Multidimensional Sentiment Cube Mining for Process Monitoring

**Nanthawadee Sucharittham[1,2,*], Choochart Haruechaiyasak[3],
Hieu Chi Dam[2] and Thanaruk Theeramunkong[1]**

[1]*School of Information, Computer, and Communication Technology,
Sirindhorn International Institute of Technology, Thammasat University, Bangkok 10200, Thailand*
[2]*School of Knowledge Science, Japan Advanced Institute of Science and Technology,
Ishikawa 923-1211, Japan*
[3]*National Electronics and Computer Technology Center (NECTEC), Pathumthani 12120, Thailand*

(*Corresponding author's e-mail: d5422300383@g.siit.tu.ac.th)

Abstract

Process monitoring is essential for quality improvement because it is necessary to find the answers to which business issues need to be understood. In the era of social media, many critiques concern the business domain, including life insurance, which is one of the significant business sectors in Thailand. To utilize this useful cloud corpus for the business improvement process, we propose a novel methodology for process monitoring using the concept of multidimensional sentiment cube (MDSC) mining to raise usefulness with the business process model notation (BPMN). As the ability of MDC raise unlimited analysis perspectives merge with sentiment analysis (MDSC), this method can provide more sets of data for association rules mining and meet the needs to be analyzed. The cube analysis scenario, which uses association rules mining results, can reveal a significant hidden issue among aspects and sub-aspects associated under our design with their measurements. The results can be used for monitoring, which presents the customer's sentiment from social media in the real business case and identifying in the real process model.

Keywords: Multidimensional sentiment cube (MDSC), Aspect-based sentiment analysis, Life insurance, Business process management notation (BPMN), Association rules mining, Customer relationship management (CRM)

Introduction

Process monitoring for increase business effectiveness, which is related to customer satisfaction improvement, requires an understanding of customers' opinions. A traditional way to monitor process efficiency is the questionnaire questions designed by business domain experts in that area. They have profoundness in problems under that domain. Even if, the expertise process from their experience, however, it is quite limit viewpoint of analysis related to customer sentiment. Furthermore, it is not easily to link with the current business flow to identify the specific processes which need to monitor. Now we are in the social media era, problems have become more extensive than the domain experts' knowledge. To support this problem, the problems are easy to realize from critiques blogs that express negative sentiment on there. In recent years, sentiment analysis influences industrial analysis activities, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. It represents a massive problem space [1,2]. To utilize the concept of sentiment analysis for this task needed the method in part of aspect and sense identification. This paper chooses Feature-based or Aspect-based sentiment analysis for identifying both aspects/sub-aspects and customer sentiment pertain to a business process. The linkage of aspects/sub-aspects and customer sentiment from critique blogs into the business process to set up the correct modules' improvement process requires some methods to define features or factors that match the business process. Therefore, identifying a clear predefined operating framework will support the linking business process more accurately [3,4].

This paper introduces a novel methodology, MDSC mining, on BPMN to monitoring processes that get customer dissatisfaction. Our method employs multidimensional process mining (MPM), which has a basic idea to interpret the attributes from the event log as dimensions that form the edges of a multidimensional data cube [5]. Besides, we merge sentiment analysis into MPM to scope business process, which needs to solve problems due to customer evaluation's high negative sense. The filter log task is to derive the desired view on the data is a laborious task. It is time-consuming if done manually, especially if multiple process models should be compared [6]. Data mining rules such as association rules from MDSC provide concrete rules to support the log filtering. It is rare research to build Multidimensional cube with sentiment analysis to be MDSC as our task, especially to support process mining based on association rules mining. A MDSC mining for process monitoring uses the concept of process mining techniques on the cube, which is a process analysis to discover process models from event data records. MDSC raise unlimited analysis perspectives merge with sentiment analysis, this method can provide more sets of data for association rules mining and meet the needs to be analyzed from concrete rules for consider with BPMN.

Many previous works used this concept to finding and improving on business process. In the 1st generation, the concept made use of structural data design in the database. An idea starts from a single viewpoint of analysis on the process, and the next generation of an idea is a multi-viewpoint. Many researchers adopt a useful cube concept such as Event Cube, which introduces a multidimensional data structure to hold business dimensions. Ribeiro and Weijters used Event Cube to improve business analysis quality [10]. Alfredo and Aalst formalizes the notion of process cubes where the event data is presented and organized using different dimensions and gives a notion connected to OLAP (Online Analytical Processing) data cube and adapting the OLAP paradigm to event data through MPM [11].

To demonstrate in a specific domain, Vogelgesang and Appelrath made a prototype in the healthcare domain to distinguish between different groups of patients, defined by the patients 'properties such as age and gender, to get more precise insights into the treatment process. PMCube Explorer tool for MPM to support viewers from various views used data warehouse [12]. Vogelgesang and Appelrath derived several general research questions addressed for MPM by a literature review. They make comparable on many steps to point out its limitations, requirements and limitations [13].

In the next era of process mining cube, many researchers started an experiment on unstructured data such as text document, chat blog on Twitter, etc. For example, Ravat *et al*. introduced an OLAP multidimensional conceptual model without facts. This model is based on the unique concept of dimensions, and they are utilized for multidimensional document analysis. They also provided the operation set of cube explanations [14]. Liu *et al*. [15] discussed a text cube approach to studying different kinds of human, social and cultural behavior (HSCB) embedded in the Twitter stream. They showed how to organize data from text to multiple dimensions and hierarchies and make visualization with statistical reports and perform OLAP. Zhang *et al*. proposed the topic cube to combine OLAP with probabilistic topic modeling and enable OLAP on the dimension of text data in the multidimensional text database. They also proposed 2 heuristic aggregations to speed up the iterative EM algorithm for estimating topic models by leveraging the models learned on component data cells to choose a good starting point for iteration [16].

Our tasks implemented a MDSC using critique blogging from websites. It is 1 kind of unstructured data of cube generation. Besides, our multidimensional cube is a sentiment cube that also used a hierarchical concept. In the other related papers, some researchers studied hierarchical concept mergers to sentiment analysis, such as Kim *et al*., which presented a hierarchical aspect sentiment model (HASM) to discover a hierarchical structure of aspect-based sentiments from unlabeled online reviews. They used the Bayesian nonparametric model recursive Chinese Restaurant Process (rCRP) to extract the tree's structure and parameters [17]. Dave *et al*. reported the development and initial results of a multidimensional analysis agent for an online learning environment to show students feedback on different levels, including identifying potential problems during the course delivery [18]. To combine multidimensional cube with sentiment analysis theory, Dayal *et al*. and Umeshwar defined patent of MDSC that a sentiment cube system is a disclosed system which stores sentiment elements inside. Its data structure has a set of cells arranged by a group of dimensions [19,20]. We utilize the sentiment analysis concept to merge into a multidimensional cube, especially in aspect, sub-aspects identification for cube dimensions. In previous work of sense categorization in opinion sentences, many techniques are used knowledge extraction from that context of reviews such as Feature-based (Aspect-based) summarization, which Bing Liu implemented to extract sense on products review [1,2,18].

This research studies the case of the Thai life insurance industry on the service process. It is one of the critical financial industries in Thailand; however, the volume of policies still has a small scale.

Moreover, this business gets a lot of complaints in many business processes that appear on critique websites in Thailand [7-9]. Therefore, our implementation concern analysis way of monitoring the business model's inefficient process from customer dissatisfaction in-service process [7]. It was studied to identify the process that needs to improve using the negative sentiment of users on current business processes. Therefore, the analysis in the concept of CRM is required to use in aspect-based sentiment analysis tasks [7-9,21-25]. Lexicon, such words or phrases are used for opinion extract from the corpus is one of the main elements to constructing aspect-based sentiment analysis. A well-organized lexicon gets accuracy on sentiment extraction. Many researchers are interested in automatically making lexicon construction, such as using a machine learning technique or data mining method. Previous works of Wiebe and Riloff have proposed different approaches for automatically constructing the lexicons for feature-based (aspect-based) opinion mining [26]. Next, mining data extract from a different layer of the cube makes use of association rules mining in negative sentiment pattern on customer service. Then utilize these rules to monitor the significant processes refer to BPMN. Previous task about data mining to process mining, many tasks also use association rules mining such as Sarno *et al.* proposed anomaly detection approach based on association rule mining in hybrid association rule learning for fraud detection [27]. Next task of Sarno *et al.* also proposed fuzzy association rule learning develops from association rules for anomaly detection in business processes [28]. In addition, Suzhen made use of Apriori algorithm to detect business process management [29]; Bohmer and Rinderle-Ma also use association rules for detect anomaly in dynamic process runtime behavior [30].

In summary, this task presents a methodology to monitor processes from a MDSC. The process of aspects/sub-aspects identification deploys the benefits from the CRM concept in-service process in each customer life cycle to generate a MDSC. Then finding the concrete rules of the concerned process from event cube logs from association rules mining and identify them on BPMN diagram. The next is the section of the background on life insurance in-service process related to CRM concept. Then follow with the section to present the methodology. The results are explained in next section. Finally, the conclusion is made in the last section.

Background

Life insurance in the service process is closely the relationship between service providing from company and customer who is possible to leave comments on social web-blogging. To identify the life insurance service process, we have to review the service process on this business in the general practice [19-23]. We explained the customer life cycle's concerns on the management information, consisting of 3 stages: Customer acquisition, customer intention and customer engagement.

Stage 1: Customer acquisition (pre-purchase → purchase process)

Service in the customer life cycle starts from the company finding a list of prospects (expected customer) in order to offer a new plan (product) or contract. Service touchpoint proposes the proper plan to the customer to meet customers' requirements. After the prospect decided to apply for a new life insurance contract, a new application will be submitted to the underwriting department to verify customer qualification and then issue a new policy contract and deliver it to the customer by service touchpoint.

Stage 2: Customer intention (purchase → service support process)

After the new policy contract was approved, the customer will be obtained a service under contract conditions such as service on policy operation process and claim settlement. An operation service has a sub-process of policy status changing (make extended term insurance (ETI), reduced paid-up insurance (RPU) or cancel policy), make loan from policy when cash value of policy occurred. A claim settlement has a claim type that includes a major claim (death claim) or a minor claim (illness claim). Also, there is a problem with a claim such as long service, conceal health record, payment reject, etc. [22].

Stage 3: Customer engagement (re-purchase process)

When policy status reaches to termination stage, we will also process customer engagement by trying to keep the customer with active status or recall ex-customer by resale with a new product or retain customer before leaving the company. We classify the dissatisfaction type of customer into 3 groups of impact to life insurance situation. There are high impact (make sue on the court, make accuse to agent, make a complaint to OIC), medium impact (make policy cancellation, policy reject, etc.), or low impact (express negative feeling in social media) as **Figure 1**.
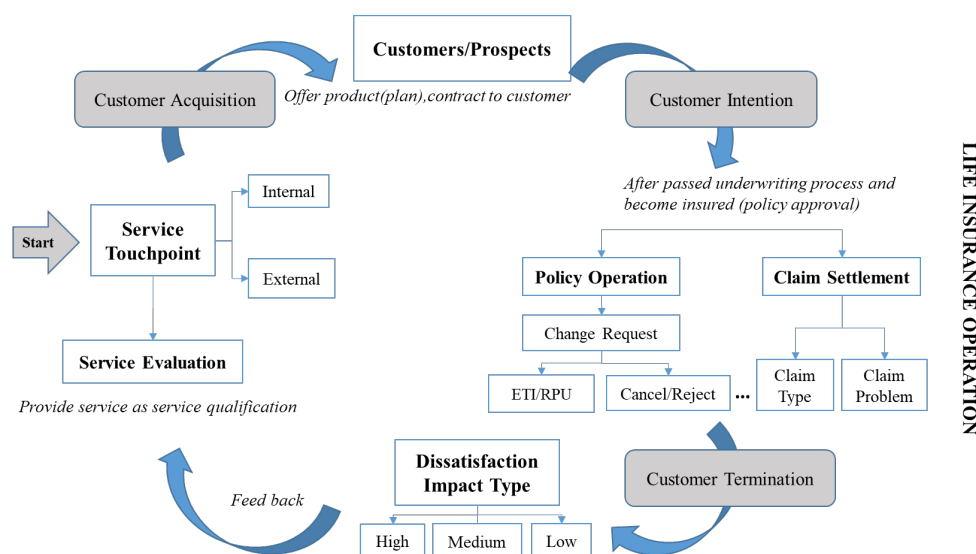
Figure 1 Customer life cycle in life insurance with CRM stages.

As **Figure 1**, we set up some monitoring points to be examples of our demonstration.

Question 1: Does it have any dissatisfaction issues from customers that occurred in the customer acquisition stage? Moreover, what are the kind of problems and related issues?

Question 2: Does it have any dissatisfaction issues from customers that occurred in the customer retention stage?

Question 3: Does it have any dissatisfaction issues from customers that occurred in the customer engagement stage?

Business Process Model Notation (BPMN)

BPMN is a technique to exemplify business process modeling in the pattern of graphical notation diagrams based on the traditional flowchart concept [3,4]. The usefulness of BPMN is to provide a notation to business users intuitively with technical users. BPMN is a way to define, identify and analyze the business process, and BPMN provides a standard notation that is promptly understandable and helps bridge the gap of communication. One of the good points of BPMN characteristics easily understands the association of process flows and structures of various features for both technical and non-technical observers. We draw 2 BPMN flows with 6 swimlanes, including Social media, Customer, Service Touchpoint, Underwriting or policy operation service and Agent service. Claim department & Hospital and Policy operation service (POS). We draw both flows were drawn having a purpose by providing linkage procedures on the relevant points in which the customer expressed sentiment on social media. Because we will use these flows to identify the related process of each aspect which customers expressed a high volume of negative sense in social media. These flows show connecting points, which use to identify aspects and sub-aspects and then use them to design dimensions of the MDSC afterward. The monitoring points from the knowledge base which is presented by BPMN as below:

First, the BPMN flow shows a monitoring point in the customer acquisition stage (**Figure 1**).
A-line area shows the acquisition stage which is process of approaching new customer. The concerning points that cause sentiment impacts on customer opinion are: 1) Service touchpoints provide whether good enough information as knowledge standard. 2) Service touchpoint provides whether professional services. 3) Service touchpoints operate whether the standard of the underwriting process as insurance rules and regulations.
B-line area shows the process of policy approval. The concerning points which cause sentiment impacts on customer opinion are: 1) Service touchpoint provides whether ethical enough policy conditions. 2) Service touchpoint whether good enough service such as fast and polite characteristics.

The 2nd BPMN flow shows the customer engagement and retention stage (**Figure 2**).

C-line area shows the process to request the claim settlement process. Customers require after-sales service. The concerning points which cause sentiment impacts on customer opinion are: 1) Service touchpoint explain whether enough knowledge to make customer understanding in rules and exception of the claim process. 2) Claim department collaborate with the hospital whether fast enough as customer expectation. The customer can post feelings and opinions in both C.1 and C.2 related to the feeling of this service (**Figure 3**).

D-line area shows the claim verification process. In general, a customer expects to get compensations to cover all of their payment. The concerning points which cause sentiment impacts on customer opinion are: 1) In case of reject payment, customers will criticize when compensation is not received or not received as expected. 2) In case of concealing, such as hidden of serious sick before making a policy contract, the insurer (company) entitles by law to or makes policy cancellation (Life). 3) In the case of the claim, results did not meet customers' expectations, and sometimes customers decide to change policy status such as ETI, RPU, cancel or make a loan, etc. The customer can post feelings and opinions in both D.1 and D.2 related to this service's feeling (**Figure 3**).
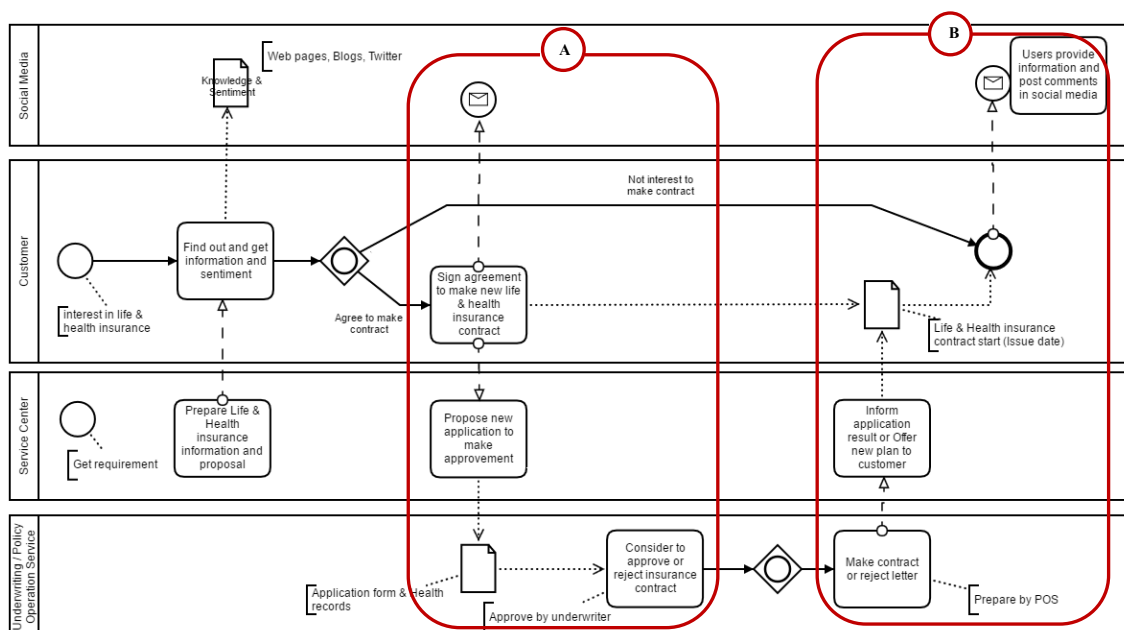

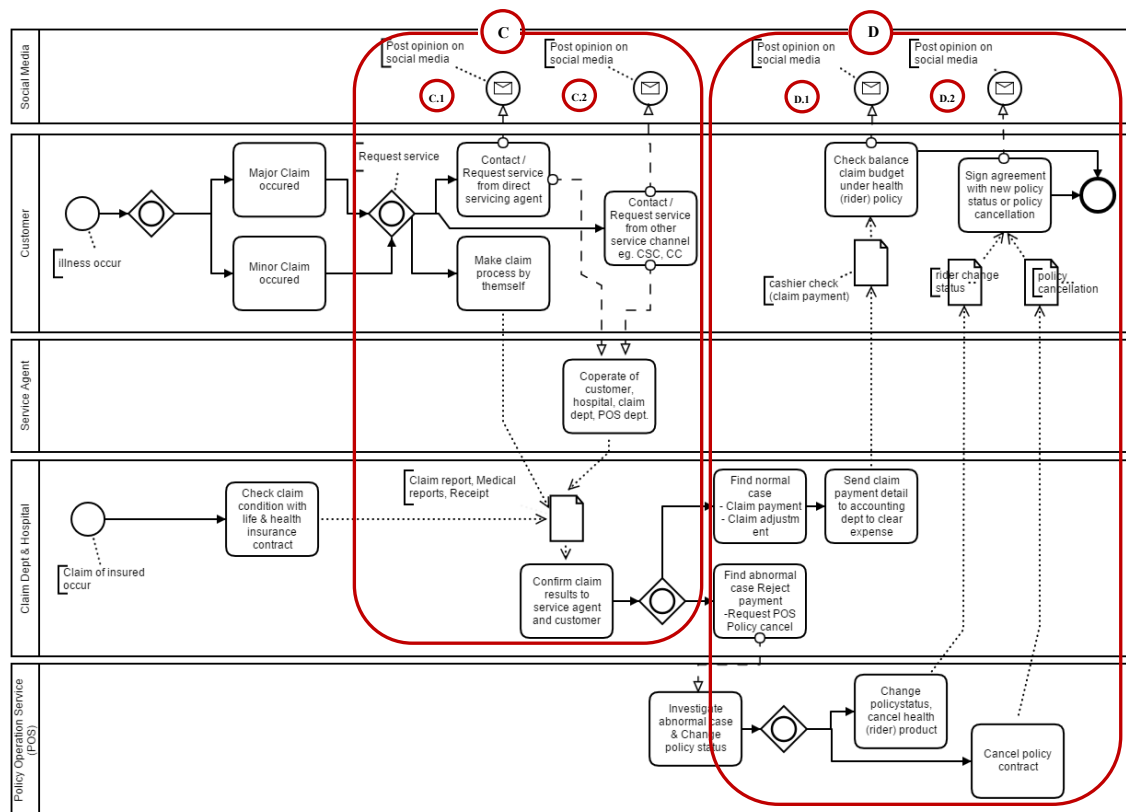**Figure 2** Customer acquisition stage in life insurance.

Figure 3 Customer engagement and retention stage in life insurance.

Materials and methods

In this research, we deploy the useful of technology to help identify the concerned processes which should be improve in the real work due to got customer dissatisfaction which provide on social media. A methodology has been proposed to broaden the perspective of analysis to identification the concerned processes. By taking advantage of technology named "sentiment analysis" is now popular with Aspect-based sentiment analysis, which is a subject-based analysis combined with OLAP technology, makes it convenient and adds unlimited new perspectives to on-demand analysis from multiple functions such as roll-up/down, crosstab, drill-up/down, etc. Previous works, many researchers have attention to focus on topic level, paper level or basic hierarchical level. It is rare research to build Multidimensional cube with sentiment analysis as our task, especially to support process mining. We called Multidimensional sentiment cube (MDSC) which provide unlimited point of views between multi-aspects and customer sentiment. Furthermore, sentiment analysis has been incorporated into the design of the business process, which is based on the BPMN diagram. The event logs of cube are also used in the process MPM in conjunction with Association rules mining for finding concrete rules of concerned processes which needs to monitor.

Our task presents a process monitoring using multidimensional sentiment mining on Business Process Management Notation (BPMN) by taking the benefit of MDSC and data mining methods. The methodology starts from Knowledge Base preparation and being used in the structure of a BPMN which we deploy to get a clearer picture of the aspect/sub-aspects related to business processes. Then the process performed by starting with the corpus preparation based on the text preprocessing process on retrieved data from websites. The tagged corpus was selected by remove unrelated theme and chooses only relevant text by selecting the sentence level in order to avoid too much junk information. Next process is to identify dimensions and sub-dimensions of MDSC and taking sentiment scoring on keywords in sentence level. Before the MDSC generation, the data preparation such as cleaning and populated tables is required. After that, create sub-MDSC as the analyst viewpoints for generating the relevant event logs and providing a set of data for association rules mining. The results of this research has purpose to figure out which processes should be monitored and should be improved by focus on negative feeling of customer.

Finally, we can identify the concerned processes in BPMN diagram which should be improved to get better customer satisfaction. This method is the following steps (**Figure 4**):

Knowledge base preparation: Study business knowledge and preparation of the relationship of features in our case study.
This process is possible to make participate with a domain expert.

Business Process Management Notation (BPMN)
Present an association of multi-features of the case study; we draw a notation of business process from an intuitive knowledge base by display linkage procedures between relevant points which customers expressed sentiment on social media. The importance of this process is not only for support aspect/sub-aspect identification, which transform from features of knowledge base but also manifest clearly on complicated business processes that are associated directly to customer's sentiment.

We draw BPMN 2 levels, the top (parent) and details (child) level of the process. We assume (based on understanding) that to find out some service processes which get negative sentiment should be concerned to improve. The design of BPMN in the different levels, such as the sub-feature level, is for expanding functions of main features and supporting different process layers. Therefore, we draw BPMN in the sub-feature level to explain more details of that business process and present it in the hierarchy. The design of BPMN in the different levels, such as the sub-feature level, is for expanding functions of main features and supporting different process layers (**Figure 5**). Therefore, we draw BPMN in the sub-feature level to explain more details of that business process and present it in the hierarchy.
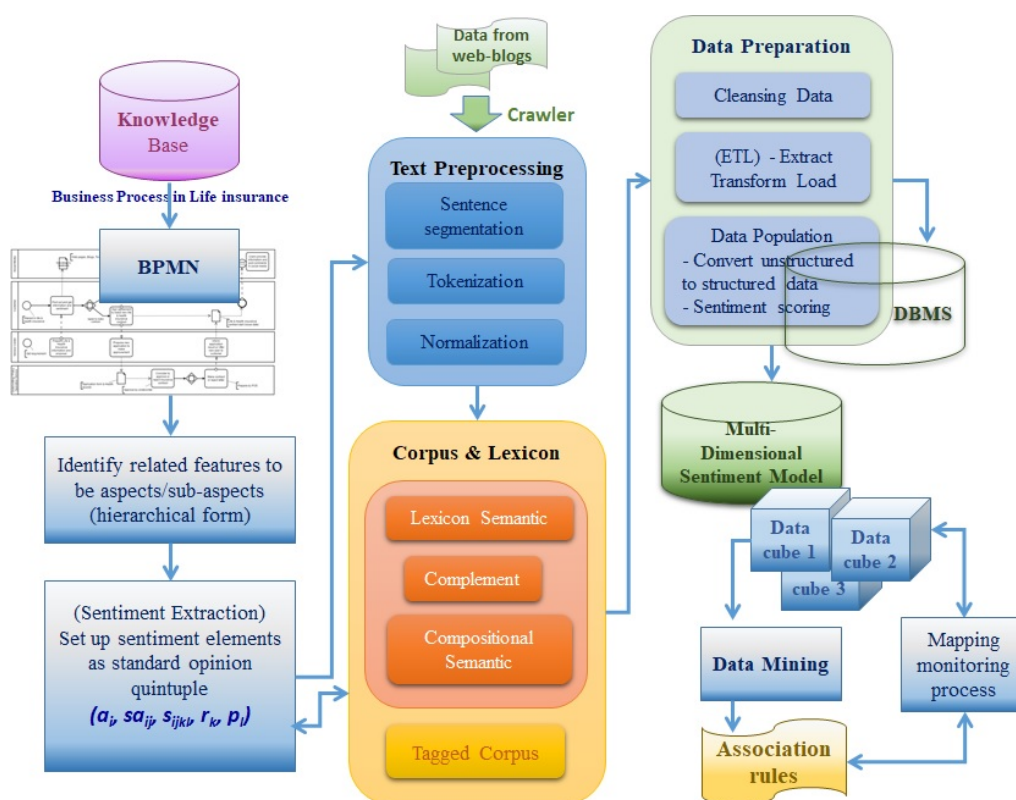

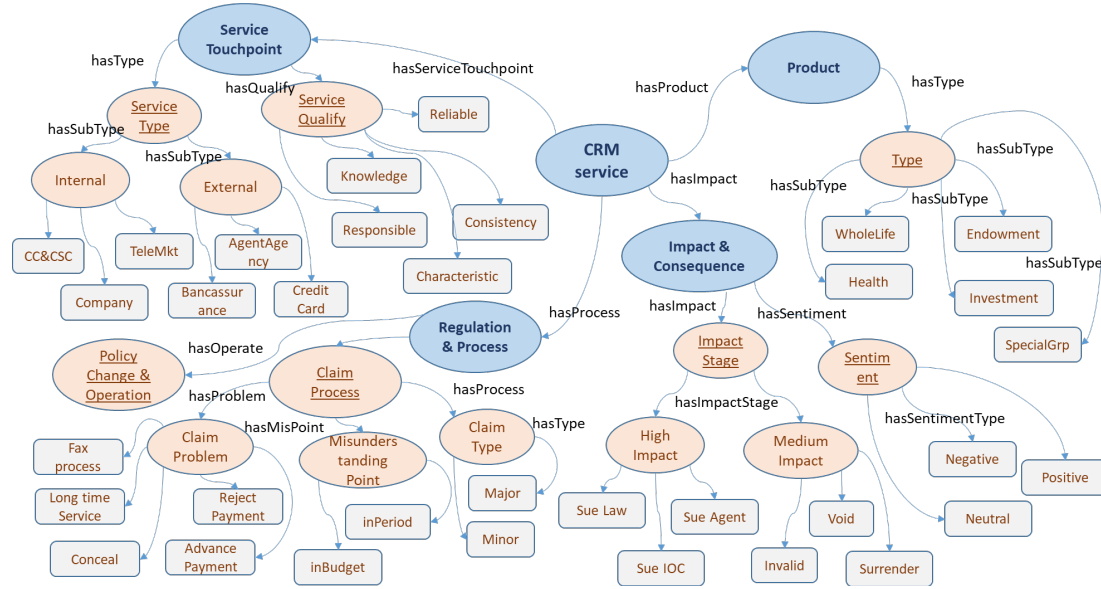**Figure 4** MDSC for process monitoring methodology.

Figure 5 Aspect and sub-aspects relationship hierarchy.

Aspect-based sentiment analysis

We deploy the concept of sentiment analysis for extracting sentiment and aspects elements from unstructured data by using standard opinion are a quintuple [1,2,17];

$$(a_i, sa_{ij}, s_{ijkl}, r_k, p_l)$$

Where a_i is an aspect name, sa_{ij} is a sub-aspect of a_i, s_{ijkl} is the sentiment on sub-aspect sa_{ij} of aspect a_i, r_k is the opinion reviewer, and p_l is the period when the opinion is expressed by r_k. The sentiment s_{ijkl} is *positive*, *negative*, or *neutral*. This research requires total sentiment of customer satisfaction, so

$$\text{Total score of customers' sentiment} = \sum_{n=1}^{m} \left(S_{in,jn,kn,ln} \right)$$

Whereas $\boldsymbol{S_{i_n j_n, k_n, l_n} \in (ai_n, sai_n j_n, si_n j_n, l_n, rk_n)}$

Where a_i is an aspect name, sa_{ij} is a sub-aspect of a_i, s_{ijkl} is the sentiment on sub-aspect sa_{ij} of aspect a_i, r_k is the opinion reviewer.

The sentiment s_{ijkl} is positive, negative and n is number of critiqued word chunk.

    When $a_i = \{prd, sp, is\}$ is a set of aspects (parent dimension of a cube)

    $r_k = \{st, cm\}$ is a set of opinion reviewer (parent dimension of a cube)

    $sa_{in} \in (sa11_{ij}, sal2_{ij})$

    $sal1_{ij} = \{lf_{prd}, rd_{prd}, am_{st}, sp_{st}, po_{sp}, stt_{st}, sv_{st}, po_{sp}, ca_{sp}, hi_{is}, mi_{is}, si_{is}\}$ is a set of sub-aspects (child dimension of a cube level 1)

    $sa_{l2ij} = \{ct_{ca}, cp_{ca}\}$ is a set of sub-aspects (child dimension of a cube level 2)

    $s_{ijkl} = \{negative, positive\}$

Aspect in parent level:

    st = service touchpoint, cm = company, prd = product (plan),

    sp = service process, is = impact & sentiment

Sub-aspect in child level 1:

    lf_{prd} = life (Basic) of prd, rd_{prd} = rider of prd, am_{prd} = affinity marketing of prd,

    sp_{prd} = special product of prd, stt_{st} = service type of st, sv_{st} = service evaluation of st,

    po_{sp} = service operation of sp, ca_{sp} = claim assessment of sp, hi_{is} = high impact type of is,

    mi_{is} = medium impact type of is, si_{is} = low impact type of is

Sub-aspect in child level 2:

    ct_{ca} = claim type of sp, cp_{ca} = claim problem of sp

Information extraction with text preprocessing and Natural Language Processing (NLP)

The corpus (lexicon) preparation for extracting opinion quintuple (a_i, sa_{ij}, s_{ijkl}, r_k, p_l) is the main task for extracting the elements of aspects, sub-aspects and sentiment from unstructured data and then rearranging them into a structured platform, which will be treated as an attribute dimension of the cube. To concern about cube usage, a hierarchical platform need to be recognized. Lexicon characteristics should be declared words, phrases, or sentiment words for extracting at the lowest level of a cell of sub-aspects, supporting the cube properties such as slice/dice, drill-down, or roll-up. Not only standard dictionary but also special provided dictionary (lexicon) with well-identified contribute the sentiment analysis accuracy. The process starts from crawled unstructured data of user-generated content in "www.pantip.com," which is one of the famous Thai critiqued web-blogs in Thailand, including the life insurance domain. The data are utilized for preparing the corpus and lexicon database as our task makes data extraction from Thai unstructured data, which is not a standard language as English. It has an inclusive characteristic such as no stop-words at the end of a sentence, so it is quite challenging to manage long sentences because recognizing the end of a sentence is difficult [31,32].

Collections of corpus from pantip.com which the system crawled from many topics related to Thai life insurance have different volume data. So, we made preparing in unit group of words on sentence segmentation with conditions. One divided sentence contains at least 1 of subjects and 1 verb. The number of divided sentences which only related to Thai life insurance in our concerned aspects is 2,994. The number of words of this corpus is 21,016. Tagged corpus was prepared from random discussion topics and was annotated into each aspects and sub-aspects.

We do the text preprocessing by making sentence segmentation by divided the sentence as a chunk that contains at least 1 subject and 1 verb. A group of a chunk is separated from user-generated contents within blogs. The estimated sentence segmentation is around 200 - 300 Thai characters per 1 sentence (chunk). The important process as tokenization, normalization and word segmentation with long string matching, including POS tagging, also requires text preprocessing and NLP tasks [37,38]. After that, we made the process of **lexicon preparation**. Next, the lexicon preparation step, the special characteristic of non-standard language is challenging to split word with a standard dictionary. Therefore, sentiment extraction from the Thai language requires a particular lexicon for tracking free-form format from word-of-mouth in user-generated contents like pantip.com, full of transliterated words, chat language or slangs, etc. We prepare 3 types of the lexicon. The 1st lexicon is a **semantic term** of words that match with **LEXiTRON**, a Thai-English electronic dictionary provided by *NECTEC* [35] including transliterated words, ambiguous words, slang words and technical words. For example, "Retirement, Term" is an example of a basic product name referred to 'Product' aspect. "ETI - Extended Term Insurance," "RPU - Reduce Paid-up" is policy status referred to 'police operation' aspect. Next lexicon, **complement clue terms** consist of linguistic and non-linguistic form such as "ในช่วงเวลา, ในระหว่าง/during, อีกฝั่งหนึ่งของ/across from" is a preposition, "ความ, การ/~ing" is an abstract noun, "อย่างไรก็ตาม/however, nevertheless" in conjunction, "ควรจะ/should," "ต้อง/must" is an auxiliary verb, etc. [21]. The 3rd lexicon is a **compositional semantic clue term** that was tracking for pragmatics in words or phrases for special meaning and/or special sentiment, including sentiment words. For example, "ผลประโยชน์ทุพพลภาพ/ disability benefit," "ตรวจสอบเจอประวัติสุขภาพ/check and found abnormal health record," "สถานะสิ้นสุดกรมธรรม์/lapse," "ปฏิเสธการจ่ายสินไหม/reject the payment," เคลมไม่มีปัญหาอะไร/no problem on claim process" referred to "Claim assessment" aspect. This kind of lexicon is one of the main parts to increase semantic and accuracy on sentiment analysis. We utilized "Sansarn Corpus Tagging tools," which had designed and implemented to support this task [36].

Figure 6 Sansarn Tagging tools.

Example: Sentence (Chunk) "*ถ้าไม่แถลงไป ถ้าเขาสืบได้ว่าเราโกหก เขาจะไม่จ่ายค่ารักษาครับ มีสิทธิยกเลิกประกันได้ครับ และมีสิทธิ์เรียกร้องค่าสินไหมเขาเคยจ่ายไปแล้ว คืนมาได้ทั้งหมดครับ*/If we conceal and then company realizes that we are lying. The company can reject payment for our treatment. Have the right to cancel the policy. Moreover, the company can request to return all of the expenses which the company paid for us."

- Clue word (Regulation (Topic)-CLM): '*ไม่แถลงไป*/conceal,' '*สืบได้ว่าเราโกหก*/company realizes that we are lying,' '*ไม่จ่ายค่ารักษา*/reject payment for our treatment,' '*มีสิทธิ์เรียกร้องค่าสินไหมเขาเคยจ่ายไปแล้ว คืนมาได้ทั้งหมด*/company can request to return all of the expense which company paid for us.'
- Clue word (Operation (Topic)-PCR): '*มีสิทธิยกเลิกประกัน*/have the right to cancel the policy.'

Information extraction of unstructured data with a standard dictionary and special lexicon as our design aspects/sub-aspects provide the group with a MDSC dimension. In a part of measurements, they are identified by the number of matching words and total score polarity of sentiment words under matching words in each chunk and domain lexicon. The next process is **sentiment extraction**; the polarity of sentiment words is identified to increase the power of result analysis in each analysis scenario, especially when analyzing on MDSC. **Polarity calculation** makes more increase on the semantic intensity of words such as worse is much stronger modifier than bad. Giving score in sentiment lexicon was preparing by hand-tagging for giving score by applying a concept of *Turney* named *'SO-CAL'* [37]. SO-CAL concept determines the score using the same five (5) scale to extraordinarily positive and minus five (–5) scale for extremely negative when 0 is a neutral sentiment on a noun, verb, adverb and phrase. The grading scale of feeling expressions shows in the SO-value column in **Table 1**.

Table 1 Sentiment scoring adjustment.

Feeling Level	SO-value	Intensifier	Modifier%
excruciatingly	–5	somewhat	–30 %
inexcusable	–3	pretty	–10 %
foolishly	–2	really	+15 %
satisfactorily	1	very	+25 %
purposefully	2	extraordinarily	+50 %
hilariously	4	(the)most	+100 %

The total sentiment words scoring in each chunk can be measured as the sum of scores of the s_j is computed score (w_j) of all words $w_j \in s_i$ multiplied with their respective weights weight (w_j) under the rules of sentiment scoring adjustment (**Table 1**):

$$\text{Score} (s_i) = \Sigma score (w_i) \times weight (w_j)$$

For example: "*สุดระอา* กับการเคลมประกันของบริษัทA"/"**Very fed up** with claim process of insurance company A."

⇨ Very = a lot (fed up: –5) = –3 + (–5×100 %) = –8.

⇨ Aspect: Claim assessment get –8 score on this chunk of words.

After we had already prepared in unstructured data part, it is the process of managing data in terms of structured data into a database with the OLAP concept. After loaded unstructured data into the database by extract, transform and load (ETL), data preparation is required. The next process is cleansing data and data population as our designed fact tables. This format was designed for supporting MDSC generation.
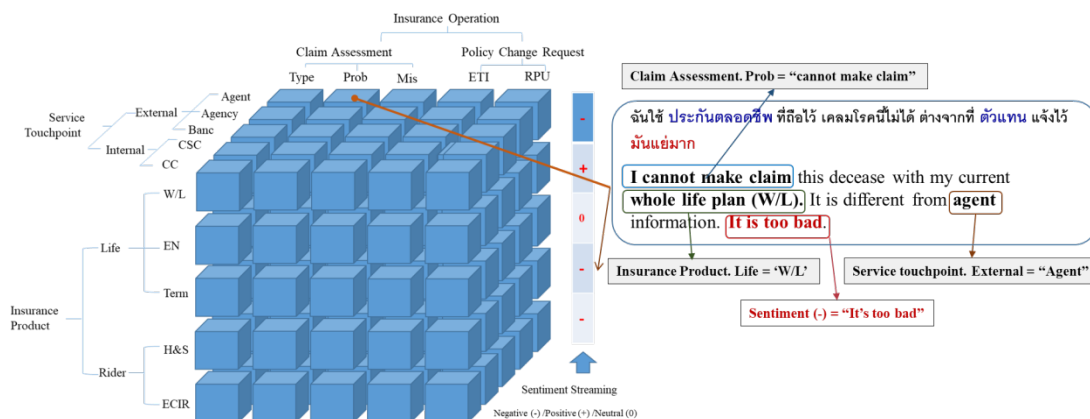
Multidimensional sentiment cube generation

Figure 7 An example of a MDSC.

The traditional multidimensional model is a pattern of the multidimensional fact-based structure generated from complex queries. A fact set up from analysis viewpoints of some concerning events called 'dimension' (e.g., service touchpoint and service process.) The data can be quantified from measures called 'measurement' (e.g., number of complaint words). The measurements can be aggregated into different levels of concepts across different layers of multi-dimensions as hierarchical forms.

A data cube provides multidimensional perspective viewpoints through multi-measurements combination to multi-dimension values called 'multidimensional cell.' An analysis point of view can vary by typical OLAP operators such as 'drill-down,' 'roll-up,' 'slice and dice,' 'pivot' or 'top-k selection.'

MDSC emerges in the era of social media, which raises sentiment analysis techniques. The sentiment is deployed to be one of the analysis dimensions. That makes it more powerful on analysis viewpoints. From our case study in Thai life insurance, a multidimensional sentiment conceptual model is generated as our design aspects/sub-aspects to be dimensions and measurements as defined below.

```
define lifeINS_cube [COM, PRD, SVT, SKP, CMT, CMP, IML]:
SKE_CNT = sum(SKE_WC),
SKE_SCR= sum(SKE_CR),
CMP_CNT=sum(CMP_WC),
CMP_SCR= sum(CMP_CR),
IMP_CNT= sum(IMP_WC),
IMP_SCR= sum(IMP_CR),
NEG_CNT= sum(NEG_WC),
NEG_SCR= sum(NEG_CR),
POS_CNT= sum(POS_WC),
POS_SCR= sum(POS_CR),
SEN_CNT = sum(SEN_WC)
define dimension COM as (COM_NM),
define dimension PRD as (LF, RD, AFF, SPC, OTH)
define dimension SVT as (INS, EXS)
define dimension SKE as (REL, CNS, RSP, KLG, CHR)
define dimension CMT as (MNC, MJC)
define dimension CMP as (ADM, RJP, CNC, FAX, LSV, MSC, MST, MSP, MSE)
```

```
define dimension IML as (MOC, MPO, MAA, LON, INV, VOD, SUR, REJ, PLC)


Remark:
COM:COMPANY; PRD:PRODUCT;SVT:SERVICE_TOUCHPOINT; SKE:SERVICE_KEY_EVALUATION; CMT:CLAIM_TYPE; CMP:CLAIM_PROBLEM;
IML:IMPACT_LEVEL;
SKE_CNT:SERVICE_KEY_EVALUATION_COUNT; SKE_SCR:SERVICE_KEY_EVALUATION_SCORE;
CMP_CNT:CLAIM_PROBLEM_COUNT;
CMP_SCR:CLAIM_PREBLEM_SCORE;
IMP_CNT:IMPACT_LEVEL_COUNT;
IMP_SCR:IMPACT_LEVEL_SCORE;
NEG_CNT:NEGATIVE_COUNT; NEG_SCR:NEGATIVE_SCORE;
POS_CNT:POSITIVE_COUNT; POS_SCR:POSITIVE_SCORE;
COM_NM:COMPANY_NAME;LF:LIFE,RD:RIDER;AFF:AFFINITY_PROD; SPC:SPECIAL;OTH:OTHER; INS:INTERNAL_SERVICE;
EXS:EXTERNAL_SERVICE; REL:RELIABILITY;
CNS:CONSISTENCY; RSP:RESPONSE;
KLG:KNOWLEDGE; CHR:CHARACTERISTIC;
MNC:MINOR_CLAIM; MJC:MAJOR_CLAIM;
ADM:ADVANCE_MONEY; RJP:REJECT_PAYMENT;
CNC:CONCEAL; LSV:LONG_SERVICE;
MSC:MISUNDERSTANDING_IN_COVERAGE;
MST:MISUNDERSTANDING_IN_PROTECTION;
MSP:MISUNDERSTANDING_IN_PERIOD;
MSE:MISUNDERSTANDING_IN_EXCEPTION;
MOC:MAKE_SUE_ON_COURT;
MPO:MAKE_PETITION_FOR_OIC;
MAA:MAKE_ACCUSATION_AGENT;
LON:LOAN; INV:INVALID_POLICY;
VOD:VOID_POLICY; SUR:SURRENDER_POLICY;
REJ:POLICY_REJECT; PLC:POLICY_CANCELLATION;
SEN_CNT:SENTENCE_COUNT
(WC:WORD_COUNT;SC:SCORE)
```

Results and discussion

The system generated the cube using IBM Cognos version 10.2 BI application with Microsoft SQL Server version 11.0.2 database. **Figure 8** shows an example result of our methodology called MDSC. This analysis results reveal the negative score in the sentiment of a customer on a single problem and co-occurrence problems in the claim process such as 'Advance money‖Reject payment,' 'Reject payment‖Conceal,' etc. We deploy this decision point's analysis to generate the event logs on the hierarchical processes to be a source of classification rules for process mining.
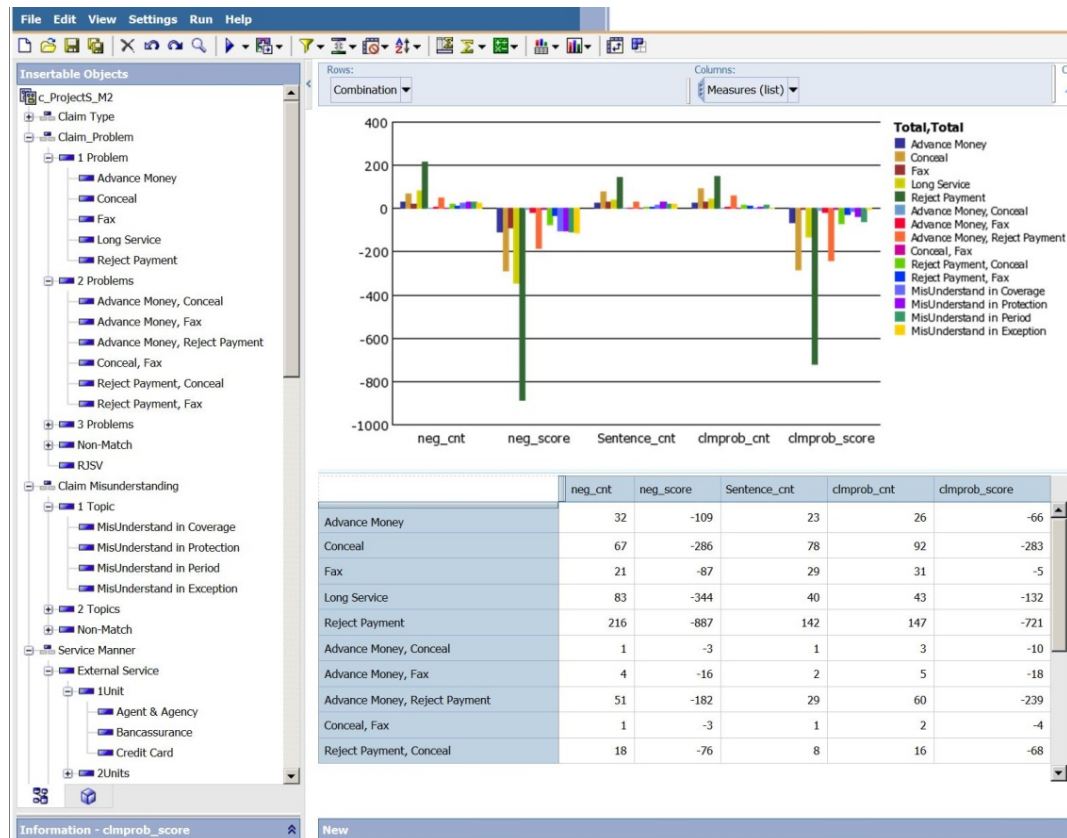

Figure 8 An example result of MDSC.

Data mining rules as each cube scenario

We use a MDSC to scope the problems and set up analysis scenarios. We then find out data mining rules using association rules, one of the data mining techniques that find a data relationship, from the large data available to find frequent patterns for analyzing relationships or prediction phenomena.

We use Weka data mining tools to generate the rules related to the acquisition process, which related features under MDSC scenario are company name, product, service touchpoint, service evaluation and sentiment. The rules are found with minimum support of 10 %, and maximum confidence is 60 % below (**Table 2**).

Remark: XXNEGLV1 summarizes negative scores in the initial level, usually express with common complaint words. XXNEGLV2 expresses a negative score in medium level typically use with severe complaint word. XXNEGLV3 expresses a negative score in high level usually use with aggressive complaint word, XX is a feature.

Customer acquisition

The data mining rules related to the acquisition process, which related features under MDSC scenario are company name, product, service touchpoint, service evaluation, impact and sentiment. The rules are found with minimum support is 10 %, and maximum confidence is 80 % as below.

Dissatisfaction in service of data mining rules shows service in life insurance requires improving on consistency (s_consist), reliability (s_realize) of external service manner (c_es). The rules mean that service quality is less reliable, and service is inconsistent.

Table 2 Association rules on acquisition stage.

LHS	NLH		RHS	NRH	Conf (%)
s_consist = CSNEGLV1	59	→	s_neg_service = NSERNEGLV1	40	67.8
s_realize = RLNEGLV1	100	→	s_neg_service = NSERNEGLV1	67	67
s_consist = CSNEGLV1	59	→	company = XXX	38	64.4
s_neg_neg = NGNEGLV1	182	→	c_es = ES	117	64.3

Customer intention

The data mining rules related to the intention process, which related features under MDSC scenario are company name, product, service touchpoint, service evaluation, claim problem, impact and sentiment. The rules are found with minimum support is 10 %, and maximum confidence = is 80 % as below.

Dissatisfaction in service of data mining rules shows claim (c_claim) process mostly minor claim (c_mnc) related to claim problem on 'reject payment' and 'rejected policy' got the customer's high negative sentiment. From rules mean if reject payment in the claim process has occurred, it is a high opportunity to make policy cancellation or policy reject (**Table 3**).

Table 3 Association rules on intention stage.

LHS	NLH		RHS	NRH	Conf (%)
company = XXX_claim = CLM s_reject = RJNEGLV1	66	→	s_rejpm = RJPNEGLV1	66	100
s_neg_neg = NGNEGLV2 s_reject = RJNEGLV1	63	→	s_neg_service = NSERNEGLV1	63	100
c_claim = CLM s_neg_claim = NCLMNEGLV1	48	→	s_rejpm = RJPNEGLV1	48	100
s_cancel = CANNEGLV1	67	→	s_surrender = SRDNEGLV1	64	95.5
c_mnc = MNC	54	→	s_rejpm = RJPNEGLV1 s_reject = RJNEGLV1	44	81.5

Customer engagement

The data mining rules related to the engagement process, which related features under MDSC scenario are company name, product, service touchpoint, service evaluation, claim problem, misunderstanding point and sentiment. The rules are found with minimum support is 10 %, and maximum confidence is 80 % as below.

Dissatisfaction in service of data mining rules show claim (c_claim) process with misunderstanding in budget (s_misBDG) frequency occur with company XX, Product H & S in a minor claim (c_mnc) has negative sentiment (NGNEGLV1) with high frequency occur with misunderstanding in protection (s_misprotect). Besides, misunderstanding in period (s_mistm) has negative (MTMNEGLV1) frequency occur in company XX. The data mining rule means that customers require more understanding of the policy condition. Mainly, in claim procedure which lacks insurance knowledge in term of the exceptional on "total assured (return money) under different condition of claim," "the different of exceptional decease in each customer," and "waiting period of the claim process in each decease" (**Table 4**).

Table 4 Association rules on engagement stage (misunderstanding analysis).

LHS	NLH		RHS	NRH	Conf (%)
product = HS c_mnc = MNC	7	→	s_misprotect = MPTNEGLV1	7	100
company = XX product = HS c_claim = CLM	10	→	s_misprotect = MPTNEGLV1	9	90
product = HS c_claim = CLM s_misprotect = MPTNEGLV1	11	→	company = AIA	9	81.8

Misunderstanding analysis

Dissatisfaction in service of data mining rules shows a request petition on OIC (s_sueoic) come from negative service (s_neg_service), reject payment (s_rejpm) with sue on-court (s_suelaw) frequency occur policy reject status (s_reject), company XX with reject payment (s_rejpm) return to the customer, usually make negative (RJNEGLV1) and make customer said about sue on-court (s_suelaw), and frequency occur policy reject status (s_reject). The rule shows Thai life insurance company XY frequency occurs make accuse to the agent (s_sueagent) with negative sentiment (SAGTNEGLV1) (**Table 5**).

Table 5 Association rules on engagement stage (high impact analysis).

LHS	NLH		RHS	NRH	Conf (%)
s_neg_service = NSERNEGLV1	9	→	s_sueagent = SAGTNEGLV1	9	100
s_sueoic = SOICNEGLV1	5	→	s_neg_service = NSERNEGLV1	5	100
s_rejpm = RJPNEGLV1 s_suelaw = SLWNEGLV1	13	→	s_reject = RJNEGLV1	13	100
company = XX s_reject = RJNEGLV1 s_suelaw = SLWNEGLV1	10	→	s_rejpm = RJPNEGLV1	10	100
company = XY	6	→	s_sueagent = SAGTNEGLV1	5	83.3

Identify process monitoring on BPMN

All event logs generated from the MDSC in our case study have the analysis result on event logs of cube and data mining rules. This section uses the benefit of monitoring the current process in the BPMN flow. To apply the monitoring results for improvement in change process mining, consider event cube logs in customer acquisition, intention and engagement scenario, including data mining rules. We found the related processes which get dissatisfaction from customer and require reviewing for make process improvement.

To analyze along with BPMN flow in **Figure 7** to specific weak process, the recommendations for improvement in this process on customer acquisition stage are related to processes below.

Question 1: Does it have any dissatisfaction issues from customers that occurred in the customer acquisition stage? And what are the kind of problems and related issues?

Answer: Dissatisfaction in Service of Data mining rules shows service in life insurance requires improving consistency, reliability and external service manner.

From the results of association rules mining found that service quality is less reliable, and service is inconsistent. Therefore, the processes that need to concern for improving the reliability is *service manner when contact to the customer*, especially external service such agents (green color in **Figure 9**).

From the A-line area of **Figure 9**, the service manner process has participation with the customer. It requires satisfaction improvement on the process of 'sign agreement to make new life & health insurance contract' and 'propose a new application to approve the new policy.' B-line area has concerned process on 'inform application result or offer a new plan to the customer.' C-line area in the claim process has concerned process on 'contact/request service from a direct servicing agent by customer.' All concerned processes require the service agent to action more reliability.
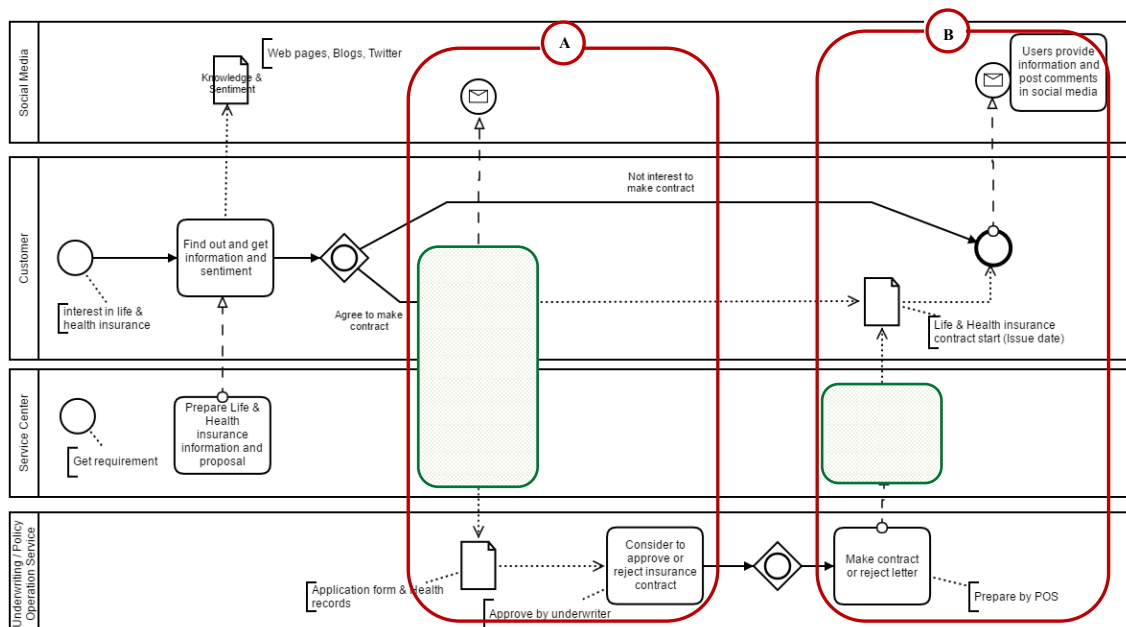

Figure 9 To identify dissatisfaction processes in the customer acquisition stage.

Question 2: Does it have any dissatisfaction issues from customers that occurred in the customer retention stage?

Answer: Dissatisfaction in service of data mining rules shows the claim process, mostly minor claims related to claim problem on reject payment, and rejected the policy, got the customer's high negative sentiment (orange color in **Figure 10**).

From the results of association rules mining found that reject payment in the claim process has occurred, it is a high opportunity to make policy cancellation or policy reject. Therefore, the processes that need to concern for improving the process of reject payment to the customer when claim assessment occurred, the claim department will verify policy conditions matching with claim activities, including the customer's health records, before paying money return to the customer. After investigation, if there are any abnormal or suspect events, the company can pay some money return, reject the payment, or cancel the policy. Reducing this problem to reduce dissatisfaction requires a clear understanding of the

customer's conditions, including recommending the customer to reveal the accurate health record without concealing.

D-line area, the concerned processes are 'check balance claim budget under health(rider) policy,' 'Sign an agreement with new policy status or policy cancellation' which is the impact from 'find abnormal case reject the payment,' 'request POS policy cancel,' 'investigate abnormal case and change policy status,' 'change policy status,' 'cancel health (rider) product,' 'cancel policy contract.'
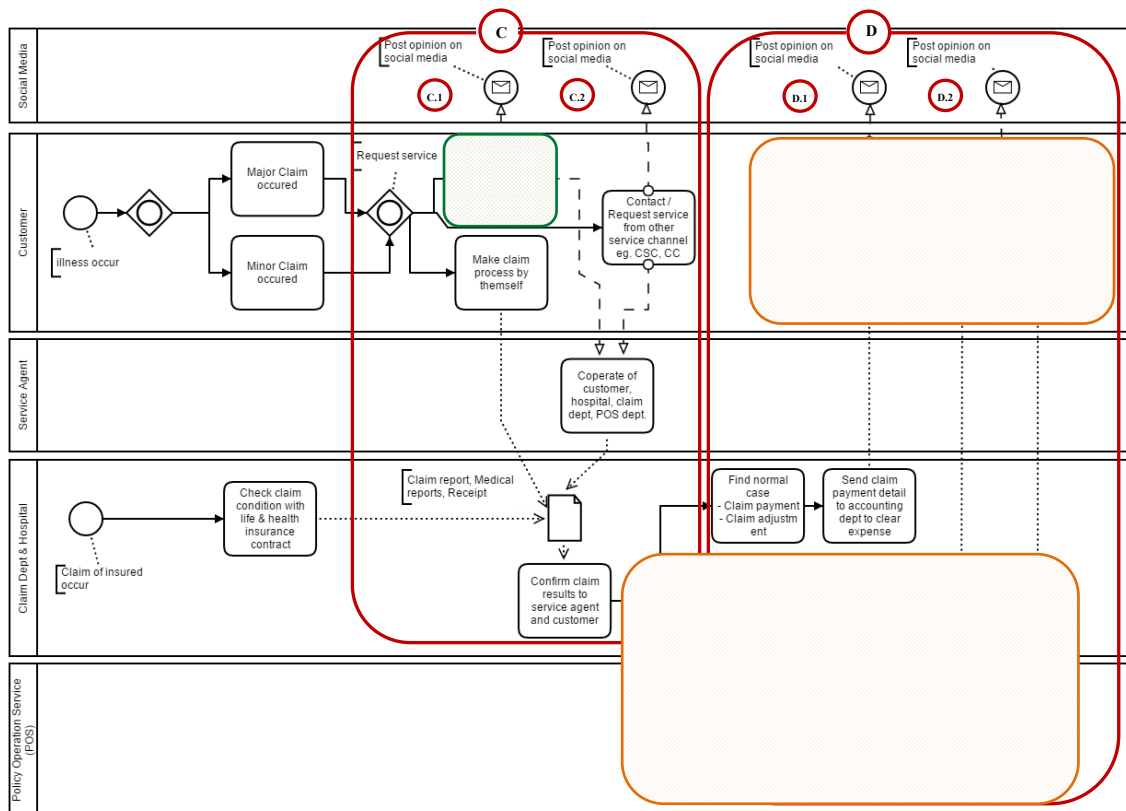

Figure 10 To identify dissatisfaction processes on Customer Intention and Engagement stage.

Question 3: Does it have any dissatisfaction issues from customers that occurred in the customer engagement stage?

Answer: Dissatisfaction in service of misunderstanding from the results of association rules mining shows that the claim process with *misunderstanding in budget* frequency occurs in minor claims has negative sentiment with high frequency that occurs with *misunderstanding in protection*. Moreover, dissatisfaction in service of high impact feature shows a *request petition on OIC* come from hostile service, *reject payment with sue on-court* frequency occur policy reject status with reject payment return to the customer, usually make harmful and make customer said about sue on court and frequency occur policy reject status.

High impact occurs from *rejecting payment*, so the concerned processes are the same as item 2 (D-line area in orange color in **Figure 10**). However, the rule shows reject payment high related to hostile service and 1 event log from sentiment cube show the make accuse to agent high related to concealing the problem. To solve this problem for getting better customers' satisfaction needs to improve from agent behavior, an agent should strongly recommend to the customer to reveal the accurate health records, including all of the severe conditions when giving service to the customer. Furthermore, the misunderstanding problems can also solve the problem if the agent or company provides more exceptional customer details or increases some business flow process to confirm customers' understanding before customer sign acceptable in a new insurance contract.

This methodology can adjust not only in the other processes of life insurance but also this method can adapt to the other business. However, the results depend on which scope of analysis. The MDSC and BPMN requires to design in the same direction in order to get different event logs generated from MDSC for data set of association rules mining process.

Conclusions

This paper presents a MDSC mining that deployed text mining and NLP to consider social CRM on Thai life insurance business to monitor business processes with BPMN. The MDSC results can reveal the significant association among aspects and sub-aspects, especially the design with sentiment analysis elements as a dimension that can help increase analysis viewpoints for recognizing in-process monitoring. Association rules mining results generated from MDSC data provide concrete evidence to support the cube analysis viewpoint. The rules also explicitly point out the related processes for monitoring tasks in the real business model case via BPMN. However, one of the importance of this method is the designation among aspects and sub-aspects in in-depth details and correctly synchronized with process notation can provide more precise results for monitoring the significant processes which require improving due to get the high negative sentiment.

Acknowledgements

References

[1]   L Zhang and B Liu. *Sentiment analysis and opinion mining. In*: C Sammut and GI Webb (Eds.). Encyclopedia of machine learning and data mining. Springer, Boston, Massachusetts, 2016, p. 1-10.
[2]   M Hu and B Liu. Mining opinion features in customer reviews. *In*: Proceedings of the 19th National Conference on Artifical Intelligence, San Jose, California. 2004, p. 755-60.
[3]   A Bolt and WMPVD Aalst. Multidimensional process mining using process cubes. *In*: Proceedings of the International Conference on Enterprise, Business-Process and Information Systems Modeling, Stockholm, Sweden. 2015, p. 102-16.
[4]   Business Process Modeling Notation. TechTarget, Available at: http://searchcio.techtarget.com/definition/Business-Process-Modeling-Notation, accessed June 2020.
[5]   S Mansmann, T Neumuth and MH Scholl. Multidimensional data modeling for business process analysis. *In*: Proceedings of the 26th International Conference on Conceptual Modeling, Auckland, New Zealand. 2007, p. 23-38.
[6]   T Vogelgesang, G Kaes, S Rinderle-Ma and HJ Appelrath. Multidimensional process mining: Questions, requirements, and limitations. *In*: Proceedings of the 28th International Conference on Advanced Information Systems Engineering, Ljubljana, Slovenia. 2016, p. 169-76.
[7]   N Sucharittham, TH Nguyen, HC Dam and T Theeramankong. Customers' sentiment in Thai life insurance. *Int. J. Manag. Bus. Econ*. 2020; **7**, 55-74.
[8]   N Sucharittham, T Theeramunkong, C Haruechaiyasak, TH Bao and HC Dam. Data mining for life insurance knowledge extraction: A survey. *In*: Proceedings of the 2nd Asian Conference on Information Systems, Phuket, Thailand. 2013, p. 111-8.
[9]   N Sucharittham, T Theeramunkong, C Haruechaiyasak and HC Dam. Sentiment and topic analysis on life insurance business in Thailand. *In*: Proceedings of the 3rd Asian Conference on Information Systems, Nha Trang, Vietnam. 2014, p. 239-46.
[10] JTS Ribeiro and AJMM Weijters. Event cube: Another perspective on business processes. *In*: Proceedings of the On the Move to Meaningful Internet Systems, Crete, Greece. 2011, p. 274-83.
[11] A Bolt and WMPVD Aalst. Multidimensional process mining using process cubes. *In*: Proceedings of the 20th International Conference on Exploring Modeling Methods for Systems Analysis and Design, Stockholm, Sweden. 2015, p. 102-16.
[12] T Vogelgesang and HJ Appelrath. Multidimensional process mining with PMCube explorer. *In*: Proceedings of the BPM Demo Session 2015 Co-located with the 13th International Conference on Business Process Management, Innsbruck, Austria. 2015, p. 90-4.
[13] T Vogelgesang and HJ Appelrath. PMCube: A data-warehouse-based approach for multidimensional process mining. *In*: Proceedings of the Business Process Management Workshops, Innsbruck, Austria. 2015, p. 167-78.

[14] F Ravat, O Teste, R Tournier and G Zurlfluh. A conceptual model for multidimensional analysis of documents. *In*: Proceedings of the 26th International Conference on Conceptual Modeling, Auckland, New Zealand. 2007, p. 550-65.

[15] X Liu, K Tang, J Hancock, J Han, M Song, R Xu and B Pokorny. A text cube approach to human, social and cultural behavior in the Twitter stream. *In*: Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Washington, DC. 2013, p. 321-30.

[16] D Zhang, C Zhai and J Han. Topic cube: Topic modeling for OLAP on multidimensional text databases. *In*: Proceedings of the SIAM International Conference on Data Mining, Sparks, Nevada. 2009, p. 378-95.

[17] S Kim, J Zhang, Z Chen, AH Oh and S Liu. A hierarchical aspect-sentiment model for online reviews. *In*: Proceedings of the 27th AAAI Conference on Artificial Intelligence, Bellevue, Washington. 2013, p. 526-33.

[18] K Dave, S Lawrence and DM Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *In*: Proceedings of the 12th International Conference on World Wide Web, New York. 2003, p. 519-28.

[19] U Dayal, C Gupta, M Castellanos, S Wang and M Garcia-Solaco. Of cubes, DAGs and hierarchical correlations: A novel conceptual model for analyzing social media data. *In*: Proceedings of the 31st International Conference on Conceptual, Florence, Italy. 2012, p. 30-49.

[20] D Umeshwar. 2012, Sentiment Cube, U.S. Patent 20120197950 A1.

[21] TechTarget. Managing the customer lifecycle: Customer retention and development, Available at: http://media.ttgtmedia.com/searchCRM, accessed June 2020.

[22] JV Doorn, KN Lemon, V Mittal, S Nass, D Pick, P Pirner and PC Verhoef. Customer engagement behavior: Theoretical foundations and research directions. *J. Serv. Res*. 2010; **13**, 253-66.

[23] MR Yaakub, Y Li and J Zhang. Integration of sentiment analysis into customer relational model: The importance of feature ontology and synonym. *Procedia Technol*. 2013; **11**, 495-501.

[24] BD Weinberg and PD Berger. Connected customer lifetime value: The impact of social media. *J. Direct Data Digit. Mark. Pract*. 2011; **12**, 328-44.

[25] N Behravan and M Sabbirrahman. Customer relationship management constructs under social networks towards customers' retention. *Aust. J. Basic Appl. Sci*. 2012; **6**, 271-82.

[26] J Wiebe and E Riloff. Creating subjective and objective sentence classifiers from unannotated texts. *In*: Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing, Mexico City, Mexico. 2005, p. 486-97.

[27] R Sarno, RD Dewandono, T Ahmad, MF Naufal and F Sinaga. Hybrid association rule learning and process mining for fraud detection. *IAENG Int. J. Comput. Sci*. 2015; **42**, 59-72.

[28] R Sarno, F Sinaga and KR Sungkono. Anomaly detection in business processes using process mining and fuzzy association rule learning. *J. Big Data* 2020; **7**, 5.

[29] Q Suzhen. Data mining and business process management of Apriori algorithm. *IOP Conf. Ser.: Mater. Sci. Eng.* 2018; **439**, 032012.

[30] K Bohmer and S Rinderle-Ma. Mining association rules for anomaly detection in dynamic process runtime behavior and explaining the root cause to users. *Inf. Syst*. 2020; **90**, 101438.

[31] C Haruechaiyasak and A Kongthon. Constructing Thai opinion mining resource: A case study on hotel reviews. *In*: Proceedings of the 8th Workshop on Asian Language Resources, Beijing, China. 2010, p. 64-71.

[32] C Haruechaiyasak, A Kongthon, P Palingoon and K Trakultaweekoon. S-Sense: A sentiment analysis framework for social media monitoring applications. *Inf. Technol. J.* 2018; **14**, 11-22.

[33] N Tongtep and T Theeramunkong. Multi-stage annotation using pattern-based and statistical-based techniques for automatic Thai annotated corpus construction. *In*: Proceedings of the 9th Workshop on Asian Language Resources, Chiang Mai, Thailand. 2011, p. 50-8.

[34] N Tongtep and T Theeramunkong. Multi-stage automatic NE and PoS annotation using pattern-based and statistical-based techniques for Thai corpus construction. *IEICE Trans. Inf. Syst.* 2013; **E96-D**, 2245-56.

[35] LEXiTRON: An online Thai-English electronic dictionary, NECTEC, Available at: http://lexitron. nectec.or.th, accessed June 2020.

[36] Sansarn.com, NECTEC, Available at: http://sansarn.com/corpus-tagging/health/tag.php, accessed June 2020.

[37] M Taboada, K Voll and J Brooke. Extracting sentiment as a function of discourse structure and topicality. *Tech. Rep*. 2008; **20**, 1-22.