6-18-2013

Sensor-based Knowledge Discovery from a Large Quantity of Situational Variables

Norihiko Moriwaki
Hitachi, Ltd., norihiko.moriwaki.qw@hitachi.com

Miki Hayakawa
Hitachi, Ltd., miki.hayakawa.qv@hitachi.com

Norio Ohkubo
Hitachi, Ltd., norio.ohkubo.jh@hitachi.com

Kazuo Yano
Hitachi, Ltd., kazuo.yano.bb@hitachi.com

Dai Senoo
Tokyo Institute of Technology, senoo.d.aa@m.titech.ac.jp

SENSOR-BASED KNOWLEDGE DISCOVERY FROM A LARGE QUANTITY OF SITUATIONAL VARIABLES

Norihiko Moriwaki, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, norihiko.moriwaki.qw@hitachi.com

Miki Hayakawa, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, miki.hayakawa.qv@hitachi.com

Norio Ohkubo, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, norio.ohkubo.jh@hitachi.com

Kazuo Yano, Central Research Laboratory, Hitachi, Ltd., Tokyo, Japan, kazuo.yano.bb@hitachi.com

Dai Senoo, Department of Industrial Engineering and Management, Tokyo Institute of Technology, Tokyo, Japan, senoo.d.aa@m.titech.ac.jp

Abstract

A new methodology called "sensor-based knowledge discovery", which utilizes wearable sensors and statistical analysis, is proposed and evaluated. This methodology facilitates identifying new knowledge that can improve business outcome. It utilizes wearable sensors to unobtrusively capture people's location, motion, and social interaction with others. The captured data is converted into multi-dimensional situational variables and then statistically analyzed to deliver a "rule set," which forms the basis of new knowledge related to business outcome. The methodology was evaluated through a case study at a retail store. A hypothetical rule, that is, a particular area (a so-called "hot spot") in the store where employee's presence correlates with average sales per customer, was identified. Based on the identified rule, a measure to concentrate employees in that area was initiated. Consequently, increasing employees' presence ("staying time") in the hot spot by 70% increased average sales per customer by 15%. This result demonstrates the effectiveness of the methodology; namely, the new sensor-based knowledge discovery can improve actual business performance.

Keywords: Knowledge Discovery, Wearable Sensor, Situational Variables, Statistical Analysis

1 INTRODUCTION

Managing knowledge is critically important for improving business performance in companies (Grant 1996; Teece 1998). Moreover, to cope flexibly with the changing business situation, a methodology for discovering knowledge ("knowledge discovery," hereafter) on the basis of factual data is needed (Hey et al. 2009). In a situation where a vast amount of data is available, a "hypothesis-free" mode of knowledge discovery is more advantageous than hypothesis-driven discovery (namely, "first hypothesize, then experiment") (Ozdemir, 2011). Accordingly, an efficient "data-intensive" knowledge discovery should be developed.

Although several "knowledge-discovery and data mining" (KDD) technologies have been proposed (Fayyad et al., 1996; Piatetsky-Shapiro 2007; Kriegel et al., 2007; Turban et al., 2008), few of them automate knowledge discovery from data. For example, "process mining" deals with log data in a company to identify business-process models. Several modelling algorithms for reproducing business processes by utilizing data characteristics and event orders have been proposed (van der Aalst, et al., 2003; van der Aalst et al., 2004). Since process mining utilizes a limited data set accumulated in a business-information system, such as a workflow-management system, the lack of data regarding real-world situations has inhibited new knowledge discovery (Leyer, 2011; Van der Aalst, et al. 2003).

For discovering knowledge without reliance on prior assumption or hypothesis, accurate and comprehensive understanding of a situation is required (Dahlbom et al., 2009; Ye at al., 2012). For example, the importance of situation has been studied on the basis of "situational-awareness" (SA) theory (Endsley, 1995; Adams, 1995). SA deals with human abilities for monitoring and grasping the self and surroundings in a current situation so that the near future can be predicted and recognized with precision. It is a critical element in high-quality decision making. Moreover, automating situation analysis has mainly been studied from the viewpoint of "context awareness" (Boytsov, 2012), which aims to provide services anytime and anywhere by capturing people's location, attributes, purpose, etc. For example, using the idea of process mining, Layer (2011) investigated how context affects the performance of a business process. That investigation, however, is limited because of a lack of situational data regarding employees' activities other than documented activities.

Although the above-described attempts aimed at data-intensive knowledge discovery have been made, no practical hypothesis-free knowledge discovery that improves business outcomes has been developed so far. To fill this gap, in this study, a new methodology, called "sensor-based knowledge discovery," which unobtrusively recognizes situations and statistically discovers rules that relate to business outcome, is proposed. Sensor-based knowledge discovery employs wearable sensors to digitally capture people's location, motion, and social interaction with others (Wakisaka at al., 2009). The captured data is then converted into situational variables. Finally, statistical analysis is applied to the variables to deliver a "rule set" related to a business outcome. The obtained rule set includes a knowledge source that is related to the business outcome. Since the rule set indicates a quantitative relationship between "situational elements" and a business outcome, they help managers to select practical rules from the rule set. The managers can therefore discover "new knowledge" after the rules are applied in a business process.

The proposed methodology, namely, sensor-based knowledge discovery, is demonstrated by using empirical data from a retail store. Among the derived rule set, a hypothetical rule was discovered that there is the particular "hot spot" in the store where the presence of employees strongly affects the average sales per customer. On the basis of that rule, in particular, increasing employee presence time in the hot spot by 70% led to a 15% increase in average sales per customer. The effectiveness of sensor-based knowledge discovery was thus recognized by the managers of the store.

The paper is structured as follows: Section 2 discusses related work concerning knowledge discovery. Section 3 explains the proposed methodology, namely, sensor-based knowledge discovery. Section 4 demonstrates the methodology by means of a case study. Section 5 presents the results of the case study. Finally, Section 6 offers some concluding remarks.

2 RELATED WORK

2.1 Knowledge Discovery

“Knowledge discovery and data mining” (KDD) is a research area focusing on methodologies for extracting useful knowledge from data (Fayyad et al., 1996; Piatetsky-Shapiro, 2007; Kriegel et al., 2007; Turban et al., 2008). Some research regarding automatic knowledge discovery employs statistical pattern matching targeting text analysis (Kearns & Schapire, 1994; Lin et al., 2001; Joachims, 2002). Other researches have focused on activity recognition. One focus area has been on detecting gestures and relatively low-level activities such as running, walking, sitting, shaking hands, and vacuuming (Bao & Intille, 2004; Bailey & Elkan, 1994). These researches aim at reproducing human intelligence algorithmically as opposed to discovering new knowledge.

As for process mining, several algorithms for process discovery taking heuristic and genetic-algorithm (GA) approaches have been proposed (van der Aalst et al. 2003; de Medeiros et al. 2006; Medeiros et al. 2006; Weijters & Ribeiro 2011). Some organizational aspects of business processes, such as social networks and interaction patterns of the participators in the processes, have also been analyzed (van der Aalst & Song 2004). Process mining makes it possible to produce objective pictures of how a process has been executed by utilizing event logs such as documented activities. Moreover, those event logs have been gathered and utilized to clarify how context affects performance of business processes (Leyer 2011; Ramos et al. 2010). In those analyses, however, social activities of employees in the real world (like face-to-face interactions) were neglected. In other words, situational factors and their impact on a business process in the real world have not been analyzed sufficiently (Rosemann et al. 2008; Tiwari & Turner 2008; Ploesser et al. 2009; van der Aalst 2011; Schonig et al. 2012; Wil et al. 2012) .

2.2 Situation and Context Awareness

Situational awareness (SA) is the perception of environmental elements with respect to time and/or space (Endsley 1995; Adams 1995). SA involves being aware of what is happening in the vicinity in order to understand how information, events, and one's own actions will impact goals and objectives, both immediately and in the near future. Having complete, accurate, and up-to-the-minute SA is essential when technological and situational complexity is involved in decision-making. SA has been recognized as a critical, yet often elusive, foundation for successful decision-making. Though SA has mainly been studied in areas like aviation and air-traffic control, the idea holds true for business management (Durso 2008; Liu et al. 2008).

In a similar manner to automated SA, context awareness targets flexible services for people using mobile-computing devices. Some recent research has focused on contextual issues associated with syntactic relationships. In many situations, context parameters are available from the activity in the environment (Kofod-Petersen et al. 2005; Kofod-Petersen & Mikalsen 2006). It is known that focusing on activities leads to a better understanding of context. Several approaches for examining activities, such as “situated action” (Suchman 1987) and “activity theory” (Vygotski 1978; Leont’ev 1978), have been proposed. These researches mainly cover activities regarding an individual’s context. Nevertheless, they lack the viewpoint of grasping a group of individuals and their social interactions.

3 METHODOLOGY OF SENSOR-BASED KNOWLEDGE DISCOVERY

3.1 Overview

A new methodology, called "sensor-based knowledge discovery," which aims at filling the gap identified in the previous section, is proposed in this section. This methodology identifies situational elements (including people's behavior and social interaction) that relate to business performance. Aiming to replace theory-based hypothesis construction, the methodology focuses on utilizing situation capturing by wearable sensors and comprehensive relational computation. As shown in Figure 1, the methodology involves three stages. In the first stage, the situation involving a social activity is precisely captured by sensors. Utilizing a wearable sensor makes it possible to digitize a situation such as location, face-to-face (F-to-F) interaction, and motion with a time stamp. In the second stage, the digitized data is aggregated and aligned into situational variables. The digitized data concerning each person's activity time-frame is then aggregated and categorized into multi-dimensions of the situation. In the third stage, situational variables that strongly relate to business outcome are extracted statistically. For example, regression analysis is applied to situational variables satisfying the correlation significance between a predetermined outcome variable and all the situational variables.

The requirements concerning sensor-based knowledge discovery are summarized as follows:

- Situational data regarding people's activities, such as location, interaction, and motion with a precise time-stamp.
- Explicit outcome variables concerning business performance.
- Sufficient computational power for comprehensive statistical calculation.

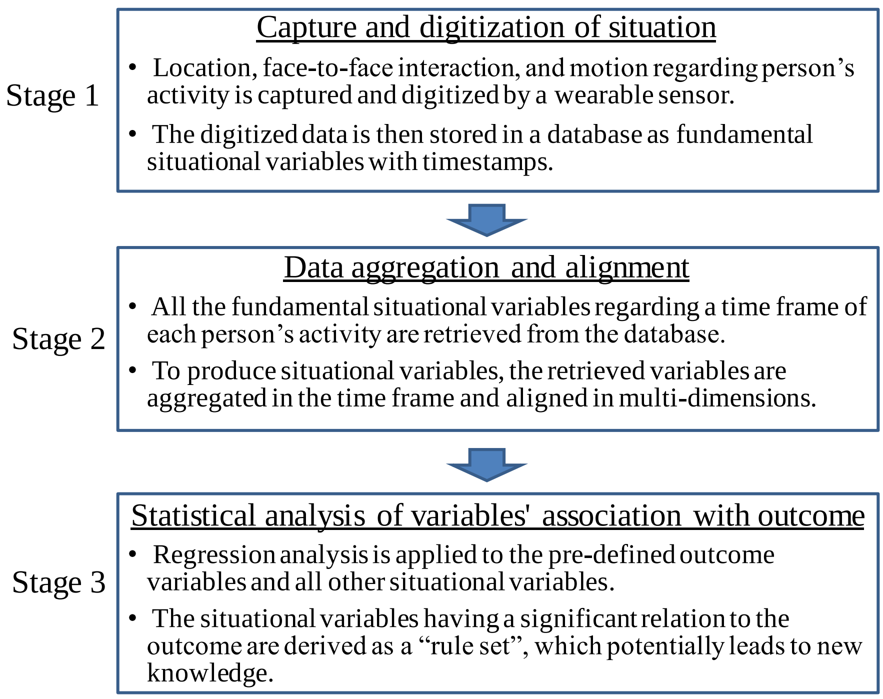

Figure 1. Three-stage procedure of sensor-based knowledge discovery

3.2 Stage 1: Capture and digitization of situation

3.2.1 Wearable sensor technology

A business situation is digitized by utilizing wearable sensor technology. Hitachi's "Business Microscope" is one such technology (Wakisaka 2009), namely, a name-tag-type sensor node, for measuring face-to-face interaction and body motion (Figure 2). Interactions between people are captured automatically and unobtrusively when the people wear the sensor node. Face-to-face interactions between people are captured and stored as follows: sensor nodes send and receive IDs, which are uniquely pre-assigned to all nodes, by infra-red signal automatically when they are within a range of two meters and an angle of 120 degrees. An infra-red signal has a directional characteristic and is therefore suitable for detecting face-to-face interactions between people wearing the sensor node. In this manner, the nodes capture the IDs of each other with timestamps. This ID exchange with time information enables analysis of face-to-face interaction in terms of quantity and frequency.

"Body-motion rhythm" of the wearer is captured by an accelerometer on the sensor node. The accelerometer precisely captures slight movements like nodding as a frequency rhythm. From characteristics of the frequency rhythm, people's emotional states, as well as activities like walking and running, can therefore be inferred (Pentland 2008; Ara et al. 2009). These data are the unique situational data obtained by the wearable sensor.

IR (infra-red) beacons for detecting sensor-node wearers around them are installed at specific locations. Each IR beacon transmits a unique ID continuously over a range of approximately two meters so that the sensor nodes capture the ID with a timestamp when they are around an IR beacon. Situation digitization is thus attained, and the dynamics of a business situation is traced as objective, quantifiable data. This measurement can capture real-world situations concerning peoples' activities, which system-event logs fail to capture.

Function	Description
Motion	Captures 3-axis acceleration at 50 Hz.
Face-to-Face Interaction	Detects interaction between nodes within 2 m at an angle of 120 degrees via infrared sensor.
Sound	Measures frequency and energy of sound through a microphone. Content of voice is not captured.
Other sensing	Temperature/brightness.

Figure 2. Wearable sensor and its functions

3.2.2 Situational variables

The obtained situational data are converted into situational variables. One way of classifying situational variables is to use the categories proposed by Belk (1975). Although his work was actually conducted in relation to consumer behavior and the factors that influence shoppers to make purchases, his categories can be applied to other business situations. Belk listed five situational variables, namely, "physical surroundings," "social surroundings," "time," "task definition," and "antecedent states." In relation to a taxonomy related to contextual knowledge (Kofod-Petersen 2005), a basic activity is defined as follows: "personal," "task," "spatio-temporal," "environmental," and "social." Moreover, Han et al. (2008) divided context simply into "physical", "internal", and "social context". Physical context, made up of physical things, refers to real-world nearby people. Internal context is composed of abstract things inside people, such as feeling, thought, task, action, and interest, and is related to people. Social context means a user's social relationships. The situational variables shown above are categorised in Table 1. According to the table, the categorization of Han et al. (2008) (which does not include time) handles a situation comprehensively thanks to its simplicity.

Han et al. (2008)	Belk (1975)	Kofod-Petersen (2005)
(Not Available)	Time	Spacio-Temporal
Physical	Physical surroundings	
		Environmental
Social	Social surroundings	Social
Internal (feeling, thought, task, action, and interest)	Antecedent states	Personal
	Task definition	Task

Table 1. Categorization of situational variables from previous literature

As listed in Table 2, based on these categorizations, fundamental situational variables, which can be calculated from sensor data, are defined as “environment” and “location” from the “physical” category. “Motion” is defined in relation to the “Internal” category. As mentioned above, however, “motion” from the sensor data can be used to infer emotional states of people. Each variable is defined per unit time and calculated per person. Note that task definition and internal context are difficult to define objectively from sensor data.

Category	Variables	Calculation method using sensor data
Time	Start time	Start time of sensor measurement
	End time	End time of sensor measurement
Environment	Sound	Sound energy from microphone
	Temp.	Temperature sensor
Location	Stay	Detection of IR beacon ID
Social	F-to-F/solitary	Detection of other wearable sensor ID
Motion (behavior)	Walking	Walk/still judgment from accelerometer
	Walking distance	Summation of detected IR beacon’s physical distance
	Moving or still	Move/Still judgment from beacon ID
	Motion rhythm	Frequency from accelerometer
	Motion energy	Energy from accelerometer
	Activeness	Active judgment from motion rhythm

Table 2. Fundamental situational variables for each person

3.3 Stage 2: Data aggregation and alignment

On the basis of the fundamental situational variables obtained in stage 1, i.e., situation digitization, data is aggregated and aligned. Firstly, a timeframe of each person’s activity (“start time” and “end time” in Table 2) are retrieved. For example, a timeframe would be work start and finish times of company employees or shopping start and finish times of customers. Next, all fundamental situational variables, including other people’s, during the time-frame are obtained and aggregated. The data are then aligned in a way that social and motion variables are combined and sorted at each location. The result of the aggregation and alignment contains each person’s multi-dimensional situational valuables, which could affect his or her behavior. An example of the multi-dimensional situational valuables per person in the case of customers in a store is shown in Figure 3. Here, “activity time” means the time period covering a certain activity of a person, e.g. shopping from 9:00 to 9:30. L_1 to L_n show location areas, and L_{all} means all locations covering the sum of activities.

Person (customer)	Activity Time (for shopping)	Stay(sec.)					Other customer stay(sec.)					Employee stay(sec.)				
		Lall	L1	L2	···	Ln	Lall	L1	L2	···	Ln	Lall	L1	L2	···	Ln
A																
B																
C																

F-to-F total (sec.)					F-to-F w/ employee(sec.)					F-to-F + motion rhythm					F-to-F + motion energy				
Lall	L1	L2	···	Ln	Lall	L1	L2	···	Ln	Lall	L1	L2	···	Ln	Lall	L1	L2	···	Ln

Figure 3. An example of multi-dimensional situational valuables (only a portion of the valuables are shown).

3.4 Stage 3: Statistical analysis of variables' association with outcome

The aim of stage 3 is to derive situational valuables that significantly affect a business outcome. Firstly, the significance of the correlation between outcome variables and all the other situational variables and its correlation coefficient is calculated comprehensively. Secondly, regression analysis is applied to those variables satisfying the correlation significance. For simplicity, linear-regression analysis is used (Kenney 1962). Thus, the situational variables having significant relation with the outcome variable are selected. These variables are called a "rule set." The rules in the set with statistical significance are thus selected as candidates for performance improvement. The selected rules are sources of new knowledge that people cannot presume but is closely linked to business outcome.

In summary, the proposed sensor-based knowledge discovery leads to a data-driven rule set forming the basis of new knowledge. Since the rule set explains statistical significance in relation to improving a business outcome, it helps managers make practical decisions beyond their experience and intuition. The rule set thus supports managers in taking measures and discovering new knowledge from their experience.

4 DEMONSTRATION OF THE METHODOLOGY

4.1 Background and data collection

To demonstrate the effectiveness of the proposed sensor-based knowledge discovery, it was applied to a typical retail store. The target store was a mid-size home center (one of a home-center chain in Japan). According to interviews and observations, the primary task of an employee is stacking and shelving items, order placement, and servicing customers when needed. Each employee is pre-assigned a specific area in the store. Their task operation is not routine but rather discretionary in manner. The following are details of data collection:

- Data collection period: 20 days (10 days each for before and after implementation of measures)
- Employees: 17 people (regular staff and part-time staff)
- Customers: 608 people (before 304 and after 304)
- Collected data: location, F-to-F interaction, motion in the store (employees and customers)
- Business outcome: average sales per customer

Situational data such as customer-service activity and locations were collected by wearable sensors and IR beacons (as explained in Section 3.2). First, data was collected over 10 days from the employees working in the store and 304 samples of all customers who visited the store during the study period. Hypothetical measures that increase business outcome were then derived from the data. After one of the measures was implemented (namely, employees' presence at particular locations was increased, as explained in Section 4.3), data was collected again over 10 days from the employees working in the store and 304 samples from the customers in the same way as the previous data collection. For collection of sample data, customers were randomly chosen from the customers who visited the store during the study period. The number of the sample customers corresponds to approximately 2% of all the customers who visited the store during the experiment.

Since the sample customers are asked to cooperate in the experiment and wear the sensor node when they enter the store, this treatment may affect their behavior. It is therefore anticipated that the sample customers are motivated to purchase more than they would do in their normal shopping situation.

According to the experimental setup explained above, information such as where and when the customer stood still, received service, what was purchased, and where the sales employees were positioned at the time was recorded.

To evaluate business outcome, POS (point-of-sales) data during the data-collection period was gathered. The POS data contain the same information as a sales receipt, namely, commodity code, quantity of purchase, and purchase amount. Therefore, the combined analysis of receipt number and sensor ID makes it possible to analyze the purchase process. For example, relations between stopping in front of store shelves, servicing customers, and purchase results can be analyzed quantitatively.

4.2 Data analysis and results

In stage 1 of the sensor-based knowledge discovery, the fundamental situation variables (Table 2) for the employees and customers are obtained every 10 seconds during the opening hours of the store from the first 10 days of data. In stage 2, data for all customer samples is prepared for statistical analysis. Specifically, all of the fundamental situation variables corresponding to the period that a customer is in the store are aggregated and aligned in a way that social, behavioral, and motion variables are sorted according to each of 25 locations in the store. It is reasonable to align the data concerning the customers because the outcome variable (average sales per customer) is related to the customer's situation in the store. In the third stage, the situational valuables that are related to average sales per customer are statistically obtained as a "rule set." Table 3 lists the derived rule set concerning the employees' operations. The rules satisfying the criteria, that is, significance of correlation (p-value < 0.001) and correlation coefficient larger than 0.2, are selected from the table of all the statistical relations. A linear-regression slope is also calculated for each rule. In this study, to reform the business process, the rule set for the group of employees is considered. Note that, however, segregation of individual employees is out of the scope of this case study.

Regarding F-to-F interaction activity, F-to-F interaction between employees and customers significantly correlates with total average sales per customer ($r=0.39$, p-value<0.001). Regarding location, an employee's stay at specific locations, such as "Magnet 2-1", "Magnet 1-3", "Magnet 2-3", and "Home electronics," relates to average sales per customer. Here, "Magnet" means a point in a store that attracts customers' attention after appropriate products are arranged so as to promote sales. Among those locations, "Magnet 2-1" affects average sales per customer most, indicating that an employee staying at this specific location (called a "hot spot" hereafter) for one second relates to average sales per customer of 14.5 yen ($r=0.27$, p-value<0.001).

It was confirmed that, from the viewpoint of the managers of the store, it does not make sense that an employee's staying at the hot spot relates to average sales per customer. It was also confirmed that the discovered rule is hard to intuitively understand, even for experienced managers, and that there is no such common belief regarding in-store marketing theory.

Situation variable (employees operation)		Correlation coefficient r	Linear regression slope (yen/unit)	Unit
Interaction activity	F-to-F with customer	0.39***	4.6	sec.
	F-to-F+motion rhythm (sum)	0.28***	3.52	-
	F-to-F+motion energy	0.28***	1.01	-
	F-to-F+motion rhythm	0.27***	1.43	Hz
	F-to-F+motion energy (sum)	0.27***	99.9	-
	F-to-F with employee	0.27***	0.38	sec.
Solitary activity	Walk time	0.35***	0.0614	sec.
	Motion energy (sum)	0.32***	38.2	-
	Motion rhythm (sum)	0.32***	1.33	-
	Solitary time	0.32***	0.198	sec.
	Walk+motion energy (sum)	0.31***	12.3	-
	Walk count	0.31***	2.57	Step
	Walking speed	0.30***	0.252	1 m/s
	Still	0.29***	0.379	sec.
Stay at location	Magnet (2-1)	0.27***	14.5	sec.
	Magnet (1-3)	0.22***	10.4	sec.
	Magnet (2-3)	0.29***	4.52	sec.
	Home electronics	0.26***	3.08	sec.
	Paint	0.28***	2.03	sec.
	Garden	0.24***	1.67	sec.
	Magnet (Total)	0.27***	1.65	sec.
	Tools & electric materials	0.29***	1.35	sec.
	Back of service counter	0.27***	1.25	sec.
	Service counter	0.3***	1.09	sec.
	Casher	0.33***	0.501	sec.
	Whole store	0.3***	0.202	sec.
	Worker' supplies	0.21***	2.86	sec.
	DIY	0.2***	0.982	sec.

*Table 3. The rule set: the effect of situation variable on average sales per customer (n=304, ***p-value <0.001).*

4.3 Evaluation of knowledge discovery

Though the managers of the store were not convinced of the validity of the derived rule set (namely, an employee's stay at the hot spot (magnet 2-1) affects average sales per customer), they made a decision to try the rule set by implementing a certain measure; that is, employees' stay at the hot spot, which strongly influences the customer's purchase in the whole store, as much as possible to see if the measure effectively improves average sales per customer. This measure is intended to be implemented without increasing costs in a way that a regular employee, working three- to four-hour shifts in charge of the area including the hot spot, is encouraged to be at the hot spot even when doing regular tasks.

The results of 10 days of measurement after implementing the above-described measure are analyzed as follows. Table 4 lists the results of sales improvement; namely, it compares average purchase per customer before and after the measure was taken. The results are shown separately for the sample customers and for the total number of customers during the study. It is apparent that the sample customers tend to purchase higher-price items and contribute to total sales more than other customers. As anticipated, asking for the customers' cooperation in the survey and giving them wearable sensors worked positively in terms of increasing purchases. Nevertheless, the number of sample customers consists of only about 2% of all the customers. The effect of sample customers on total sales increase is therefore small enough (increase in average sales due to the non-sample customers is calculated as 14.7%).

As for the total number of customers, increasing an employee's presence at the hot spot by 1.7 times, i.e. 70% increase in "staying-time," increased average sales by 15%. It is therefore concluded from this result that the measure caused the increase in customers' sales without increasing labor cost.

Average Purchase/customer	Before→After (sample customers; n=608)	Before→After (all customers)
(a) Number of items	+14.6%	+13.2%
(b) Price of items	+11.1%	+1.6%
(c) Total amount of sales	+29.9%	+15.0%

Table 4. Result of implementing sales-improvement measure

The statistical analysis of situation variables was thus shown to be effective in discovering new operation rules. Nevertheless, to enable the managers to gain knowledge, it is still necessary to understand the cause-and-effect relationship concerning the hot spot in an explainable way. Visualization is one way to understand a complicated situation. Therefore, a before-and-after analysis in terms of the customers' staying time in locations in the store was performed, and the analysis results were visualized as shown in Figure 4.

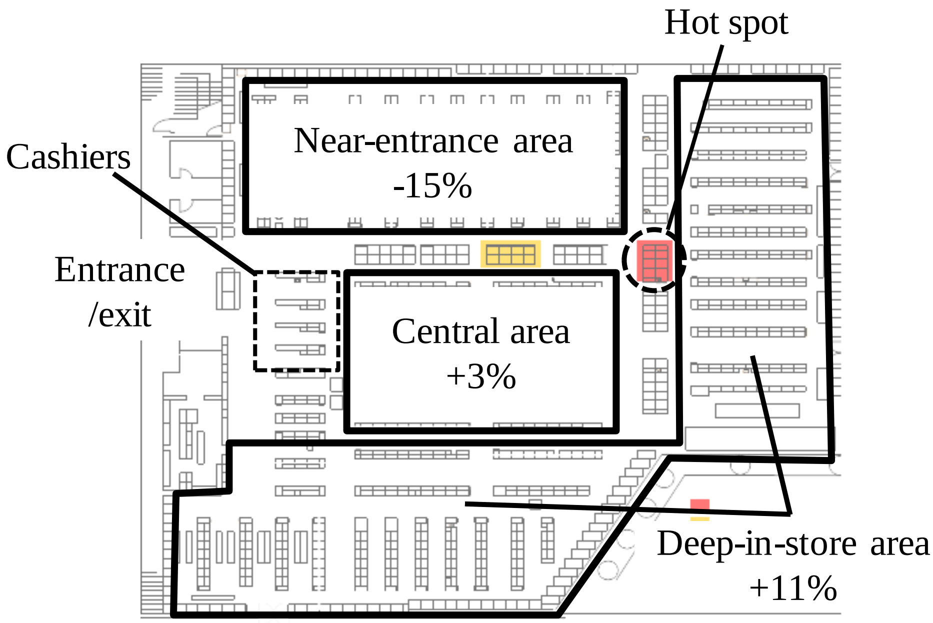

Figure 4. Location of hotspot (Magnet 2-1) and increase of customers staying time on the shop layout

The visualization shows that when employees stay at the hot spot for a longer time, the customers tend to go deeper into the store. The employees' presence at the hot spot thus seemingly functions as a "magnet." Given that result, namely, 15% increase in sales and visualization of customers' circulation change, the managers understood the effectiveness of the implemented measure. Moreover, it is necessary to consider that seasonal fluctuation of average customers' purchases is around 5% for this store. Since 15% increase in sales is sufficiently high and cannot be attained in regular routine activities, it is concluded that discovery of new knowledge improved the business outcome.

5 DISCUSSION

5.1 Implications

By using the proposed methodology, the managers in the store were able to discover new knowledge and improve business outcome. The hypothetical rule regarding employees' stay in the hotspot was therefore successfully interpreted as new knowledge. Since the rule set derived in stage 3 could provide both new and ordinary knowledge to the managers, the rules from the set have to be sorted out manually by managers in terms of their novelty, feasibility, and expected returns. As an example of ordinary knowledge, the variables regarding the employees' interaction with customers and the employees' activeness reasonably correlate to the business outcome, as shown in Table 3. However, even in the case of ordinary knowledge, the proposed methodology is effective in providing substantial reasons that some rules are more effective than others.

In the future, correlation analysis of variables can be performed more effectively by extracting new knowledge candidates automatically. Moreover, the variables concerning individual employees or the specific time of the day may yield other knowledge on the condition that enough samples are available. As for assuring statistical validity of the methodology, simple linear regression was employed in the present study. For more sophisticated knowledge discovery, however, multi-regression analysis or other known statistical techniques can be applied.

5.2 Limitation

The proposed methodology is effective in discovering new knowledge. Nevertheless, some limitations should be taken into account. Firstly, although the methodology proved to be practical and ready to be used for helping managers discover knowledge, it utilizes wearable sensors to obtain situation data unobtrusively. Privacy issues should therefore be taken care of, and the purpose of data usage should be clarified to diminish a wearer's privacy concerns. Secondly, the methodology requires explicit and measurable outcomes, such as sales and turnaround time, which are directly related to the situations. When causal relations are more complex, such as relations between R&D process and eventual profits, some process modelling needs to be integrated in the analysis. Thirdly, the period of validity concerning the discovered knowledge should be considered. Because the situational variables themselves change in time, different results might be obtained at similar settings in different situations. Similar types of stores should therefore be studied to gain generalization of the results.

5.3 Extension of knowledge-creation theory

Knowledge-creation theory is widely known for the spiral process of explicit knowledge and tacit knowledge (Nonaka 1994). Traditionally, in the socialization mode of the SECI process, people obtain tacit knowledge mainly through interaction with other people, such as observation, and training. Then, the tacit knowledge should be externalized using peoples capabilities such as metaphor and analogy.

As an extension of this process, a new process of knowledge co-creation between humans and computers can be realized by computational support which overcomes people's cognitive limitations (Figure 5). In this process, people's behavioral experiences associated with location, face-to-face interaction, motion, and physical environments are automatically digitized and accumulated in a database. The accumulated data is then statistically analysed to derive the particular behaviour that relates to the KPI indicated by people. Using these statistical relations, people make a decision to initiate or reinforce specific behaviours, suggested by a computational machine, to gain new experience.

In summary, the computer can play a leading role in knowledge creation and encourage people's activities to help them discover knowledge from new experience. In other words, as opposed to the traditional view that deep thought and hypothesis should precede behaviour, new behavioral experience triggers creative thinking. This extension to the conventional process explores a new paradigm of knowledge creation by human-machine cooperation.

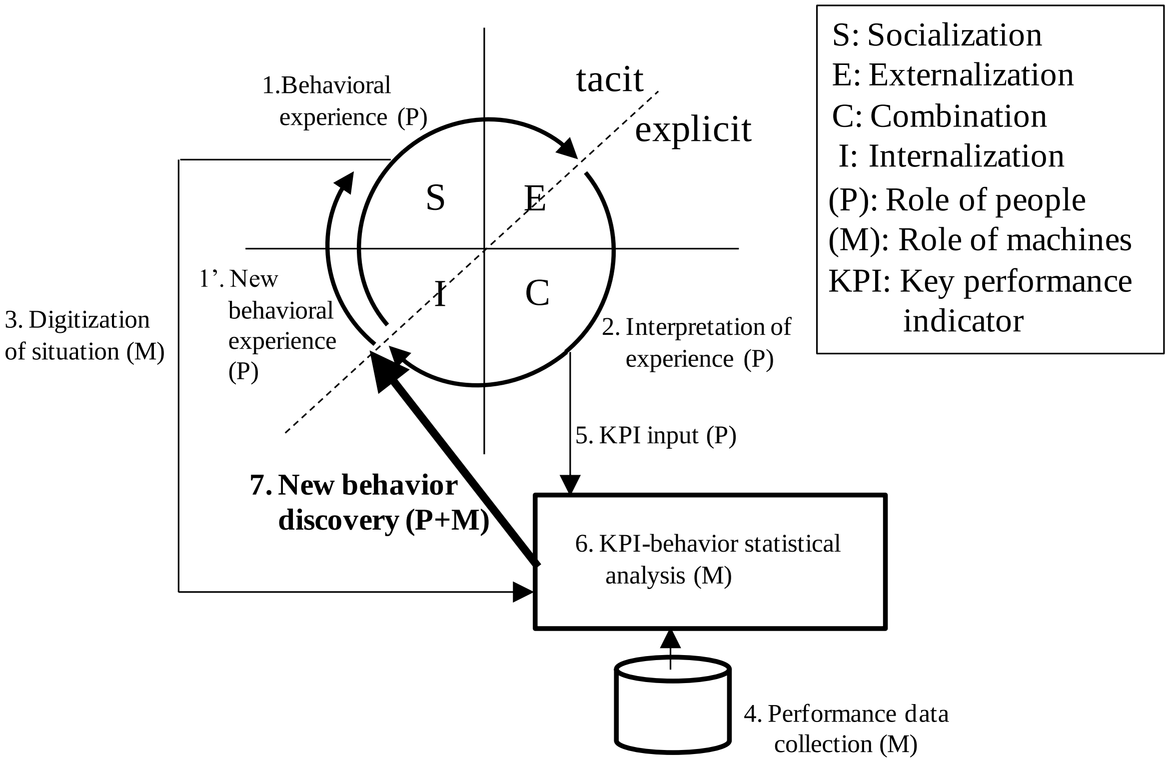

Figure 5. Knowledge-creation process supported by computing machine

6 CONCLUDING REMARKS

The proposed sensor-based knowledge discovery aims at discovering knowledge without relying on hypothesis or existing theory. The features of this methodology are as follows: Wearable sensors are utilized to unobtrusively capture people's location, motion, and social interaction with others. The captured data is converted into multi-dimensional situational variables and then statistically analyzed to derive a "rule set," which forms the basis of new knowledge related to business outcome.

A case study at a retail store derived a rule and demonstrated that increasing employees' staying time in a "hot-spot" area by 70% increased average sales per customer by 15%. The effectiveness of the rule might be momentary and is not assured for a long time period. Nevertheless, the result indicates that the methodology successfully discovered knowledge that can increase business outcome.

The contribution of the methodology is the practical discovery of knowledge without the need for particular hypothesis. The methodology identifies situation variables related to business outcome with statistical significance. It thus helps managers make proactive decisions with confidence rather than adopting a passive "wait-and-see" approach.

The methodology so far identified new knowledge at the group level. Nevertheless, it can also be applied at the individual level or specific spacio-temporal situations. As a future scenario for adding more value, a feedback system with automatic rule filtering, which does not rely on a manager's manual decision making, will enable delegation to lower employees.

To sum up, it was shown that rules discovered by computer, by using innovative sensor and computational technologies to capture real-world situations and conduct statistical analysis, can become practical human knowledge that leads to real business outcomes. This finding represents some of the first evidence that the methodology discovers practical knowledge by overcoming cognitive limitations of people.

Acknowledgements

We would like to thank all the members of the SocioInfo Project, led by Hitachi High-Technologies and Hitachi Central Research Laboratory, for their useful comments and technical support.

References

Adams, M.J., Tenney, Y.J. and Pew, R.W., (1995). Situation awareness and the cognitive management of complex systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37 (1), 85-104.

Ara, K., Sato, N., Tsuji, S., Wakisaka, Y., Ohkubo, N., Horry, Y., Moriwaki, N., Yano, K. and Hayakawa, M. (2009). Predicting flow state in daily work through continuous sensing of motion rhythm. In Sixth International Conference on Networked Sensing Systems (INSS), 1 – 6.

Bailey, T. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Second International Conference on Intelligent Systems for Molecular Biology, pp. 28-36, AAAI Press.

Bao, L. and Intille, S.S. (2004). Activity Recognition from User-Annotated Acceleration Data. In Pervasive (volume LNCS 3001), pp. 1-17, Springer-Verlag, Berlin Heidelberg.

Belk, R.W. (1975). Situational variables and consumer behavior. Journal of Consumer Research, 2(3), 157-164.

Boytsov, A. and Zaslavsky, A. (2012). Formal verification of context and situation models in pervasive computing. Pervasive and Mobile Computing, 9(1), 98–117.

Dahlbom, A., Niklasson, L., Falkman, G., and Loutfi, A. (2009). Towards Template-based Situation Recognition. In Intelligent Sensing, Situation Management, Impact Assessment, and Cyber-Sensing, Proc. SPIE, 7352, doi:10.1117/12.818715.

de Medeiros, A.A., Weijters, A. and Van Der Aalst, A. (2006). Genetic Process Mining: A Basic Approach and Its Challenges. In Business Process Management Workshops, pp.203-215, Springer.

Durso, F.T and Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments, human factors. The Journal of the Human Factors and Ergonomics Society, 50, 442-448.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems, human factors. The Journal of the Human Factors and Ergonomics Society, 37(1), 32-64.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17 (3), 37-54.

Grant, R.M. (1996). Toward a knowledge-based theory of the firm. Strategic Management Journal, 17(7), 109-122.

Han, L., Jyri, S., Ma, J., and Yu, K. (2008). Research on Context-aware Mobile Computing. In Proceedings of Advanced Information Networking Applications Workshops, pp 24-30, Beijing.

Hey, T., Tansler, S. and Tolle, K. (2009). The Fourth Paradigm: Data Intensive Scientific Discovery. Microsoft Research Publishing.

Joachims, T. (2002). Learning to Classify Text Using Support Vector Machines. PhD thesis, Dortmund University.

Kenney, J.F. and Keeping, E.S. (1962). Linear Regression and Correlation. Mathematics of Statistics. Third Edition. pp.252-285, Van Nostrand, Princeton, NJ.
Kearns, M.J. and Schapire, R.E. (1994). Efficient distribution-free learning of probabilistic concepts. Computational Learning Theory and Natural Learning Systems. Journal of Computer and System Sciences, 48(3), 464-497.
Kofod-Petersen, A. and Mikalsen, M. (2005). An Architecture Supporting Implementation of Context-Aware Services. Workshop on Context Awareness for Proactive Systems (CAPS 2005), pp.31-42, HIIT Publications, Helsinki, Finland.
Kofod-Petersen, A. and Cassens, J. (2006). Using activity theory to model context awareness. Modeling and retrieval of context. Computer Science, 3946, 1-17.
Kriegel, H., Borgwardt, K.M., Kröger, P., Pryakhin, A., Schubert, M. and Zimek, A. (2007). Future trends in data mining. Data Mining and Knowledge Discovery 15(1), 87-97.
Leont'ev, A.N. (1978). Activity, Consciousness, and Personality. Prentice-Hall.
Leyer, M. (2011). Towards a Context-Aware Analysis of Business Process Performance. In Proceedings of PACIS 2011, paper 108.
Lin, D. and Pantel, P. (2001). Dirt: Discovery of Inference Rules from Text. KDD 2001.
Liu, J., Yan, S., Li, Y. and Qi, J. (2008). The support model of situation awareness and business intelligence to virtual enterprise partner selection. Second International Symposium on Intelligent Information Technology Application, 2, 1025-1029.
Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. Organization Science, 5(1), 14-37.
Ozdemir, V., Smith, C., Bongiovanni, K., Cullen, D., Knoppers, B.M., Lowe, A., Robert, M.P., Stewart, E., Yee, G., Yu, Y-K. and Kolker, E. (2011). Policy and data-intensive scientific discovery in the beginning of the 21st century. OMICS: A Journal of Integrative Biology, 15(4), 221-225.
Pentland, A. (2008). Honest Signals: How They Shape Our World. MIT Press, Cambridge, MA.
Piatetsky-Shapiro, G. (2007). Data mining and knowledge discovery 1996 to 2005: Overcoming the hype and moving from "university" to "business" and "analytics". Data Mining and Knowledge Discovery 15(1), 99-105.
Ploesser, K., Peleg, M., Soffer, P., Rosemann, M., and Recker, J.C. (2009). Learning from context to improve business processes, BPTrends, 6(1), 1-7.
Ramos, E., Santoro, F. and Baiao, F. (2010). Process Improvement Based on External Knowledge Context, In Proceedings of the 21th Australasian Conference on Information Systems, paper 34, Brisbane.
Rosemann, M., Recker, J., and Flender, C. (2008). Contextualization of business processes. International Journal of Business Process Integration and Management, 3(1), 47-60.
Schonig, S., Gunther, C. and Jablonski, S. (2012). Process Discovery and Guidance Applications of Manually Generated Logs. In the Seventh International Conference on Internet Monitoring and Protection, pp.61-67, Stuttgart, Germany.
Suchman, L.A. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press, New York, USA.
Teece, D.J. (1998). Capturing value from knowledge assets: the new economy, markets for know-how, and intangible assets. California Management Review, 40(3), 55-78.
Tiwari, A. and Turner, C.J. (2008). A review of business process mining: state-of-the-art and future trends. Business Process Management Journal, 14(1), 5-22.
Turban, E., Sharda, R., Aroson, J.E. and King, D. (2008). Business Intelligence: A Managerial Approach. Pearson Prentice Hall, Upper Sadle River, NJ.
van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G. and Weijters, A.J.M.M. (2003). Workflow mining: A survey of issues and approaches. Data and Knowledge Engineering, 47(2), 237-267.
van der Aalst, W.M.P. and Song, M. (2004). Mining social networks: Uncovering interaction patterns in business processes. Business Process Management, 3080, 244-260.

van der Aalst, W.M.P., Weijters, A.J.M.M. and Maruster, L. (2004). Workflow mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering, 16(9), 1128-1142.

van der Aalst, W. (2011). Process mining: making knowledge discovery process centric. ACM SIGKDD Explorations Newsletter archive, 13(2), 45-49.

Vygotski, L.S. (1978). Mind in Society. Harvard University Press, Cambridge, MA.

Wakisaka, Y., Ara, K., Hayakawa, M., Horry, Y., Moriwaki, N., Ohkubo, N., Sato, N., Tsuji, S. and Yano, K. (2009). Beam-scan sensor node: Reliable sensing of human interactions in organization. In Proceedings of the Sixth International Conference on Networked Sensing Systems, pp.1-4, Pittsburgh.

Weijters, A., van der Aalst, W.M.P. and de Medeiros, A.K.A. (2006). Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP, 166.

Weijters, A and Ribeiro, J.T.S. (2011). Flexible heuristics miner (fhm). In IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2011, pp.310-317.

Wil, M.P., van der Aalst, W. and Dustdar, S. (2012). Process mining put into context. IEEE Internet Computing, 16, 82-86.

Ye, J., Dobson, S. and McKeever, S. (2012). Situation identification techniques in pervasive computing: A review Pervasive and Mobile Computing, 8(1), 36-66.