A Review of KDD-Data Mining Framework and Its Application in Logistics and Transportation

Fauziah Abdul Rahman[#1], Siti Mariyam Shamsuddin[*2], Shafaatunnur Hasan [#3] Norhaidah Abu Haris[*4]

[#1]*Faculty of Techncial Foundation, Universiti Kuala Lumpur- Malaysian of Institute of Industrial Technology (MITEC),81750, Johor, Malaysia*

[#23]*Faculty of Computing, Universiti Teknologi Malaysia(UTM) Skudai, 81310, Johor Bahru Johor, Malaysia*

[*4]*Faculty of Software Engineering, Universiti Kuala Lumpur Citi Campus- Malaysian of Institute of Information Technology (MIIT),1016, Jalan Sultan Ismail, 50250 Kuala Lumpur, Malaysia*

[1]fauziahar@unikl.edu.my

[2]mariyam@utm.my

[3] Shafaatunnur@utm.my

[3]norhaidah@unikl.edu.my

***Abstract*— In this paper, an understanding and a review of Knowledge Discovery Database Data Mining (DM) development and its applications in logistics and specifically transportation are highlighted. Even though data mining has been successful in becoming a major component of various business processes and applications, the benefits and real-world expectations are very important to consider. It is also surprising fact that very little is known to date about the usefulness of applying data mining in transport related research. From the literature, the frameworks for carrying out knowledge discovery and data mining have been revised over the years to meet the business expectations. The paper is concluded by proposing a framework for actionable knowledge discovery and data mining to be applicable in real life application such as within the context of transportation industry.**

***Keywords*—** Data Mining, Knowledge Discovery Database-Data Mining (KDD-DM), Domain Driven Data Mining (DDDM) Knowledge Discovery, Domain Driven Data Mining- Actionable Knowledge Discovery (AKD-KDD), Logistics, Fleet Maintenance.

1. INTRODUCTION

Logistics can be understood as a subset of supply chain management performance. Logistics can be defined as strategically managing the procurement, involve the movement of materials as well as storage of materials, parts and finished products inventory and related with the information flows, through the organization at a maximum profits with minimum costs in fulfilment of orders [3]. American Council Logistics Management also defined logistics as the process of planning, implementing and controlling the efficient, cost effective flow and storage of raw materials, in-process inventory, finished goods and related information from point of origin to point of consumption for the purpose of conforming to customer's requirements. Logistics became as a planning orientation and framework that seek to create a single plan for the flow of product and information through a business. Organization deals with supplier and customer to measures the quantity of material that passes through a given network per unit of time. Thus to achieve the logistics objectives, it builds upon the logistics framework to achieve the linkage within particular organization and with the processes of other organization. The incessant economic and industrial activities around the globe and the splurge of exports and imports continue to impose greater demands on shipping and cargo industry. Hence, the traditional transportation vendors not only strive to deliver cargo securely and accurately to customers on time but also consider reducing cost and flexibility dispatching vehicles as well as staff [2]. Thus, in order reducing costs at time related positioning resources, logistics scheduling problem has gained increasing importance with the development of supply chain management. Logistics scheduling has to deal with job delivery and transportation issues [4]. This includes minimize the sum of weight job delivery and the total transportation cost. The world wide cost to industry of outsourced logistics problems in 1995 have been highlighted in three (3) areas, maintenance of fleets, distribution and delivery stock. It is estimated to be AUS$ 900 Billion dollars of which the cost to Australian organizations is two billion.

International Journal of Supply Chain Management
IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

2. LOGISTICS AND FLEET MAINTENANCE PROBLEMS

Given the proliferation and complexity of some logistics problems, the application of computer systems, particularly decision support system is expected to increase significantly. However, current software tools for decision support in logistics do not totally address the combination of these characteristics. Software developed specifically for logistics problems are usually offer the reporting and tracking variety and or automate routine tasks but offer little support for decision making. Tools currently available for decision support are usually problem specific or too general to complement the decision process in logistics problems. There were four characteristics that have been identified as the most real world logistics problem. This includes large decision space; consists of a large of number of decision variables and possible options or strategies, availability of real time data; business modern that have extensive data collection capabilities that provides operational data which can be used for effective optimization and communication in real time which uncertainty problems in making decision because of the uncertain and incomplete knowledge about future circumstances, and numerous decision makers and complexity; dynamic behavior among the logistics components and interconnectivity. A general decision support system framework for logistics has been done and a mapping between research areas and logistics problem characteristics has been highlighted [10]. Three (3) outsourced logistics problems identified. It was fleet maintenance, distribution and delivery stock are related to uncertainty research area; thus proved that those areas are among the most relevant to logistics problem in a real business world.

3. PROBLEM'S BACKGROUND

Previously, there are many researches related to transportation involving vehicle routing, vehicle scheduling, fleet preventive maintenance related with time windows in job delivery and transportations using statistical method. Using statistical method, sometimes one can find patterns are not significant in reality. Data mining is a legitimate activity as long as one understands how to do it correctly. But very little is known to date about the usefulness of applying data mining in logistics and transport related research. Nowadays, the computer based systems are being used to automatically diagnose problems in vehicles in order to overcome some of the disadvantages associated with relying completely on experienced personnel. Typically, a computer based system utilizes a mapping between the observed symptoms of the failure and the equipment problems using techniques such as table look-ups, a symptom problem matrices and production rules. These techniques work well for simplified systems having simple mappings between the symptoms and problems. However, complex equipment and process diagnostics seldom have simple correspondences. In addition not all symptoms are necessarily present if problem has occurred, thus making other approaches more cumbersome [21]. These approaches either take a considerable amount of time before a failure are diagnosed or provides less than reliable results, or are unable to work well in complex systems. There is a need to be able to quickly and efficiently determine the cause of failures occurring in the vehicle maintenance system, while minimizing the need of human intervention [18].Having a direct access to systems data from remote vehicles would helpful in optimizing vehicle maintenance scheduling, route planning and minimize downtime from unexpected breakdown such as track vehicles with artificial intelligence but depending on it alone was costly [18]. The research also shown that, the existence fleet management can only analyze records after incident-occurrence and cannot analyze vehicle status in a real time. Even though the future system can be integrated with real time technology such as Global Positioning System (GPS), that can provide more valuable information, it will lead to data accumulation [6]. Thus in identifying imminent system failures or failure prognostics, better diagnostics data in the system is another way to help in enhancing the capability of maintainers at minimal cost where data mining is applied.

4. DATA MINING

4.1 Data Mining in Transportation

Data mining (DM) can be defined as the science of extracting useful information from large data sets or databases. With the help of data mining, derived knowledge, relationships and conclusions are often represents as models or pattern [18]. It is also can be defined as a spatial data mining that is useful in extracting useful information from huge amounts of data and is highly relevant to applications in which tremendous data volumes are involved, thus exceeding human analytical capabilities [6]. A recent review by Kohavi [11] stated that data mining serves two (2) goals, insight and prediction. Nowadays, data mining in various forms is becoming a major component of business operations. Almost every business process involves some form of data mining. In term of transportation, several researchers have been developing a unique approach to road traffic management and congestion control, monitoring drowsy drivers, road accident analysis, Pavement Management Data, Geographic Information Systems for Transportation data, GPS Data, Roadway Video logs, Spatial data and Road Roughness Data Analysis using data mining tools to identify these complex relationship between the data nature of logical, physical, real and virtual world [6]. However the data captured from the various information technologies are not fully utilized as deliverable information and knowledge. Additionally, using DM as a tool alone, failures in real business environment makes the analysis results not interpret as the whole picture of business perspectives. Recent critiques state that DM does not contribute to business in a large scale [7][17]. To meet this requirements, the DM process alone has been merged with Knowledge Discovery Database (KDD) has been revised over the years to meet the business expectations by supporting decision support making and action specifically in transportation reduction cost.

Table 1. The Existing Studies on Application of Data Mining in Transportation

Case studies	Framework and Data mining methods
1.Using Data Mining Techniques on Fleet Management System By: Chang-Yi Chen, Tien Yin Chou, Ching Yun Mu, Bing-Jean Lee, Magesh & Hsien Chao. Year:2003	Objectives: - To explore the useful data of vehicle behaviors that help to understand the status of vehicle or driver such as being out on duty, driving against traffic regulations and deviating from routes. - To alert to abnormal conditions releasing burden on fleet mgt. -Data Mining methods used: (1) Sequential Pattern Data Mining: To locate the regular routes based on "Checkpoint" & compared with current status & system identifies deviating routes. (2) Cluster Analysis Data mining: To detect vehicle halting and staying around some place. -Result: The table reveals the sequential characteristics wherein every record is maintained or regular time basis. Cluster analysis result is able to detect whether drivers involved in illegal matters. -Framework used: CRISP-KDD.
2. Using simulation, Data Mining, and Knowledge Discovery Techniques for Optimized Aircraft Engine Fleet Management By: Michael, Madhav and Gary L. Hogg. Year: 2006	-Objectives: To determine the near-and long term impacts of candidate aircraft engine maintenance decision, particularly in terms of Life-Cycle cost (LCC) estimation and operational availability. These will combines the approach of data mining, knowledge based techniques & simulation. The project then called as Cost Projection Simulator (CPS).In simulation-based cost projector, these parameters such as the domain expert knowledge, mission scenarios, fleet configuration & status, supply system posture, maintenance policies and individual engine reliability characteristics are updated to reflect imposed changes. -Data mining methods used: (1) Regression-Linear regression: understand variables influencing LCC. (2) Classification (MDA, CARD, ANN and Bayesian Network): used to analyze the parameters that influence LCC. Data classified as low or high cost engine based on their LCC. (3) Clustering (K-means): segment low-medium-high LCC engines and then study them to understand the variables or factors influenced their costs. -Results: 1. Replacing a new engine module within 100 hours of phase did not seem to influence the cost. 2. Repairing a module within 50 hours of phase seems to have effect on cost differences. -Framework used: DDDM
3. Effective Data Mining for a Transportation Information System By:P.Haluzo Year:2003	Objectives: -To indentify the reasons why accidents happened between tram and car in the electric tramway net of Prague Public Transit Company. -One accident increasing delays to other trams in the affected area. -Data mining methods: Association Rules Data Mining Results: If a high number of accidents occur in a given day and hour then a high percentage of trams delayed more than 180 seconds will appear with a probability of 89% in the same day and hour. Framework used: CRISP-KDD
4. Study on the Application of Knowledge	Objective: -To demonstrate the useful of KDD in finding out the potentially useful

Discovery in Databases to Decision Making of Railway Traffic Safety in China By: Cao Zhang, Yanchun Huang and Gang Zong Year:2010	information from mass database so that decision will be more accurate based on the problem statements. -To demonstrate how case-based reasoning can address accident treatment problems effectively. Mapping from structure information of the information to CBR. -Results:Decision tress is reasonable where the accuracy rate of classification is 80%. But no knowledge on classification decision tree is extracted. -Data mining methods: Decision trees classification. -Framework used: DDDM-AKD.
1. Utilizing Data Mining to Influence Maintenance Actions **2.** By: Thomas Young and et. Al(2010)	-Objective: 1. To improve maintenance practices by determine where and how maintenance procedures can be changed and enhanced for Aircraft Launch and Recovery Equipment (ALRE). -Results: 1. Corrective maintenance happens more often in environments that have either nighttime landing or more F-18 landing. This can help in making decision for example: Can invest more time doing preventative maintenance to avoid the necessity of corrective actions on carriers with higher occurrence of nighttime flights or f-18 flight. 2. Saving man hour due to reduction in overhead and increased operational time if the components can be replaced parallel. Te result not only to identify what types of maintenance should be done and with their frequency but also how to do maintenance (Actions). -Data mining methods: Apriori Algorithm and statistics. -Framework used: DDDM-AKD

4.2 Evolution of Data Mining Framework

4.1.1 CRISP-Knowledge Discovery Database

The whole process is sometimes called as knowledge discovery databases (KDD). This was the first generation of KDD where DM process attached together in the KDD life cycle to ensure a discover knowledge can meets the business requirements. Nowadays researchers with strong industrial engagement realized the need from DM to KDD to deliver useful knowledge for the business decision making. Traditionally, one standard, named CRISP-DM (Cross-Industry Standard Process for Data Mining) Methodology[8], determine the process step helps to avoid common mistakes[7][22]. It is important to understand each phases before implementing DM process.

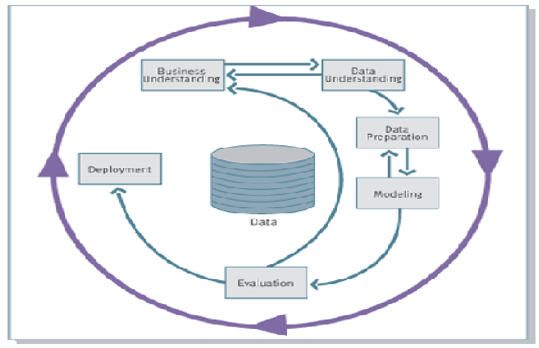

Figure 1. Cross-Industry Standard Process for Data Mining Methodology (CRISP-DM))

Refer to Figure 1 above the first phase is business understanding where to understand what is really to be accomplished. This task involves more detailed fact-finding about all the resources, assumptions and other factors that should consider in determining the data analysis goal. Second phase is data understanding that investigates a variety of descriptive data characteristics (count of entities in table, frequency of attribute value, average values and etc.). Third phase is data preparation which is the most difficult and time-consuming element in KDD process. The goal is to choose relevant data from available data, and to represent it in a form which is suitable for the analytical methods that are applied. Data preparation includes activities like data selection, filtering, transformation, creation, integration and formatting. The fourth phase is modeling which is the use of analytical methods (algorithms). There are many different methods and most suitable one must be chosen. This phase is also verifying the quality of the model such as testing in the independent data matrix, cross validation and others. The fifth phase is evaluation where the interpretation and evaluation of the discovered knowledge. In a decade, CRISP-DM life cycle representation of DM process seems to become more dominant [15]. However using this traditional framework represented some issues when the deployment stages are taken. The framework life cycle is sequential and linear. Even though the feedback loops are mentioned the sequential, natures of the representation suggest an ordering of the knowledge space and its exploration is not appropriately characterized the hierarchical and interactive network features of corporate knowledge space or can be as dynamic of DM [15]. CRISP-DM is a data centered-heavily depend on data itself [25] or data methodology or called as Data-oriented base framework. Current dominant situations are narrow focus and over emphasized by innovative data-driven and algorithm-driven research. In the real world scenarios, challenges always come from specific domain problems which back to the goal of DM towards business concerns, hence the objectives and goals of applying KDD are basically problem solving to satisfy real user needs.

4.2.2 Domain Driven Data Mining Knowledge Discovery (DDDM) Knowledge Discovery

In solving the problems that come from specific domains problem in a real world, next generation framework, Domain-Driven Data Mining (DDDM) has been developed specifically highlight the importance of data and domain intelligence [1]. Fundamentally, DDDM was including domain expert and domain knowledge as refer to the figure 2. Domain knowledge consists of the involvement of domain knowledge and experts. But usually in DDDM existing work often stops a pattern recovery which is mainly based on technical significance and interestingness which including objectives and subjective technical measures. Interestingness basically refers to the pattern of result or rules at the end of KDD and is unexpected or desired to expert and being useful or meaningful [1]. Therefore, it is important to have the involvement of domain knowledge in each phases of DDDM framework.

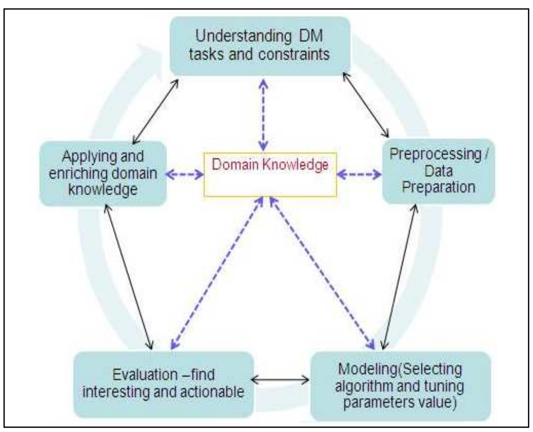

Figure 2. Data Mining Integrated with Domain Knowledge

However, different user may have different measures of interestingness pattern. Therefore interestingness is strongly depends on the application domain, expert knowledge and experience. Therefore, actionable pattern has been added instead of just interesting pattern. In business, actionable pattern is more important than interesting pattern. This is because actionable is refers to the mine rules or pattern that suggest valid and profitable actions to the decision makers [1][12][15][16][22]. This framework included two technical measures or metrics which are objective and subjective measures. The objectives measures are based on statistical strengths or properties of the discovered rules (data from database) and subjective measures are derived from the user’s belief or expectations of their particular problem domain [15]. In order to encode the domain knowledge manual method, currently, semi-automatic and automatic methods have been used. The automatic method requires some knowledge discovery tools such as Ontology Learning, Knowledge Acquisition based on Ontology and Semantic Web [13]. However these popular methods provide a conceptual or mapping representation of the application domain mainly elicited by analyzing the existing operational databases. Hence the interesting patterns and actionable patterns are still based on the technical interesting pattern which refer to data and user’s belief (domain knowledge) in particular domain. It shows that although this framework highlighted the involving of domain knowledge, the business concerns are not considered in assessing patterns. There are often many patterns mined but they are not informative and transparent to business people who did not know which are truly interesting and operable for their business. Furthermore business people often do not know and also did not informed, how to interpret them and what straight forward actions can be taken on them to support business decision making and operation. Therefore the studies on DDDM have been extended to effective and practical methodologies for Actionable Knowledge Discovery (AKD).

4.2.3 Domain Driven Data Mining- Actionable Knowledge Discovery (DDDM-AKD Discovery)

AKD framework was based on DDDM framework as refer to figure 3[23]. This framework targets knowledge that can be delivered in the form of business friendly and decision making actions, and can be taken over the business people seamlessly. Fundamental of AKD is therefore necessary to cater critical elements such as environment, expert knowledge and operability. To this end, AKD must cater for domain knowledge, environmental factors, balance technical and business expectations from both objectives and subjective perspectives (technical interestingness and business interestingness concerns) and support automatically converting patterns into deliverable business which are friendly and operable forms. This framework involves four major stages, constraints analysis, post-process and in-depth mining phases. Domain knowledge is involved into a system as constraint format which are data constraints, domain business process and business rules (domain constraints), interest or gap between academia and business (interestingness constraints) and deployment constraints where refer to interesting pattern must be able to integrate with domain environment for instance business rules, process, information flow and etc. [15]. Business interesting may refer to specific social and economic measures in term of problem domain. For instance profit, return and return on investment are usually used [15].Post-process deals with expert manual and required to re-mining actionable knowledge after a lots of pattern are mined. In-depth mining phase will make sure DM process as a humanmachined cooperated, loop-closed iterative back until obtain satisfied and actionable knowledge [23]. AKD is important because of multiple requirements expectations. However there are some issues need to be considered on.

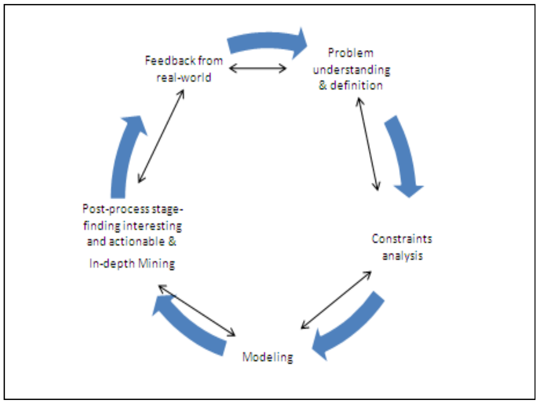

Figure 3.Research on Domain Driven- Actionable Knowledge

4.2.4 Proposed Framework

Based on AKD recently framework, there are few sub-process that we enhance and more detail about the proposed framework DDDM-Knowledge AKD. In this framework, Business understanding (BU) and Domain Knowledge (DK) play significant roles in real-world data mining.

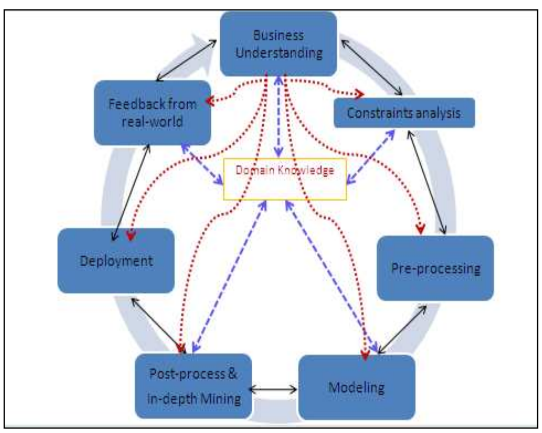

Figure 4. DDDM-Knowledge AKD

Business understanding. This phase is more important than other phases given that a number of decisions about other phases be made during BU phase. It is consists of four sub-processes as Table 2 below.

Table 2. Business Understanding Phases

Tasks	Outputs
Determination of business objectives	Background information, business objectives and business success criteria.
Assessment of situation	Inventory of resource, requirements, assumptions and constraints, risks and contingencies, terminology, and costs and benefits
Determine data mining goals	Data mining goals and data mining success criteria
Produce project plan	Project plan, initial assessment of tools and techniques.

Constraint Analysis. It is involve technical, economic, and social aspects in the process of developing and deploying actionable knowledge. There are data constraints, domain constraints, interestingness constraints and deployment constraints. All these constraints will be consider and concludes as hypothesis.

Pre-processing. It is based on human (domain knowledge) is in the circle of data mining process. This phase involve the collection of various of potential sources that can be integrated such as meta-knowledge from expertise in DM and other related field, meta-knowledge from DM practitioner, meta-knowledge from laboratory data experiments and meta-knowledge from field experiments in the real-world. Some related research presented few methods of represented these sources using Ontology, semantic, domain model and case based reasoning (CBR).

Modeling. Discover the interesting and actionable pattern using DM algorithm, learning the pattern results and using techniques of manual, semi-automatic or automating pattern recognition.

Post-process and in-depth mining pattern. Re-mining actionable knowledge after a lot of patterns are mined and refinement the process until obtains satisfies and actionable knowledge [23].

Deployment. The actionable patterns should be implementing in the real-world as based in BU and DK.

Feedback from the real world. There should revised and make sure the data being updated and correct for certain period of times.

5. CONCLUSION AND FUTURE WORK

In the next stage of research, the researchers will test and explore the afore-mentioned proposed DDDM-Knowledge AKD framework for determining effective maintenance strategy of vehicle fleet within a Malaysian logistics company. This company transport palm oil and the related products throughout the country. While there are

many issues in AKD that need to be considered in this area of application, but not much related works on the techniques to balance and combine all types of interestingness metrics specifically business interestingness components. There were studies that shows the AKD applications in areas such as stocks market, customer relationship management, supplier selections, crime identification, blog specific, search mining, social security network, telecommunication mining, financial mining and government service mining[14][23][16]. But very little is known to date about the usefulness of applying actionable data mining in transport related research. In addition, there are still unclear definitions or identifying elements of business interesting since it will depends on the domain. The idea of m-space with intelligence meta-synthesis facilities need to be explore as well. These include Domain intelligence, Data intelligence, Human Intelligence and Social intelligence [16][17]. Additionally, studies should be done by merging some proposed method of knowledge representation instead of ontology and semantics to transfer pattern to business rules. This is highly recommended to increase easy understanding by the end-users. Furthermore, there is a need to deal with possible conflict and uncertainty among respective interestingness elements especially business interestingness concerns.

ACKNOWLEDGMENTS

The researchers would like to express many thanks specifically to Universiti Kuala Lumpur (UniKL) for the financial support and a study leave granted to the first author to undertake a PhD research at Universiti Teknologi Malaysia (UTM).

REFERENCES

[1] Anqiang Huang, Lingling Zhang, Zhengxiang Zhu and Yong Shi, *"Data Mining Integrated with Domain Knowledge", in Cutting-Edge Research Topics on Multiple Criteria Decision Making*, Communications in Computer and Information Science, vol. 35, Springer Berlin Heidelberg, 2009, pp.184.

[2] A.Nesa thurai, "*Key Players in the Logistics Chain*", unpublished.

[3] Alan Harrison and Remko Van Hoek,"*Logistics Management and Strategy"*, International Logistic: A Supply Chain Approach Financial Times Prentice Hall, 2001 3- 45, 2002.

[4] Adolfo Crespo Marquez, *"Applications and Case studies, Modeling Critical failures maintenance: a case study for mining"*, Journal of Quality in Maintenance Engineering, Vol. 11 Issue 4, pp.301 – 317, 2005.

[5] Chang-Yi Chen, Tien-Yin Chou, Ching-Yun Mu, Bing-Jean Lee, Magesh and Hsien Chao, *"Using Data Mining Techniques on Fleet Management System"*, Esri International User Conference.

[6] Dan Luo, Longbing Cao, Chao Luo, Chengqi Zhang and Weiyuan Wang, *"Towards Business Interestingness in Actionable Knowledge Discovery"*, Proceeding of the 2008 conference on Applications of Data Mining in E-Business and Finance in press, 2008.

[7] Daniel T. Larose, *Discovering Knowledge Data, 2nd ed.* Publisher John Wiley& Sons, pp. 4-7, 2005.

[8] Ding Pan," *A Formal framework for Data Mining Process Model*". IEEE PACIIA 2009 in press, 2009.

[9] Ian Davidson and Ryszard Kowatczyk, *"Towards Better Approaches to Decision Support in Logistics Problem"*, unpublished.

[10] Kohavi R., "*Data Mining and Visualization*", In: Sixth Annual Symposium on Frontiers of Engineering, National Academy Press, pp. 31-40, 2001.

[11] Longbing Cao, Dan Luo and Chengqi Zhang, *"Knowledge action ability: satisfying technical and business sInterestingness."IEEE* Trans. International Journal Business Intelligence and Data Mining, Vol.2, No. 4, 2007.

[12] Longbing Cao, *"Domain Driven Data Mining: Challenges and Prospects."*Journal on Knowledge and Data Engineering. Vol. 1, No. 1 January 2008.

[13] Longbing Cao, *"Introduction to Domain Driven Data Mining", Data Mining For Business Application*, 1st ed., Springer Science and Business Media, LLC, 2009 pp. 3-10.

[14] Longbing Cao and Chengqi Zhang, "*The Evolution of KDD: Towards Domain-Driven Data Mining"*, International Journal of Pattern Recognition and Artificial Intelligence. Vol. 21, No. 4, pp. 677-692, 2007.

[15] Lily Sun, Cleopa John Mushi, *"Case-Based Analysis in User Requirements Modeling for Knowledge Construction"*, Journal of Information and Software Technology, Vol. 52, Issue 7, July 2010.

[16] P.Haluzova, "*Effective Data Mining for a Transportation Information Systems,*" Czech Technical University Publishing, in press.

[17] Sudhir Kumar Barai, "*Data Mining Applications in Transportation Engineering*", Transport, Vol. XVIII, No. 5, pp. 216-223, 2003.

[18] WenQun Wang, Haibo and Magaret, "*Vehicle Breakdown Duration Modeling.*" Journal of Transportation and Statistics. Vol. 8, Number 1, 2005.

[19] William R. King, Peter V. Marks, Jr., and Scott McCoy, "*The Most Important Issues in Knowledge Management.*" Communications of the ACM, Vol. 45, No 49, September 2002.

[20] Zhengxiang Zhu, Jifa Gu, Lingling Zhang, Wuqi Song and Rui Gao, "*Research on Domain-Driven Actionable Knowledge Discovery*", Communications in Computer and Information Science, Vol. 35, Part 5, 176-183, 2009.

[21] Sumana Sharma and Kweku-Muata Osei-Bryson, "*Framework for Formal Implementation of the Business Understanding phase of Data Mining Projects*, "Journal Expert Systems with Applications: An International Journal, Vol. 36, Issue 2, March 2009.