, JAPAN (telephone: +81-6368-1121, e-mail: yada@ipcku.kansai-u.ac.jp).

N. Matsumura, is with the Graduate School of Economics, Osaka University, Toyonaka, 560-0043 Osaka, JAPAN (e-mail: matumura@econ.osaka-u. ac.jp).

using the company data. In this paper, by integrating the methods of text mining and statistical method, we proposed the use of quantitative and general-purpose methods to discover the tacit knowledge to useful knowledge and effective persuasion methods using a massive amount of communication data. With regard to the case study described in this paper, as a result of using the methods we suggested for handling the customer negotiation data of a subsidiary of a Japanese telecommunications company, we were successful in converting a part of the tacit skills of the operators to explicit knowledge.

The organization of this paper is as follows. The following section introduces the research that was carried out in the past for content analysis. Then, we explain the framework used for discovering useful knowledge from the negotiation communication data. Following this, we explain the case where we discovered new data by using the methods we recommended researching the data involving communications with customers of a Japanese telecommunications company. Lastly, we summarize the results, state our conclusions and add some comments on future research topics.

II. PERSUASIVE COMMUNICATION CONTENT ANALYSIS

The existing research in the persuasive communication area had been already carried out involving the analysis of communication and negotiations with customers similar to the project described in this paper (Bettinghaus, 1968). Persuasion can be defined as using communication to convince the other party to accept a solution to a problem and to change the attitude and behavior in the desired direction by exerting influence on the other party (Erwin, 2001; Perloff, 1993; Seiler and Beall, 1999). Concerning research on persuasion, such aspects as the factors that promote successful persuasion, the mechanism of persuasion and methods of predicting success have been studied (McGuire, 1985). However, in most of these studies, the methods used consisted of questionnaires and experiments in laboratory. Such methods cannot be used for handling massive amounts of actual business data such as recordings of conversations to obtain quantitative and general application results (Araki et al., 2005).

As a method of extracting characteristics from text data based on conversations, up to the present time, content analysis methods have been used. Content analysis (Krippendorff, 1980) is a scientific method that investigates language and other symbols using communication data to arrive at deductions concerning the data. Most research based on questionnaires and interviews uses various forms of coding for simplification. Questionnaires are structured by using prepared answers for selection by the respondent. Thus, in effect, the generation of the data must be controlled using a priori methods. This type of method calculates the frequency of use of specific words using a computer and large amounts of data can be processed (Pool, 1959; Stone et al., 1966). This type of method is frequently used for studies in the areas of

psychology, sociology and political science to study practical questions in these fields.

In the case of most of these kinds of studies, the context to be analyzed is established in advance and the analyst has preconceived hypotheses that are to be verified. In the actual process used, multiple vocabulary word groupings are used to generate word lists in advance. Frequency, related occurrences and distance of separation in the text can also be calculated for making interpretations of the data. Unfortunately, we were not able to generate any reliable hypotheses from the interview research carried out in advance among involving the operators and their supervisors. Therefore, it was necessary to develop methods to elicit hypotheses from the amassed conversation data.

III. SKILL MINING FROM PERSUASIVE COMMUNICATION

A. The Objectives of Persuasion communication data mining

The objectives of the study were focused on extracting the differences in the negotiation skills of the high late payment collection rate operators and the low collection rate operators from the persuasion communication data. Therefore, this required an analysis framework with integration between text mining and traditional statistical methods. This involved an effort to clarify the factors such the timing and words used related to high collection of late payments based on negotiations between the operators and the customers. The persuasion communication data late payments referred to in this paper consisted of normal conversation data that was massive in size and unstructured. In order to create operator classification models, the text mining and the statistical analysis were integrated and the time series data for the conversations was added by means of developing a process for structuring the total data. In this paper, the process and system that was developed is called "SMPC" (Skill Mining from Persuasive Communication Data).

In this paper, the term "late payments" refers to unpaid charges for the company's products and services owed by the customer, also referred to as the "late payment customer". The term "operator" refers to the late payment collector personnel using the telephone to collect the unpaid charges owed by the customer. The term "call" refers to a single telephone call made by an operator to a single late payment customer and these calls are made up of exchanges of messages. Here, the term "message" refers to the start of a portion of a conversation initiated by the operator (or customer) until this portion of the conversation ends and a reply begins. A given message consists of various "words", including nouns, verbs, adjectives, exclamations, etc.

B. The Persuasion Communication Skill Extraction Framework

The SMPC consisted of four phases: 1) Classification of the operators, 2) Message quantification, 3) Vocabulary grouping and 4) Operator characteristic extraction. (See Fig. 1). Below, the procedures used for these four phases are

1742

explained in detail.

1) Operator Classification: The late payment collection operations are managed using management information concerning the assignment of collection accounts to each operator and records of the attempts to collect the charges in arrears and levels of success and are archived. Along with this type of data that was extracted from the database, the unpaid charge collection ratio (r) related to the total amount of unpaid charges owed for the specific accounts given to a specific operator was defined as per the following formula:

$$r = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} D_i}$$

A set of total record of unpaid charges: $a = \{1, 2, ..., n\}$

Sum of total value of unpaid charges = D_a

Sum of total collections of unpaid charges = d_a.

For each operator, the ratio of collection of unpaid charges in arrears was calculated and based on k-means clustering; the operators were divided into the high collection ratio operators and low collection ratio operators.
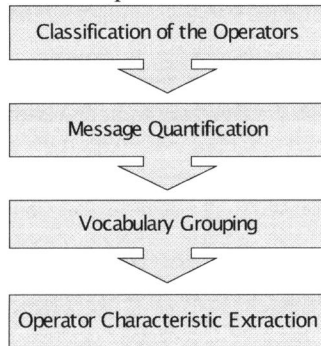

Classification of the Operators

Message Quantification

Vocabulary Grouping

Operator Characteristic Extraction

Fig. 1. The Framework of SMPC.

2) Message Quantification Based on Text Mining: The operator and customer conversation data concerning payment of unpaid charges is unstructured data that is massive in size and difficult to handle. Therefore, we focused on the vocabulary items, "words," used during the persuasion communication used by the operators and the timing involved and analyzed the data using the following procedures:

--The time series data concerning the messages in the call was generated.

--By means of morphological analysis, the various vocabulary items in the target messages in the calls were divided into each word.

--All items other than morphemes of more than two Japanese characters, verbs, exclamations, and katakana syllabary characters, personal names, location names and nouns were eliminated from the data.

--The frequencies of use of the items that remained were calculated and a keyword list of N word items in order of frequency of use was generated.

--As the vertical axis, messages, and as the horizontal axis, the key words from the key word lists were used to divide the messages into those containing one or more key words (1) and those containing no key words (0) and the results were used to make tables.

--Lastly, this table data was integrated with the timing information of a message in call.

Based on this process, it was possible to quantify the message key words and the related call message timing data.

3) Key Word tables: The vocabulary items with very low frequency of use tended to include special technical vocabulary words related to overdue charges and these words were eliminated as a target for analysis. Nevertheless, since the number of different words included in the total calls was very large, if all these vocabulary items were used it would have been impossible to carry out a useful and focused analysis. To handle this situation, a factor analysis was used to extract potential factors in word groupings and carry out an analysis of their meanings as a practical solution to this problem. Extracted factor of key word grouping (key word lists) were created and the words the messages contained and the related timing factors were used to make tables.

4) Extraction of the Special Characteristics of the Operators: As the last phase, the data produced by the preceding phases indicated above was used to extract the special characteristics of the operators. The definitions of the operators obtained in the 1st phase were used as the objective variables and classification models were generated using the word grouping words contained in the messages and the timing data as the explanation variables.

IV. EXPERIMENTS

The target data of this case was data from the conversations between the operators of a Japanese company and customers of the company concerning payments for charges in arrears aimed at collecting these charges from the related customers. The operators that carried out these negotiations with the customers to obtain payment of overdue charges for company products and services held conversations with the customers, using computers indicating the payment history and the conditions of the contracts, amount of unpaid charges, etc. and headphones. We established a group of 16 operators that were equipped with PCs that were set up to record all searches for information in the system and with recording using devices for recording the conversations. Data was collected in March of 2005 over a period of two weeks and an analysis was carried out on the data related to 108 cases of overdue charge payment negotiation. The messages in the related calls included about 15,000 messages that were generated by the operators. The average length of a call was about 6 minutes and average number of messages per call was 72 messages.

A. Operator Classification and Key Word List Generation

Based on the unpaid charge collection ratio using k-means clustering, the operators were divided into two groups (High collection rate and low collection rate). The average overdue payment collection rate of the high collection rate operators was 16.56 percent and the average rate for the low collection rate operators was 11.91 percent. For such factors as average value of charges in arrears, length of time in arrears and length of calls, there were no statistically significant differences between the two groups.

Next, the numbers of words that frequently appeared in the calls were extracted from the calls and a key word list was

generated. In this particular case, the 200 most frequently used key words were used. Then, the data concerning the timing of the messages in the call were integrated with this list data to generate a data set.

B. Grouping of Key Words Using Factor Analysis

Based on the factor analysis, from the data set key word groupings used by the operators were made and four potential factors were discovered (See Table 1). The Kaiser-Myer-Olkin (KMO) measure of sampling adequacy was 0.587, which falls into the range of being mediocre (Hutcheson and Sofroniou, 1999). The Bartlett's test of sphericity was highly significant (p-value < 0.001), indicating that there was a significant degree of covariance between the items.

Table 1. Factor analysis of key words contained in the messages.

words	Factors			
	1	2	3	4
confirmation	0.74		-0.241	
to request	0.662	0.387		
to send	0.637			
installments	0.607	0.271		0.305
remittance in	0.574			
single payment	0.495			
letter	0.486		0.34	0.285
amount	0.415		0.216	
total amount	0.226			
arrive	0.213			
make an effort		0.689		
court case		0.653		
able to pay		0.589		
will do one's best		0.515		
will be able to receive	0.316	0.496		
late			0.777	
interest			0.675	
shop			0.436	
argue in court			0.366	
can use			0.239	
promise			0.224	
no, no				0.617
documentation				0.539
understand				0.529
cancel contract				0.413
no				0.412
yes			0.228	0.267
management				0.218

Rotation method: Kaiser Method (Vari-max method for normalization, 6 rotations)

As the first factor, when concrete payment procedures were being put together with the customer in arrears, such key words as "payment by installments", "payment" and "one payment for the full amount", etc. were indicated to have high weightings. Therefore, we named this group of related key words the "Joint creation of a payment plan" factor.

As the second factor, the words that were often used at the end of the conversation that were related to confirming payment such as "able to pay", "will do one's best" and "will be able to receive" had high weightings. These words that the operators used after all varieties of negotiations at the end to confirm payment we named the "Payment confirmation" factor.

As the third factor, the words that the operators used to explain to the customer in arrears the precarious situation he/she was in included such words as "late/tardy payment", "interest charges" and "pleading" received heavy weightings. The operators used such words to remind the customers in arrears that they were late in paying or that they would be receiving a letter from the company lawyers and we named this factor the "Explanation to late payer" factor.

As the last factor, the operators often used such words as "I understand", "no..."and "yes..." in handling the late payer's excuses in a passive mode. We named this factor the "Excuse passivity" factor. The types of words that the operators are using can be broken down into four key word groupings. To use the frequency of use of the words in these four groupings to extract the special characteristics of the operators, the related messages were processed and after the necessary tabulations, were attached to the data set that was generated in the previous phase.

C. Extracting the Special Characteristics of the Operators

Here, the data for the four extracted vocabulary groups and the data concerning position in the calls were used to clarify the persuasion communication-related differences between the high collection rate and the low collection rate operators. Even though the call time data was not used at first, the data for first factor and second factor and the fourth factor frequency of existence (use) frequency within the calls indicated statistically significant differences.

Fig. 2 shows the ratio of High/Low Group for each words group that is defined as an average of word's frequency in calls of the higher collection ratio operators divided by one of the low collection ratio operators. For example, for the high collection rate operators, the average number of "Joint payment plan" factor word group words in a call was 10.28 and was higher than the low collection rate operators group by 7.18. Similarly, in the case of the "Payment confirmation" factor vocabulary word group as well, the level (of use) for the high collection ratio operators was greater. In addition, the average level of use of "Excuse passivity" factor word group words by the high collection ratio was 7.85 times, 3.11 uses greater than the level of use by the low collection ratio operators. In other words, the higher collection ratio operators spent much more time with the customers working out a payment plan and carefully reconfirmed the jointly created repayment plan and at the end of the negotiation remembered to carry out a final reconfirmation. In comparison, it appears that the low collection ratio operators, rather than having that type of conversations, tended to over-use the "Excuse passivity" mode and the negotiations tended to proceed at the customer's pace.
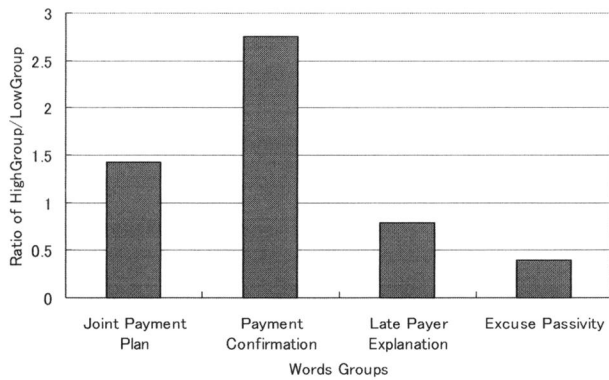
1744

Fig. 2. The ratio of an average of the words frequency in calls of High collection ratio operators group compared to that of Low group.

In addition, when the call time-related data and the four factor vocabulary group word use frequency relationship were analyzed, there was a statistically significant difference in the timing of the use of words related to the third factor, the "Late payer explanation". In the case of the low collection ratio operators, they tended to repeat the use of late payer situation explanation vocabulary into the latter half of the call. Fig 3 indicates the timing proportion of the use of these "Late payer explanation" words by the low and high collection ratio operators. As can be seen here, in over 60% of the calls they carried out, they used these "late payer explanation" factor vocabulary words during the last half of the call. In other words, it can be surmised from the data that the high collection ratio operators carefully carried out the necessary explanation during the first half of the call whereas, in comparison, the low collection ratio operators waited until the last half of the call and then tended to repeat the explanation to the customer in arrears.
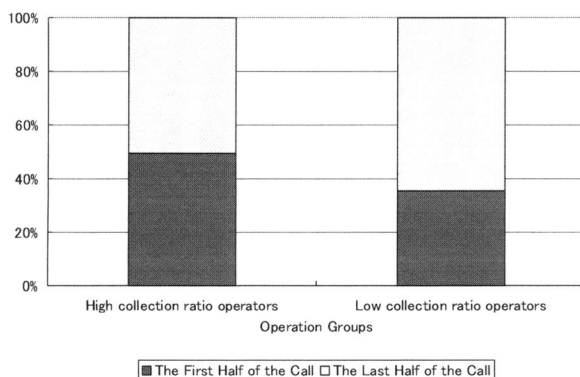

Fig. 3. The timing proportion for explaining the late payer situation in the high and low collection ratio operators.

D. Implication for the Business

These results were fed back to the operators and when interviews were carried with them, it was possible to discover the following implications for the business.

1) In order to examine the actual content of various methods of payment, it is necessary to spend sufficient time explaining these matters to the slow paying customers: One operator said that getting the customer in arrears to repeat actual amount to be paid and date of payment and method appeared to her to have a great effect on the actual payment. For this reason, she said that she "listened carefully to the customer's comments". It can be surmised that to involve the customer in arrears in establishing the payment method and getting the customer to reach a sufficient level of understanding and commitment will raise the collection ratio. It is necessary, when listening to the customer's conversation to pay sufficient attention to the content of his/her statements and to lead the customer toward statements concerning more concrete payment methods.

2) In order to get the customer to listen and understand the operator's statements, it is necessary during the negotiations not to get caught up in the pace of the other party. One must lead the communication process: Ordinarily, this type of customer is unlikely to admit his own fault and, if possible, will try to get the operator to understand the personal situations that these customers are in and to sympathize and then attempt to get out of paying. It cannot be said that the negotiation is succeeding if the operator is listing on and on to the customer's excuses. One high collection rate operator said, "In actual fact, I spend almost no time listening to the excuses of the customer. I always try to change the subject." As can be seen here, she does not let the negotiation proceed in the way the customer wants it to. She felt it was very important that she lead the communication. To achieve this, if the conversation began with excuses about late payments, she felt it was important bring the conversation back to the subject of methods of paying amount outstanding.

These implications could not be discovered using the methods that have been used in the past and a part of the high collection ratio operator tacit knowledge was converted to explicit knowledge. Using this knowledge as a basis and by reviewing and improving the training program, we can anticipate that improvements can be made in the collection rate for accounts in arrears.

V. CONCLUSION

In this paper, we have proposed a knowledge management framework and process for clarifying the tacit skills related to persuasion communication that is possessed by operators engaged in getting payment for over-due accounts. When we used the method we developed for analyzing the data of a Japanese telecommunications company, we discovered differences in the communication of the high collection ratio operators compared to low collection ratio operators that had not been discovered using the previous methods.

There are several problems still remaining that are implied by this paper. The number of operators used for respondents for this paper was relatively small. There is a need to verify the hypotheses that have been generated by using a larger number of respondent operators for analysis using our methods. In addition, in the case of the research described in

this paper, it was not possible to use the actual unpaid amounts or the respondent's personal attributes that are quantitative data that help to make it possible to clarify the relationships between the hypotheses. Lastly, we used only a small part of the time-series data in the conversations. In order to analyze the dynamic process that is inherent in the persuasion communication, it will be necessary to develop a proper framework in the future.

ACKNOWLEDGMENT

REFERENCES

[1] T. Araki, S. Hamada, N. Matsumura, S. Niwase, Y. Ohsawa and K. Yada, "Chance Discovery from Consumer Research Using KeyGraph, Readings in Chance Discovery, Advanced Knowledge International, 2005, pp. 373-384.

[2] J. B. Barny, Gaining and Sustainable Competitive Advantage, Second Edition, Prentice-Hall, 1997.

[3] E. P. Bettinghaus, Persuasive Communication, Holt Rinehart & Winston, 1968.

[4] P. Erwin, Attitudes and Persuasion, Psychology Press, 2001.

[5] G. D. Hutcheson and N. Sofroniou, The Multivariate Social Scientist, Sage Publication, 1999.

[6] K. Krippendorff, Content Analysis: An Introduction to Its Methodology, Sage Publication, 1980.

[7] W. J. McGuire, "Attitudes and Attitude Change," G. Lindzey & E. Aronson (Eds.), The Handbook of Social Pychology, Third Eddition, Random House, 1985, pp. 233-346.

[8] I. Nonaka and H. Takeuchi, The knowledge-Creating Company, Oxford University Press, 1995.

[9] R. M. Perloff, The Dynamics of Persuasion, Lawrence Erlbaum Associates, 1993.

[10] I. D. S. Pool, Trends in Content Analysis, University of Illinois Press, 1959.

[11] W. J. Seiler and M. L. Beall, Communication: Making Connections, 4th Edition, Allyn and Bacon, 1999.

[12] P. J. Stone, D. C. Dunphy, M. S. Smith and D. M. Ogilvie, The General Inquirer System: A Computer Approach to Content Analysis, MIT Press, 1966.

[13] K. Yada, "Knowledge Discovery Process and Introduction of Domain Knowledge," B. Montano (Ed.), Innovations of Knowledge Management, IRM Press, 2004, pp. 86-98.