Volume 3, Issue 9, September 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Performance Evaluation of Actionable Knowledge Discovery (AKD) Framework under the Decision Making System

Mrs.Amruta Y. More (Kapse), Mr.A. Balachandran
Computer Engg., Pune University
India

Abstract— Recently there are many algorithms and tools presented for data mining. Many of such methods stop at mining as well as delivery of patterns those are satisfying expected technical interestingness. There are frequently many patterns are minded, however failed to satisfying the requirements of business peoples either of the fact that they are interested in using them or do not know how to operate them in order to take proper business decisions. And hence this resulted into the serious effects over employment of improved techniques of data mining those are used to produce the better productivity as well as quality at work. Thus in this paper we are presenting the framework for decision making system which is consisting of different methods to data mining at enterprise level as well as presenting its performance evaluation under the decision making system. Actionable knowledge discovery (AKD) is nothing but the closed optimization problem solving method. The processes involved in AKD are widely used for business needs. In this paper we are presenting four AKD frameworks in order to support the processes of AKD such as Postanalysis-based AKD, Unified-Interestingness-based AKD, Combined-Mining-based AKD, and Multisource Combined-Mining-based AKD (MSCM-AKD). We have implemented this algorithm and investigated their performances under the decision making datasets.

Keywords— AKD, Greedy BSP, action extraction, Data mining, domain-driven data mining (D3M), actionable knowledge discovery, decision making, CRM

I. INTRODUCTION

Like most data mining algorithms nowadays, a typical drawback in current applications of data mining in intelligent CRM is that individuals tend to focus on, and be happy with, building up the models and interpreting them, however to not use them to urge profit expressly. a lot of specifically, most data processing algorithms (predictive or supervised learning algorithms) solely aim at constructing client profiles, that predict the characteristics of consumers of bound categories. Samples of these categories are [1]: What reasonably customers (described by their attributes like age, income, etc.) area unit doubtless attritors (who can attend competitors), and what kind are loyal customers? This knowledge is beneficial however it doesn't directly profit the enterprise [2]. To boost client relationship, the enterprise should recognize what actions to require varying customers from an unwanted standing (such as attritors) to a desired one (such as loyal customers). This will be worn out the telecommunications business, as an example, by reducing the monthly rates or increasing the service level for a valuable client.

Unlike spacing information, to think about actionable information one should take into consideration resource constraints. Actions, like direct mailing and commercial, price cash to the enterprise. At identical time, enterprises are progressively strained by cut [3]. There's therefore a powerful limitation on the quantity of client segments that the corporate will battle, or within the variety of actions the corporate will exploit. To form a choice, one should take into consideration the price furthermore because the good thing about actions to the enterprise [4]. However, for every client, there is also an oversized variety of attainable actions or action sets which will be applied to the client.

There are many researchers working over the area of data mining [5]. The general problem in doing the research over the KDD is nothing but the dominating focus over algorithm innovation as well as avoiding the capability of decision making under the real life settings. And hence this resulted into the major problem for data mining applications such as workability of deployed algorithms, tools, as well as resulting deliverables [6-9]. In order to overcome such conditions, and empower the workable capability and advanced data mining performance in real-world production and economy, there is an urgent need to develop next-generation data mining methodologies as well as techniques that target the paradigm shift from data-centered hidden pattern mining to domain-driven actionable knowledge delivery.

In [1], we have studied the frameworks for actionable knowledge Discovery (AKD) in which we have studied the four types of AKD frameworks presented by author. These frameworks are presented for actionable knowledge discovery as well as decision making in system. In the paper [1], to support the AKD process, four types of AKD frameworks such as Postanalysis-based AKD, Unified-Interestingness-based AKD, Combined-Mining-based AKD, and Multisource Combined-Mining-based AKD (MSCM-AKD). However their real time analysis has not yet been evaluated graphically to make strong decisions in systems like CRM [2]. Hence in this paper we are presenting the approach in which we will combine above AKD frameworks and decision making tightly by formulating the decision making problems directly on top of the data mining results in a postprocessing step [2]. Empirical tests are conducted on both a realistic insurance application domain and UCI benchmark data. In next sections first we will define this research paper objectives and outcomes, after that we will define the four investigate algorithms in details, after that practical implementation and results analysis of four algorithms of AKD framework under decision making system [11].

II. RESEARCH BACKGROUND

The main objective of this research is to evaluate the performance of AKD frameworks under the realistic decision making systems such as CRM by using the benchmark datasets. Apart from this following the major objectives of this research study:

- To present the study over the complete AKD process and frameworks.
- To present the study over AKD frameworks formations and types.
- To present the efficient technique to actionable knowledge from decision trees of decision making system.
- To implement and evaluate the performance in terms of run time vs. number of action sets, net profit vs. number of action sets, etc.

Research Outcomes

1. Efficient method of combining the concepts of AKD frameworks and decision making system together under the realistic environment.
2. Performance of such system using the realistic datasets for varying number of action sets.
3. Our approach suggests the most efficient solution for intelligent systems like CRM, ERP etc.
4. Detailed process of AKD and Decision Making system used in CRM.

III. INVESTIGATED AKD FRAMEWORK

In this section of paper, we will discuss the four algorithms or frameworks of AKD which we are analyzing during this paper for decision making system.

3.1 Postanalysis-Based AKD: PA-AKD

This framework is having two steps such as pattern extraction as well as refinement. First, generally interesting patterns (which we call "general patterns") are mined from data sets in terms of technical interestingness $(t_o(), t_s())$ associated with the algorithms used. Further, the mined general patterns are pruned, distilled, and summarized into operable business rules (embedding actions) (which we call "deliverables") in terms of domain-specific business interestingness $(b_o(), b_s())$ and involving domain (Ω_d) and Meta (Ω_m) knowledge. Following figure 1 is and algorithm explaining the working of this framework.

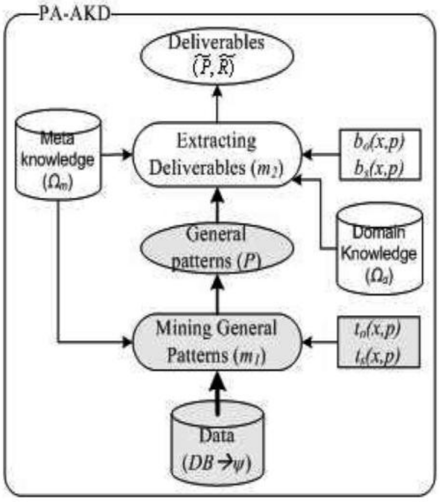

Figure 1: Postanalysis-based AKD (PA-AKD) approach

FRAMEWORK 1*:* Post Analysis-based AKD (PA-AKD)

INPUT: target data set DB, business problem Ψ*, and Thresholds* $(t_{o,0}, t_{s,0}, b_{o,0}$ and $b_{s,0})$ *OUTPUT: actionable patterns* $\tilde{P}$ *and operable business rules* $\tilde{R}$

Step 1: Extracting general patterns P;

FOR n =1 to N

Develop modeling method m_n

Interestingness $t_i()$ (i.e., $t_o(), t_b()$);

Employ method m_n *on DB and environment e;*

Extract the general pattern set P^{m_n};

END FOR

Step 2: Extracting actionable patterns $\widetilde{P}$;

$P = P^{m_1} U \ldots U P^{m_N}$

FOR j = 1 to $(count(P))$

Design post-analysis method m_2 *by involving domain Knowledge* Ω_d *and business interestingness* $b_i()$;

Employ the method m' *on the pattern set P as well as data set DB if necessary;*

Extract the actionable pattern set e P;

END FOR

Step 3: Converting pattern $\widetilde{P}$ *to business rules* $\widetilde{R}$.

3.2 Unified-Interestingness-Based AKD: UI-AKD

Fig. 2 shows the framework of UI-AKD. It looks just the same as the normal data mining process except for three inherent characteristics. One is the interestingness system, which combines technical interestingness $(t_i())$ with business expectations $(b_i())$ into a unified AKD interestingness system $(i())$. This unified interestingness system is then used to extract truly interesting patterns. The second is that domain knowledge (Ω_d) and environment (e) must be considered in the data mining process. Finally, the outputs are $\widetilde{P}$ and $\widetilde{R}$.

Ideally, UI-AKD can be expressed as follows:

$$UI-AKD : DB \xrightarrow{e, i(), m, \Omega_d, \Omega_m} \widetilde{P}, \widetilde{R}. \quad (10)$$

Based on the AKD formulas addressed before, $i()$ can be further expressed as follows:

$$i() = Int() = I(t_i(), b_i()). \quad (11)$$

Very often $t_i()$ and $b_i()$ are not dependent, thus

$$i() \rightarrow \eta \hat{t_i}() + \varpi \hat{b_i}(). \quad (12)$$

Weights η and ϖ reflect the interestingness balance tradeoff negotiated between data analysts and domain experts in terms of business problem, data, environment, and deliverable expectation. In some cases, both weights and aggregation can be fuzzy. In other cases, the aggregation may happen in a step-by-step manner. For each step, weights may be differentiated.

Patterns with $i()$ beating given thresholds (again, this must be mutually determined by stakeholders) come into the actionable pattern list. The pseudocode describing the UI-AKD process is as follows:

FRAMEWORK 2: Unified Interestingness-based AKD (UI-AKD)

INPUT: target data set DB, business problem Ψ and thresholds $(t_{O,0}, t_{s,0}, b_{O,0} \text{ and } b_{s,0})$

OUTPUT: actionable patterns e $\widetilde{P}$ and business rules e $\widetilde{R}$

Step 1: Extracting general patterns P;

FOR n =1 to N

Design data mining method m_n by involving domain knowledge Ω_d and considering Environment against unified interestingness $i()$;

Employ the method m_n on DB given e and Ω_d

Generate pattern set P;

ENDFOR

Step 2: Extracting deliverables $\widetilde{P}$;

Step 3: Converting e P to business rules $\widetilde{R}$.

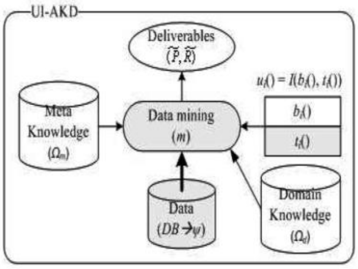

Figure 2: Unified-interestingness-based AKD approach

3.3 Combined-Mining-Based AKD: CM-AKD

For many complex enterprise applications, one-scan mining seems unworkable for many reasons. To this end, we propose the Combined-Mining [18]-based AKD framework to progressively extract actionable knowledge. Fig. 3 illustrates the CM-AKD.

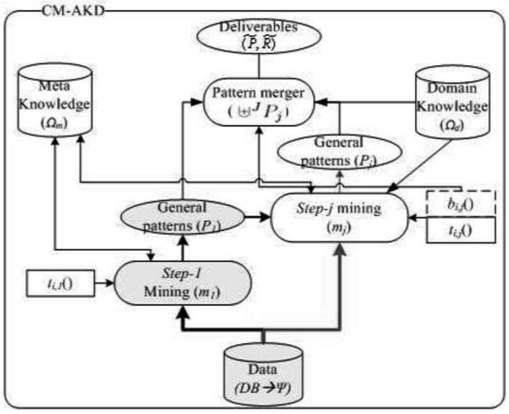

Figure 3: Combined-mining-based AKD (CM-ADK).

The CM-AKD process can be expressed as follows:

FRAMEWORK 3: Combined Mining-based AKD (CM-AKD)

INPUT: target data set DB, business problem Ψ*, and Thresholds* $(t_{o,0}, t_{s,0}, b_{o,0}$ and $b_{s,0})$

OUTPUT: actionable patterns $\widetilde{P}$ *and operable business rules* $\widetilde{R}$

Step 1: AKD is split into J steps of mining;

Step 2: Step-j mining: Extracting general patterns Pj

(j = 1; . . . ; J);

FOR j = 1 to J

Develop modeling method m_j *with technical*

Interestingness $t_{i,j}()$ (i.e., $t_o(), t_b())$ *or unified* $i_{i,j}()$

Employ method m_j *on the environment e and data*

DB engaging meta-knowledge Ω_m*;*

Extract the general pattern set Pj;

ENDFOR

Step 3: Pattern merging: Extracting actionable patterns e P;

FOR j = 1 to J

Design the pattern merger functions $\uplus^J P_i$ *by involving domain* (Ω_d)*and meta* (Ω_m) *knowledge, and business interestingness* $b_{i,j}()$*;*

Employ the method $\uplus^J P_i$ *on the pattern set Pj;*

Extract the actionable pattern set $\widetilde{P}$*;*

ENDFOR

Step 4: Converting patterns $\widetilde{P}$ *to rules* $\widetilde{R}$.

3.4 Multisource + Combined-Mining-Based AKD: MSCM-AKD

Another common situation is that the data volume is so large that it is too costly to scan the whole data set. Mining such complex and large volumes of data challenges existing data mining approaches. To this end, the approach proposes a Multisource + combined-mining-based AKD framework. Fig. 4 shows the idea of MSCM-AKD.

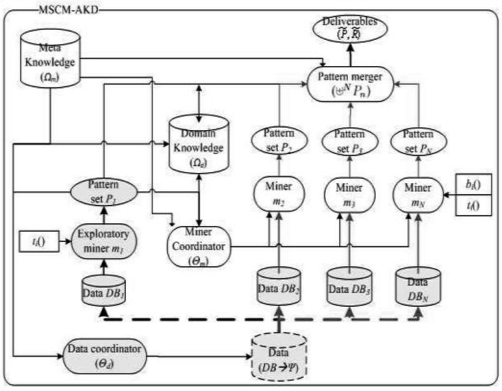

Figure 4: Multisource + combined-mining-based AKD

The MSCM-AKD process is expressed as follows:

FRAMEWORK 4: Multi-Source + Combined Mining Based AKD (MSCM-AKD)

INPUT: target data set DB, business problem Ψ, and Thresholds $(t_{o,0}, t_{s,0}, b_{o,0} \text{ and } b_{s,0})$

OUTPUT: actionable patterns $\widetilde{P}$ and operable business rules $\widetilde{R}$

Step 1: Identify or partition whole source data into N data Sets DB_n $(n = 1, \ldots, N)$;

Step 2: Data Set-n mining: Extracting general patterns P_n *on data set/subset* DB_n;

FOR $l = n$ to (N)

Develop modeling method m_n *with technical interestingness* $t_{i,n}()$ (i.e., $t_o(), t_b()$) *or unified* $i_{i,n}()$

Employ method m_n *on the environment e and Data* DB_n *engaging meta-knowledge* Ω_m;

Extract the general pattern set P_n;

ENDFOR

Step 3: Pattern merger: Extracting actionable patterns $\widetilde{P}$;

FOR $l = n$ to N

Design the pattern merger functions $\uplus^N P_n$ *to merge all patterns into* $\widetilde{P}$ *by involving domain and Meta knowledge* Ω_d *and* Ω_m, *and business interestingness* $b_i()$;

Employ the method $\uplus P_n$ *on the pattern set* P_n;

Extract the actionable pattern set $\widetilde{P}$.

ENDFOR

Step 4: Converting patterns $\widetilde{P}$ *to business rules* $\widetilde{R}$

IV. EXPERIMENTAL ANALYSIS

Experimental analysis is main aim of this paper. In this paper we have presented four recently presented AKD frameworks. However still to the date their experimental evaluation was not carried over real time datasets. We have used the real data set from various sources to test the performances of all this four algorithms of AKD. We have implemented one AKD framework in which four algorithms included to test their performances individually. Following figure 5 is showing the main screen of our implemented framework. This practical implementation is done in JAVA. Figure 6 showing the execution of first framework of AKD by giving input of tennis dataset of decision making. In the same ways we have executed different AKD frameworks with different decision making datasets which was our main aim of this project. After evaluation of all this we have get results for each framework as noted in figures 7 to 10. In this figures we are showing the run time performance vs. number of actions sets for each algorithm.

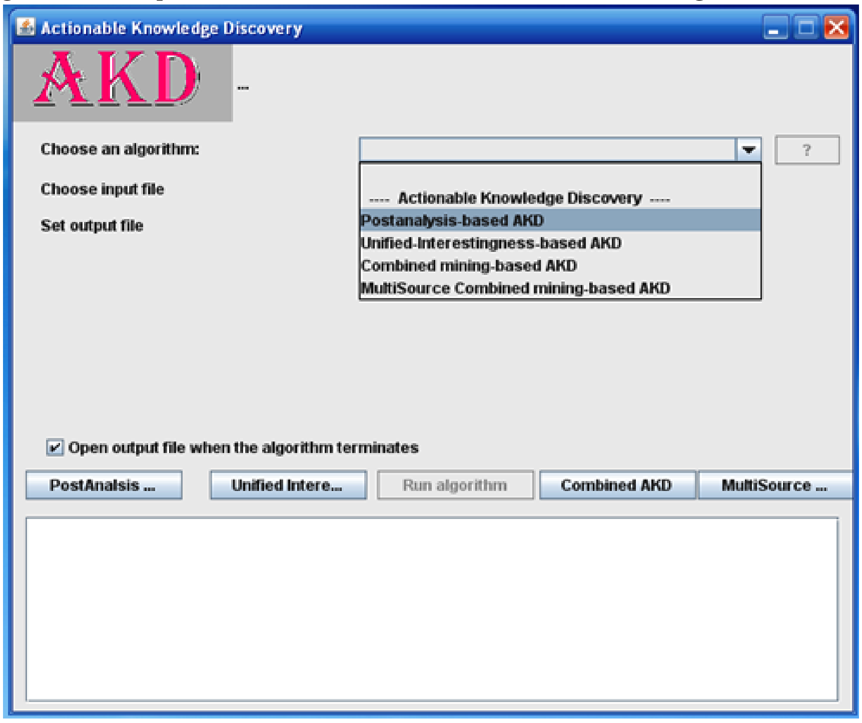

Figure 5: Main Screen of AKD Framework

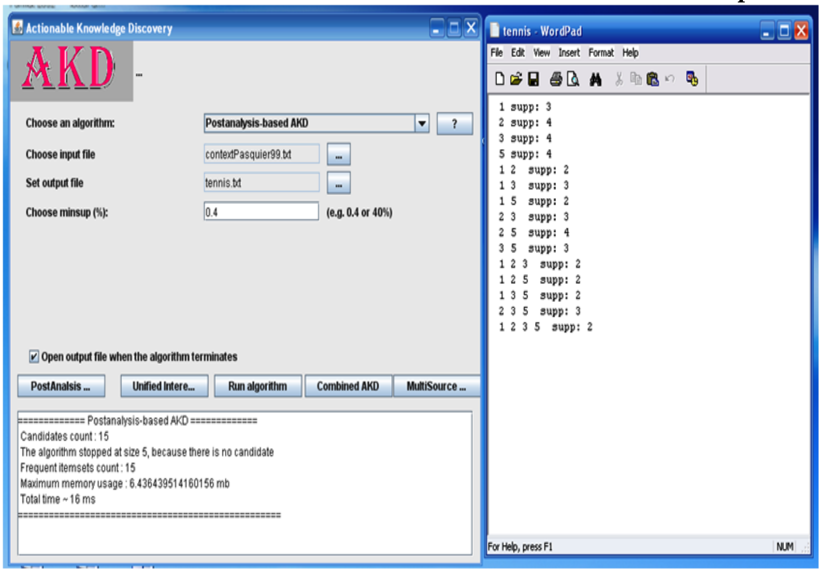

Figure 6: Outputs of PostAnalysis Based AKD over decision making system

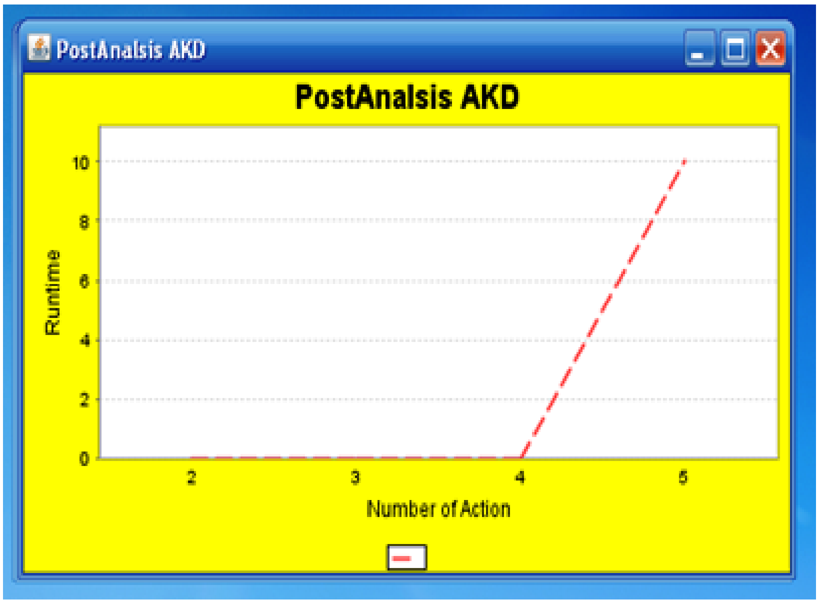

Figure 7: PostAnalysis AKD Performance

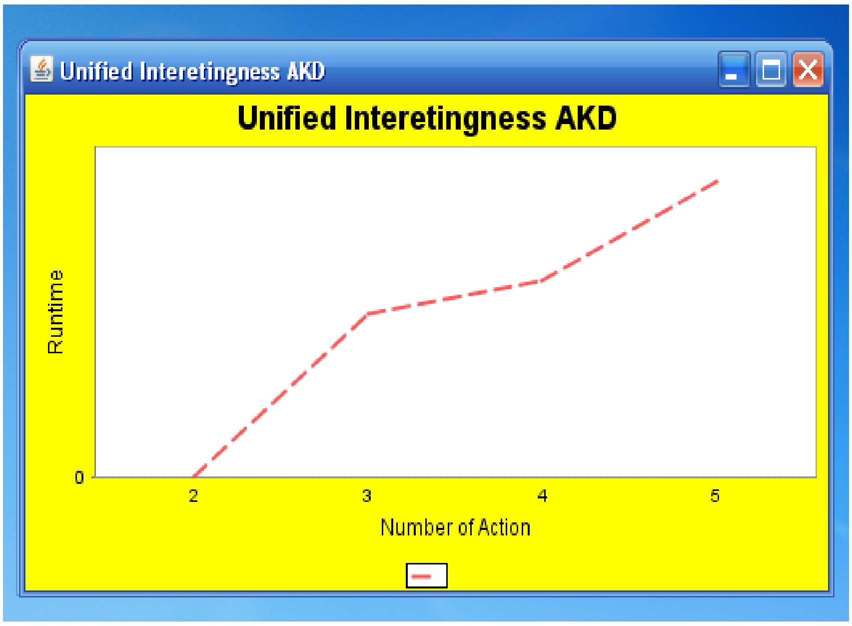

Figure 8: Unified Interetingness AKD Performance

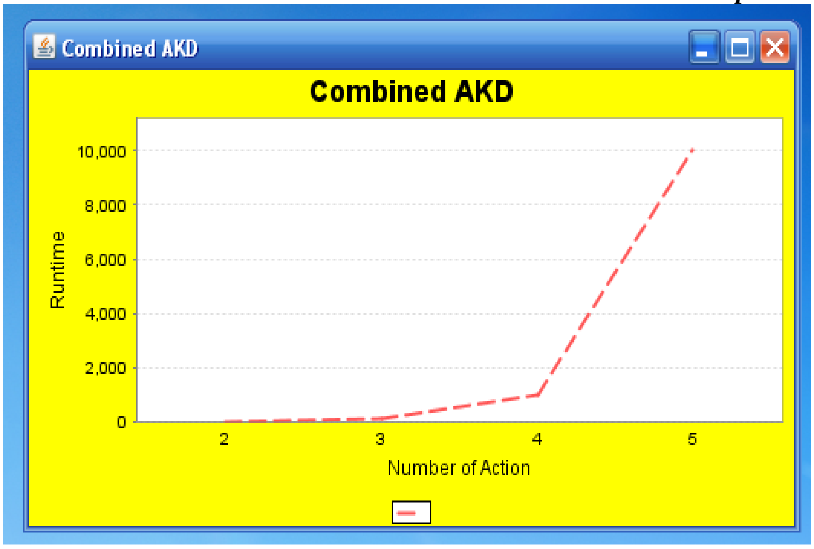

Figure 9: Combined AKD Performance

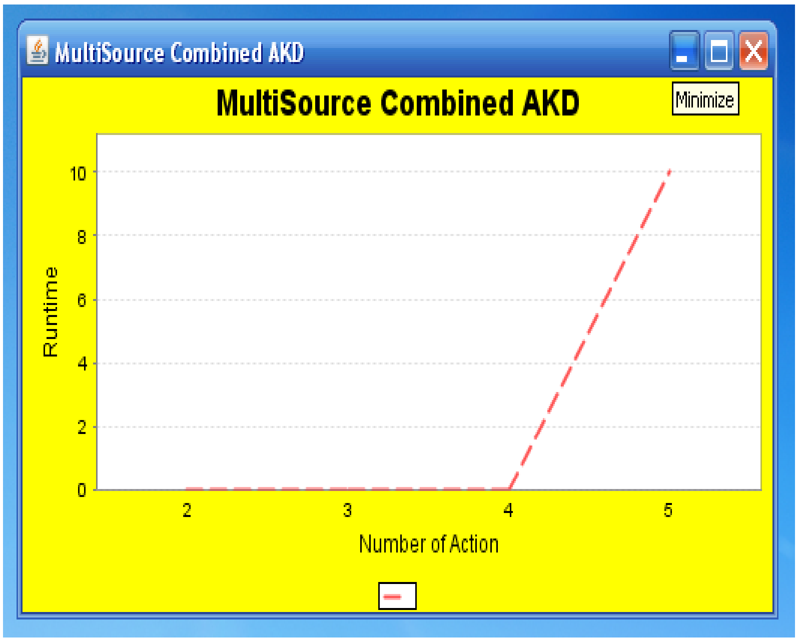

Figure 10: MultiSource Combined AKD Performance

From above performances and results, we can conclude that Framework 4 outperforms the other frameworks in case of decision making systems. It delivers the optimal solution to the problem and in quick time. This proposed framework helps to check the performances of these frameworks for business persons.

V. CONCLUSION AND FEATURE WORK

In this paper we have investigated the four AKD frameworks those are presented recently for overcoming the challenges faced while used AKD in business terms or decision making system. In this paper we have evaluated the performances of each this frameworks against the real time datasets of decision making system. There are four types of AKD frameworks those are presented here; these frameworks are capable of executing the different types of business problems as well as applications. These frameworks support closed-optimization-based problem solving from a business problem/environment definition, to actionable pattern discovery, and to operable business rule conversion. Deliverables extracted in this way are not only of technical significance but also are capable of smoothly integrating into business processes. The experimental analysis showing how we can check the performances of each of this proposed AKD frameworks, how they are generating the optimal solutions to the decision making systems in real time environment. For this future work we will suggest to address the other types of limited resources problem in order to fulfill this proposed framework further.

REFERENCES

[1] Longbing Cao, Yanchang Zhao, Huaifeng Zhang, Chengqi Zhang, "Flexible Frameworks for Actionable Knowledge Discovery", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 9, SEPTEMBER 2010.

[2] Qiang Yang, Jie Yin, Charles Ling, and Rong Pan, "Extracting Actionable Knowledge from Decision Trees", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 19, NO. 1, JANUARY 2007.

[3] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB '94), pp. 487-499, Sept. 1994.

[4] Bank Marketing Association, Building a Financial Services Plan:Working Plans for Product and Segment Marketing, Financial Sourcebooks, 1989.

[5] M. Belkin, P. Niyogi, and V. Sindhwani, "On Manifold Regularization," Proc. 10th Int'l Workshop Artificial Intelligence and Statistics, pp. 17-24, Jan. 2005.

[6] L. Cao, "Domain-Driven Actionable Knowledge Discovery," IEEE Intelligent Systems, vol. 22, no. 4, pp. 78-89, July/Aug. 2007.

[7] L. Cao, "Domain-Driven Data Mining: Empowering Actionable Knowledge Delivery," Proc. Pacific-Asia Conf. Knowledge Discovery and Data Mining (PAKDD '09) Tutorial, 2009.

[8] L. Cao, P. Yu, C. Zhang, and H. Zhang, Data Mining for Business Applications. Springer, 2008.

[9] L. Cao, Y. Zhao, C. Zhang, and H. Zhang, "Activity Mining: From Activities to Actions," Int'l J. Information Technology and Decision Making, vol. 7, no. 2, pp. 259-273, 2008.

[10] L. Cao, P. Yu, C. Zhang, and Y. Zhao, Domain Driven Data Mining. Springer, 2009.

[11] L. Cao, Y. Zhao, and C. Zhang, "Mining Impact-Targeted Activity Patterns in Imbalanced Data," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 8, pp. 1053-1066, Aug. 2008.