

# CONCEPTUAL THREE PHASE ITERATIVE MODEL OF KDD

Simmi Bagga  
Assistant Professor

Sant Hira Dass Kanya Maha Vidyalaya,  
Kala Sanghian, Distt. Kapurthala, Punjab

Dr. G.N. Singh

Department of Physics and Computer Science,  
Sudarshan Degree College, Lagaon, Rewa  
(M.P), India

## ABSTRACT

KDD process includes how data is stored and accessed, how and what algorithms can apply to large amount of data efficiently, how results can be interpreted and visualized. KDD is the process of identifying valid, interesting and understandable patterns in data. In this paper we will describe conceptual three Phase Iterative Model of KDD. The main layers of this purposed model are: Philosophy Layer, Technique Layer and Application Layer. We will also perform the comparison of Tradition KDD Model with Three Phase Iterative Model.

## Keywords

KDD, Philosophy layer, Technique Layer, Application Layer

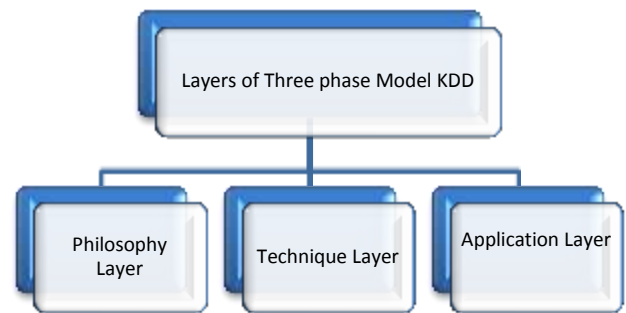
## 1. INTRODUCTION

Knowledge discovery in databases (KDD) is an emerging field that defines a set of techniques and tools for discovering useful information from large sets of database. It is the process that navigation knowledge from the data base according to user requirements. It includes how data is stored and accessed, how and what algorithms can apply to large amount of data efficiently, how results can be interpreted and visualized. The KDD process is interactive and iterative, involving various steps with many decisions being made by the user. It is the process of identifying valid, interesting and understandable patterns in data. Knowledge discovery is defined as the non trivial process of identifying valid, interesting, understandable and interesting information from the large data stored in the database.

## 2. CONCEPTUAL THREE PHASE ITERATIVE MODEL OF KDD

A conceptual framework is a particular set of rules, ideas in mind which we use in order to deal with problems or to decide what to do to solve a particular problem. KDD is multi-step process. The KDD process is interactive and iterative, involving various steps with many decisions being made by the user. KDD refers to the overall process of discovering useful knowledge from data. In the traditional KDD Steps if one of the phase take wrong decision then whole process will fail. If we modify this process by providing reverse feedback then the resulting knowledge will be more accurate than the traditional approach. The three phase conceptual framework consisting of three layers these are: the philosophy layer, the technique layer, and the

application layer. This framework represents the understanding what knowledge is required, discover it, and utilization of knowledge.



### 2.1 Philosophy Layer

The Philosophy layer mainly involves understanding the requirement of knowledge. The important issues related to this layer involve the representation of knowledge, the communication of knowledge in languages and the relationship between knowledge in the mind and with the real world and also involves organization of knowledge. Philosophical study deals with the technology that helps in understanding of our world and also establishes the operational boundaries of knowledge. This layer mainly involves the following subtasks:

Data Cleansing

Data Integration

Data Selection

Data Transformation

These steps mainly perform data selection according to the requirements, cleaning of data that involves the examining of data for completeness and integrity then data is transformed in the different format that is known as data transformation. This is one of the main steps of the KDD because the success of whole process depends upon this step. While performing this step, if any case we identified that the selected data set cannot provide

the accurate result then we can change this data set because this step can iterate. But in the case of traditional model if the wrong data set has been selected then the whole process can fail and the whole process has to start again in that case.

The subtask of this layer is further divided into two sub steps. In the first sub step we only consider those steps of KDD which includes independent parameters like Data Cleaning and Data Integration. Data cleaning cleans the data i.e. removing bad data and finding hidden correlations in the data also identifying sources of data that are the most accurate, and determining which columns are the most appropriate for use in analysis. Data may also include missing data information, duplicate data etc. It also involves the examining of data for completeness and integrity. The next subtask involve in this is Data Integration in which we merge data that was collected from the different forms. That data may have different formats.

In the second subtask only those steps are performed that contain dependant parameters. This sub task mainly involves Data Selection and Data Transformation. Data selection analyses the collected data from the different sources and decide from the collected according to some algorithms. In Data Transformation data is transformed in proper format that can be directly used for the mining process. Data transformation involves Smoothing to remove the noise from the data.

This layer of three phase iterative model repeats till we ensure the selected pattern will give the required result. The output of this step includes only selected subset of data that contain the information for which we are performing whole process. Once we get the appropriate subset we can move to the next layer of the model i.e. Technique layer.

## 2.2. Technique Layer:

This layer is mainly concerned to the study of discovering knowledge in machine. The main issues involve with this layer is how to discover knowledge. It also involves discovery methods by programming languages, techniques and algorithms used by the intelligent systems. The main step performed in this layer is Data Mining. Logical mathematical analyses are base of the technique layer.

Data mining is the iterative and interactive involves the various steps with the decision made by the user. Data mining is only the one stage in KDD process concerned with applying computational techniques to find patterns discovery in a data set from which noise has been previously eliminated and which has been transformed in such a way to enable the pattern discovery process. Data Mining required cleaned, transformed in proper format and coded data.

Data Mining is the core process that takes input cleaned & transformed data and searches patterns using some algorithms and then results patterns and relationships. These patterns and relationships are then used to the interpretation and evaluation phase to generate knowledge. This step will get only successful if the previous step performed accurate. If the any error occur in the initial stage the result of whole KDD process will not be feasible.

In this layer, we identify the patterns and relationships between data. Data Mining operation may generate thousands of

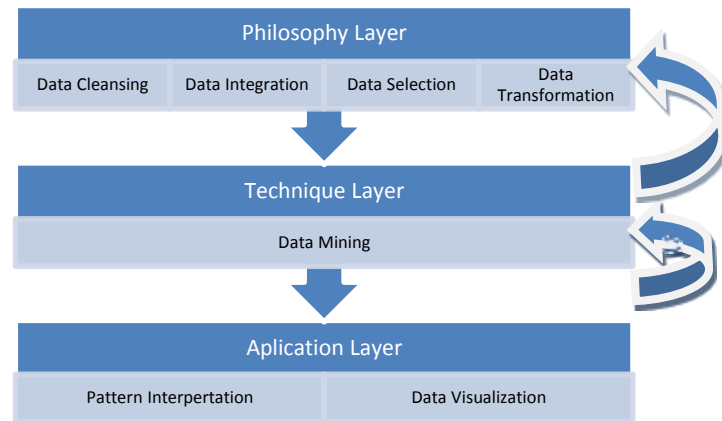
patterns but all of them are not interesting. To decide the interesting pattern we perform next step.

## 2.3 Application Layer

The overall goal of knowledge discovery is to effectively use of discovered knowledge. The application layer mainly focuses on usefulness and relevance of discovered knowledge for the particular domain. It's the last layer of the KDD process.

In this layer, results have to be interpreted and evaluated to discover knowledge from the patterns. Even the purpose of the model is to increase knowledge of the data, the knowledge will need to be organized and presented in a way that the customer can use it.

After completing Data Mining, we visualized the results. Visualization plays an important role in making discovered knowledge to understand and interpret by humans.



The ultimately goal of the conceptual model of KDD is to discover knowledge i.e. to extract knowledge to meet the requirement of the user. This conceptual model is an iterative model of KDD. It also provides the backward link from Technique layer to Philosophy layer. If any error or changes are required in the pattern selection while performing mining process we can select again by moving backward.

In this model of KDD the whole process is divided in to three layers i.e. Philosophy, Technique and Application layer. The Philosophy Layer involves task discovery, data selection, data cleaning and data transformation. The result of this stage is pattern selection to which we perform Data mining step. This model also provides the backward link from data mining step to pre-processing. If any error or changes are required in the pattern selection while performing mining process we can select again by moving backward.

## 3. COMPARISON OF TRADITIONAL AND THREE PHASE ITERATIVE MODEL OF KDD

The goal of both the models is same i.e. to extract knowledge to meet the requirement of the user. The working of both the model is totally different. In the traditional model whole process of KDD is divided in to seven steps and there is no any backward move to the previous step and if any problem detect in the later steps then there is no chance to recover that error.

In the Conceptual Three Phase iterative model of KDD, the whole process is divided in to three layers i.e. Philosophy, Technique and Application Layers. The Philosophy layer involves task discovery, data selection, data cleaning and data transformation. The result of this stage is pattern selection to which we perform Data mining step. This model also provides the backward movement from data mining step to previous layer. If any error or any changes are required in the pattern selection while performing mining process we can change.

The main comparative points of these two models are:

#### 4. SUMMARY

KDD is the process that is used to extract knowledge from data. In the traditional KDD model, whole process is divided into seven steps and on the other hand in the Conceptual Three Phase iterative model of KDD the process is divided into three steps

Traditional KDD Model	Conceptual Three Phase Iterative Model of KDD
It is straightforward and easy to use process.	This model also moves backward from the any step that provides results according to the user.
It is less time consuming.	It takes little more time but always provide accurate result.
If the requirement is not fully specified then we have to repeat whole steps again & again.	In the case of this method we can refine our process at any step.
The knowledge obtained from this model is not much reliable	The results obtained from this model is more reliable because always give accurate results according to the user requirement.
It is not much flexible because user can not even little change in the requirement after selecting the pattern for data mining process	This model is flexible because if any change occurs at any step we can more back and fulfill that changed requirements.

and each step has forward as well as backward link with each other.

Three Phase iterative KDD model provides us more accurate result than the traditional model because in traditional process if the starting step i.e. Task Analysis in which we perform selection in which an appropriate procedure is selected for generating the target data set from the database and all other steps are performed on the basis of this selection process. But if we selected the wrong data set the whole process may fail. On the other hand in the Three Phase iterative model if we noticed that selected target data is wrong at any stage then we can go backward and can rectify our problem.

#### 5. REFERENCES

- [1] Fayyad, U.M. “Data Mining and Knowledge Discovery: Making Sense Out of Data”, IEEE Expert 11(5), 1996
- [2] Elder, John F., IV and Daryl Pregibon, (1996), Advances in Knowledge Discovery & Data Mining, “A Statistical Perspective on KDD.
- [3] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R. 1996. Advances in Knowledge Discovery and Data Mining. Menlo Park, Calif.: AAAI Press.
- [4] Han Jiawei and Kamber, Micheline. “Data Mining: Concepts and Techniques”. 2001. Morgan Kaufmann. Sanfransico, CA
- [5] Lee, Sang Jun & Keng, Siau (2001), A Review of Data Mining Techniques, Industrial Management & Data Systems: Volume 101
- [6] Dunham, M.H. (2003). Data mining introductory and advanced topics. Upper Saddle River, NJ: Pearson Education, Inc.
- [7] Grossman, Robert. L., Hornick, Mark. F. and Meyer, Gregor. “Data Mining Standards Initiatives.” (Communications of the ACM, Vol. 45, No. 8, 2002,59-61)
- [8] Banerjee A, Merugu S, Dhillon I, Ghosh J (2005) Clustering with Bregman divergences. J Mach Learn
- [9] Singh G.N, Bagga Simmi (2011) Three Phase Iterative Model of KDD, International Journal of Information Technology and Knowledge Management, Volume 4, No. 2, pp.695-697.
- [10] Singh G.N, Bagga Simmi (2011), Clustering Method for categorical and Numeric Data sets, Global Journal in computer Science.