

# Logic based Pattern Discovery using the Integral Logical Derivative Rule

Prasadh.K  
Mookambika Technical Campus,  
Muvattupuzha,  
Kerala, India.

Sutheer.T  
Research Scholar ,  
Manonmanium Sundarnar university,  
Tamil Nadu, India.

## ABSTRACT

**Problem Statement:** To enhance the pattern discovery process, the multi-level propositional process work extends the pattern discovery process with coherent rule generation framework. The multi-level coherent rule structure produce rules coming from diverse levels and determine highest recurrent item sets at inferior level. The propositional logic process formed the multilevel connection rules from logical rules and utilizes bottom-up progressive extending technique. This method develops the effectiveness of rules with minimum support threshold but takes longer time. **Approach:** To overcome the above issue, we are going to implement a new technique termed Logic based pattern discovery using Integral Logical Derivative Rules (ILDR). This technique is used to efficiently produce the rule with the short span of time. **Results:** Performance of Integral Logical Derivative Rules technique to discover the logic based pattern is evaluated in terms of execution time, support threshold based on number of items and memory consumption for pattern discovery. **Conclusion:** Logical based pattern discovery considers the problem of minimum support threshold. An analytical and empirical result shows the lesser execution time with the efficient integral based pattern discovery of our proposed scheme.

**Keywords:** Pattern discovery, Derivative rules, Integral based pattern, support threshold, data mining, mining methods.

## 1. INTRODUCTION

Data conversion is an integral part of data mining and knowledge discovery. Renovate data allows for an amplified understanding of the data and discovery of new and interesting relationships between features. There are numerous methods for converting data, for example:

- Arithmetic operators: (+, -, \*, /). This method may engross arithmetic operators useful to a single or multiple features (e.g.,  $B^2$ ,  $B/C^3$ ).
- Combination of features: This renovation technique combines two or more features to form a new feature.
- Discretization of features: This method encompasses enchanting raw feature values and grouping them with comparable values.
- Denoising features: Fourier transforms, wavelet transforms, moving averages and so on.

With the development of computer hardware and software and the rapid computerization of business, huge amount of data have been composed and stored in databases. The rate at which such data is stored is rising at a phenomenal rate. As a result,

traditional ad hoc mixtures of arithmetical techniques and data management tools are no longer sufficient for analyzing this vast collection of data. Numerous domains in which the large volumes of data are stored in centralized or distributed databases such as

- Financial Investment: Stock indexes and prices, interest rates, credit card data, fraud detection.
- Health Care: A lot of analytical information are stored by hospital management system.
- Manufacturing and Production: Process optimization and trouble shooting.
- Telecommunication network: Calling patterns and fault management systems.
- Scientific Domain: Astronomical observations, genomic data, biological data.

Raw data is hardly ever of direct benefit. Its true value is predicated on the ability to take away information cooperative for decision support or exploration, and understanding the phenomenon governing the data source. In most domains, data analysis was conservatively a manual process. One or more analysts would turn into familiarly familiar with the data and, with the assist of statistical techniques, offer summaries and produce reports.

In effect, the analyst acted as a complex query processor. However, such a come up to rapidly breaks down as the amount of data grows and the number of dimensions increases. When the scale of data manipulation, exploration and inference goes further than human capacities, people look to computing technologies for automating the process.

All these have encouraged the need for intelligent data analysis methodologies, which could find out useful knowledge from data. The expression KDD refers to the overall procedure of knowledge discovery in databases. Data mining is a meticulous step in this process, involving the application of exact algorithms for extracting patterns from data. The supplementary steps in the KDD process include data preparation, data collection, data clear out, suitable prior knowledge integration, and proper interpretation of the results of mining. It guarantees that useful knowledge is resulting from the database.

The method we propose consists in analyzing the knowledge database used to produce the rule-based system. We extract well-organized subsets of features using the distinctive analysis and the Integral Logical Derivative Rule. Distinctive analysis is based on the similarity and the dissimilarity of one sample compared to the defect classes. The Integral Logical Derivative

Rule is part of the aggregation techniques based on Neuro-fuzzy integrals. Neuro- Fuzzy integrals, and the Integral Logical Derivative Rule together used as a fusion operator, and it have been successfully used as fusion operators in various applications.

## 2. LITERATURE REVIEW

The basic goal of data mining is to discover patterns occurring in the databases, such as associations, classification models, sequential patterns, and so on. Nittaya., kittisak kerdprasop., 2010, focus on the problem of frequent pattern discovery, which is the process of searching for patterns such as sets of features or items that appear in data frequently. Fuzzy logic provides a mathematical framework that is compatible with poorly quantitative yet qualitatively significant data. Mishra, D., and Satapathy, S.K., 2011, have fuzzified our original dataset and have applied various frequent pattern mining techniques to discover meaningful frequent patterns.

Ning Zhong.,et.AL., 2012 presents an innovative and effective pattern discovery technique which includes the processes of pattern deploying and pattern evolving, to improve the effectiveness of using and updating discovered patterns for finding relevant and interesting information but consumes higher cost. Man Lung Yiu., et.AL., 2012 presents techniques that transform the data prior to supplying it to the service provider for similarity queries on the transformed data. This techniques provide interesting trade-offs between query cost and accuracy.

Donghun Lee.,et.AL., 2012 uses Statistical Process Control (SPC) charts to detect performance anomalies and differential profiling to identify their root causes. By automating the tasks within the framework we were able to remove most of the manual overhead in detecting anomalies and reduce the analysis time for identifying the root causes by about 90 percent in most cases.

Kyriacos E. Pavlou., and Richard T. Snodgrass., 2010 introduces the notion of a candidate set and provides a complete characterization of the candidate set and its cardinality. An optimal algorithm for computing the candidate set is also presented with a comparison to other forensic algorithms in terms of space/time complexity and cost. To identify the query result produce in the database Weifeng Su.,et.AL., 2010, present an unsupervised, online record matching method, UDD, which is used for a given query, can effectively identify duplicates from the query result records of multiple Web databases.

K.Venkateswara Rao., et.AL., 2012, issues and challenges related to spatiotemporal data representation, analysis, mining and visualization of knowledge are presented. Various kinds of data mining tasks such as association rules, classification clustering for discovering knowledge from spatiotemporal datasets are examined and reviewed. Jyoti Jadhav., et.AL., 2012 presents a novel method for incremental discovery of frequent patterns using Main Memory database Management System to eliminate the highly dependent main memory.

Yuzhe Tang., et.AL, 2010 propose LIGHTweight Hash Tree (LIGHT) a query-efficient yet low-maintenance indexing scheme. LIGHT employs a novel naming mechanism and a tree summarization strategy for graceful distribution of its index structure. We show through analysis that it can support various

complex queries with near-optimal performance. Extensive experimental results also demonstrate that, compared with state of the art over-DHT indexing schemes.

Feng Qian.,et.AL., 2012, propose a novel iterative mining framework that discovers spatial co-location patterns without predefined thresholds. With the complete and relative prevalence of spatial co-locations, this method allows users to iteratively select informative edges to construct the neighborhood relationship graph until every significant co-location has enough confidence and eventually to discover all spatial co-location patterns.

To reduce the time consumption in the logic based pattern discovery, a new technique named logic based pattern discovery using the Integral logical Derivative Rule (ILDR) scheme is presented.

## 3. LOGIC BASED PATTERN DISCOVERY USING THE INTEGRAL LOGICAL DERIVATIVE RULE

The architecture diagram of the logic based pattern discovery using the integral logical derivative rule is shown in the Fig 3.1.

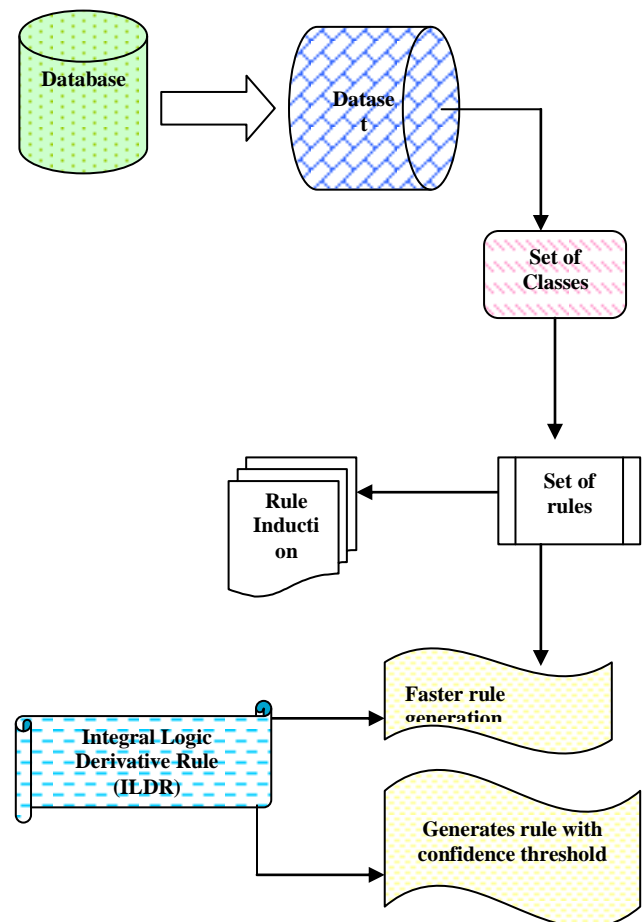


Fig 3.1 Architecture Diagram of Integral Logical Derivative Rule

The above diagram describes the Integral Logical derivative rule using the real dataset. The database contains the set of datasets which derives the set of classes. It is being observed that the dataset normally have set of classes to be processed. By mining association rule process, set of rules have been discovered for the classes specified. With the integral logic applied to the set of rules, it produces the faster rule production and generates the rule with the minimum threshold.

#### Steps to perform the Integral Logical Derivative rule:

First an initialization is done as step 0. Then an iterative global pattern discovery process is performed and can be roughly split into two steps as step 1 and step 2.

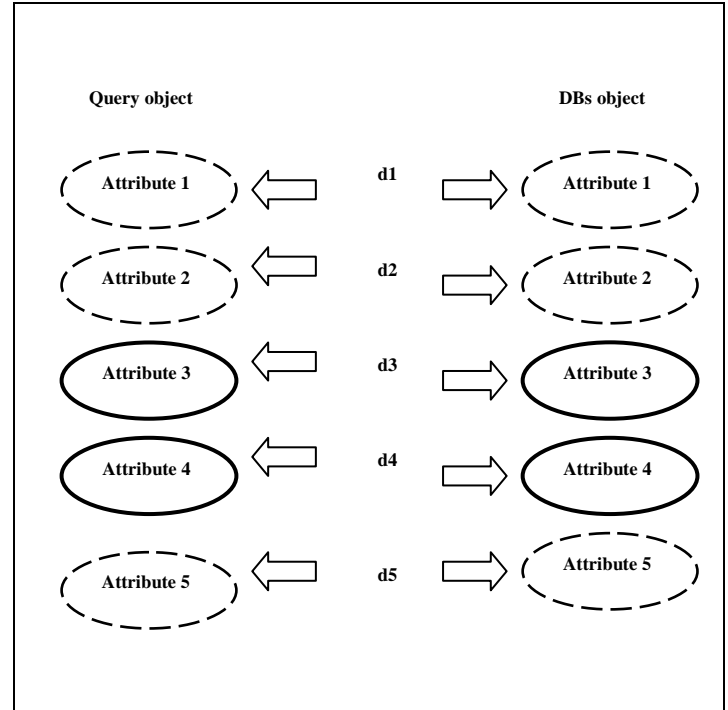
- Step 0 contains the application of distinctive analysis allows to propose a primary set of patterns, and also to validate this choice by training and testing the acknowledgment model with them. A universal acknowledgment rate is obtained and assumed to be a reference set.
- Step 1 contains the first set of patterns; a pattern interactivity process is applied to determine the less representative ones.
- Step 2 produces an acknowledgment model with no first less representative patterns and tests. The reached acknowledgment rate is stored. The process is repeated using the k next less representative patterns.

If one of the acknowledgment rates is better or equal to the early one (position rate obtained in step 0), the weaker associated pattern is removed from the list and go to Step 1 (with n-1 patterns) and so on. Else, the first set of logical patterns given by the step 0 is kept and the associated acknowledgment model is used for the final pattern acknowledgment application.

### 3.1 Determination Of Logical Patterns Using Integral Method

The suitable pattern can be analyzed using the Integral Logical Derivative Rule. In database the higher frequency of occurrences are expected to separate the dimensions. Therefore in the general cases the compound dissimilarity indicator is used to between the two objects. It will be calculated from the weighted average of individual dissimilarities along each of the dimensions. For a collection of 'n' integral patterns, the integral rule methods are accepted to obtain the dissimilarity of the patterns with respect to the integral group.

It resolves the overall dissimilarity 'D' between the DBs object and the Query object. Fig 3.2 describes the integral pattern discovery process. The Figure contains the query object and the DB object with the list of the dissimilarity values.



In the diagram

————— - describes the integral part

----- - describes the separable part

$$D = \frac{d1 + d2 + d5 + \sqrt{d3^2 + d4^2}}{\sqrt{2}}$$

The attribute 3 and attribute 4 are the integral part of the pattern discovery process. The integral dimensions are powerfully unanalyzable and classically perceived as a single stimulus. Consequently the grouping of 'n' patterns in an integral collection is supposed to yield a one dissimilarity section rather than the 'n' entity section in the complex compute. The above diagram Lists of attributes are present in the Query object in which the two integral attributes namely the attribute 3 and attribute 4 are combined together. It is combined with the help of dissimilarity values d3 and d4 into a single dissimilarity component.

#### 3.1.1 Integral Logical Derivative Rule

In particular, we consider discrete spaces only for a thorough description of the Integral Logical Derivative Rule. Let us consider n classes, M1, M2, M3..., Mn, and n Rule Induction (RI) where it is represented in the form of R1, R2, . . . ,Rn. When a new pattern y0 is observed, we wish to find the class. Labeling this unknown pattern is a three-step process.

Firstly, for each rule induction ‘1’ we need to compute  $\phi$ , the degree of assurance in the statement “According to R1,  $y_0$  belongs to the class C”. Secondly, we combine all these partial assurance degrees into a universal assurance degree by choosing a suitable aggregation operator.

It presumes that in the nonappearance of any information, the most realistic way of aggregation is the arithmetic mean, thus an Integral Logical Derivative Rule with reverence to an additive measure. Any input of information tends to move away the measure from this equilibrium point. This means that, in case of few data, coefficients of the measure which are not concerned with the data are kept as near as possible to the equilibrium point. Thus, this rule is still efficient when training data are limited. It also has a low computing time and a low memory cost.

### 3.2 Pattern Discovery Methods

In this section, we introduce the two pattern discovery methods used for our proposed pattern discovery process. The first one is the distinctive analysis which consists in extracting the initial set of patterns from the knowledge database. The second one is the pattern interactivity analysis. It consists in investigating the initial list of patterns to extract the three less delegate ones. The result of this analysis is a matrix which coefficients indicate how significant the patterns are.

#### 3.2.1 Distinctive analysis

Distinctive gives information about the delegate of a sample for a pattern. In our case, as each sample of the knowledge database is labeled, the distinctive notion returns information about the delegate of a defect class for a pattern. It is appraise with the calculation of the typicality coefficient for the dissimilar features. The distinctive calculation as described in the Eqn 3 is based on the detain dissimilarity as on Eqn 1 and extern similarity as on Eqn 2 are described.

$$D(W_a^u) = \frac{\sum_{i=1}^n 1/d(W_a^u, W_a^{fi})}{n} \quad \text{---Eqn 1}$$

$$E(W_a^u) = \frac{\sum_{i=1}^m 1/1-d(W_a^u, W_a^{ei})}{m} \quad \text{---Eqn 2}$$

$$I(W_a^u) = \frac{D.E}{D.E + (1-D).(1-E)} \quad \text{---Eqn3}$$

Where,

$W_a^u$  - Value of the pattern a for the point u;

$W_a^{fi}$  - Value of the pattern a for the “companion” point

f - Class i;

$W_a^{ei}$  -Value of the feature a for the “adversary” point e of the class i;

d (a, b) is the Euclidian distance between a and b;

n - Number of “companion” points;

m - Number of “adversary” points;

D and E correspond to  $D(W_a^u)$  and  $E(W_a^u)$

#### 3.2.2 Interactivity Analysis

From the above definitions, we can see that the behavior of the Integral Logical Derivative Rule as an aggregation operator entirely relies on the neural - fuzzy measure used. There are several methods to determine the most adequate neural -fuzzy calculate to be used for a given application. To manage with initialization problems “Integral” derivative rules have been developed. Their objective is to find an approximation of the neural- fuzzy measure that minimizes an error criterion.

Once the neural-fuzzy measure is learned, it is possible to understand the contribution of each decision rule in the final decision. Several indexes can be extracted from the neural- fuzzy measure, helping to examine its behavior.

**Magnitude index:** The magnitude index in game theory is based on the classification planned and introduced in a neural-fuzzy Measure.

**Interface index:** The interaction index represents the degree of interface between two decision rules. If the neural-fuzzy measure is non-additive then some sources interact.

The above two methods are used to discover the patterns efficiently using our proposed scheme.

### 4. EXPERIMENTAL EVALUATION

The proposed logic based pattern discovery is efficiently done by implementing integral logical derivative rule. The proposed logic based pattern discovery using Integral Logical Derivative Rule (ILDR) is implemented in Java. It is experimented with real and or synthetic data to assess the rule induction for generating the set of rules. The results are also evaluated with logic based pattern discovery with dissimilar set of classes which contains the both the amount of strong rules and weak rules produced. Effects of non sensitive rules are consequent in both the presented and proposed strategies to illustrate the effectiveness and faster way of novel rule mining process.

The representations works well with troubles connecting ambiguity in data relationships, which are denote by multilevel notions for integral logic derivative rule. The proposed ILDR mining algorithm can thus engender huge item sets level by level and then obtain notion multilevel association rules from operation dataset with the short span of time. The outcome revealed in the pattern implies that the proposed ILDR

algorithm can obtain the multiple-level association rules under diverse integral logic in an easy and valuable way. The performance of the proposed logic based pattern discovery using Integral Logical Derivative Rule (ILDR) is measured in terms of

- Support Threshold
- Execution Time
- Memory Consumption

## 5.RESULTS AND DISCUSSION

In this work, we efficiently evaluated the consistent rules established using the proposed pattern discovery using Integral Logical Derivative Rule ILDR. Two variants of thresholds are positioned for the a priori: The smallest support threshold is positioned at 5 percent and the least amount declaration threshold is positioned at 50 percent. Depends on this position, all association rules not continued by no less than 5 percent of procedure records are considered unbalanced; or else, they are considered recurrent. Here, we group the rule induction into two classes:

- Infrequent rule induction contain smaller amount of sensible classes
- Frequent rule induction have frequently observed

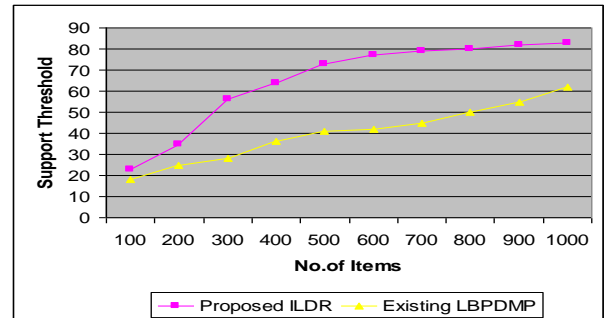
In indulging the outcomes, the attribute values of the dataset are represented as attribute names followed by its feasible values. The below table and graph describes the performance of the proposed pattern discovery using Integral Logical Derivative Rule (ILDR) scheme compared with an proposed Logic Based Pattern Detection using Multi-level Propositional logic (LBDPMP)

**Support Threshold:** The Integral Logical Derivative Rule (ILDR) performs a confidence support threshold with the large item sets generated instantly.

No. of items	Support Threshold	
	Proposed ILDR	Existing LBDPMP
100	23	18
200	35	25
300	56	28
400	64	36
500	73	41
600	77	42
700	79	45
800	80	50
900	82	55
1000	83	62

**Table 5.1 No. of items vs. Support Threshold**

The set of rules generated based on the rule induction obtained by implementing both the proposed Integral Logical Derivative Rule (ILDR) scheme and LBDPMP. The above table (table 5.1) describes the support threshold based on the number of items participate in it.



**Fig 5.1 No. of items vs. Support Threshold**

Fig 5.1 describes the support threshold based on the number of rules created. The proposed ILDR efficiently identified the rule inductions that have the properties of integral logic in terms of attaining the domain knowledge of the patterns identified. And an integral logic framework is used here to generate confidence threshold by adapting the pseudo implications of equivalences from a given data set. With the integral logic, the proposed ILDR determine all rule induction that are rationally correct devoid of having to identify the domain knowledge in a less interval of time.

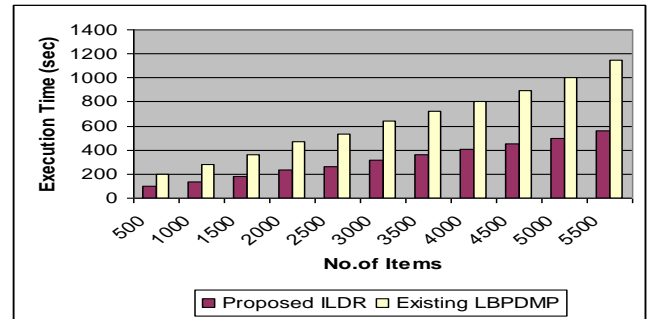
The outcome of the proposed Integral Logical Derivative Rule is compared with the Logic based pattern discovery using the Multi-level Propositional logic techniques. In the proposed ILDR scheme, each generated rule induction is decoupled into two pseudo inferences of equivalences which can be additionally mapped to a balanced equivalence. The data quality management with the threshold by J.Malar Vizhi.,et.Al., 2012 indicates the quality rules. The proposed ILDR scheme provides a confidence threshold using the integral group which yields a single dissimilarity component. The variance in confidence threshold generation is 30-40% high in the proposed ILDR.

**Execution Time:** The execution time of the Integral Logical Derivative Rule is the generation of rules based on the number of items present in the given dataset.

No. of rules	Memory consumption (KB)	
	Proposed ILDR	Existing LBPDMP
1000	1120	1300
2000	1145	1450
3000	1160	1555
4000	1178	1620
5000	1192	1740
6000	1210	1830
7000	1225	1940
8000	1242	2000
9000	1265	2100
10000	1295	2220

**Table 5.2 No. of items vs. Execution time**

Fig 5.2 described the execution time consumed to generate the rules based on the rule induction from the set of classes. In the proposed ILDR scheme, the process of generates the rule consumes less time with the confidence threshold. The outcome of the proposed Integral Logical Derivative Rule is compared with the Logic based pattern discovery using the Multi-level Propositional logic techniques.



**Fig 5.2 No. of items vs. Execution time**

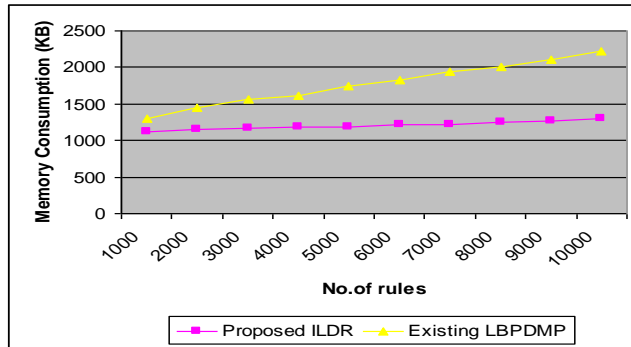
Fig 4.2 describes the consumption of time taken to produce the frequently and infrequently set of rules based on the number of items. Normally, rules contain more rigorous manacles contrast to distinctive association rules with a minimum confidence value. Although the existing framework by Alex Tze Hiang Sim et. Al., 2010 that discover association rules supported on quantify of interests other than a least support. In this case, Propositional logic for coherent rule generation produces both positive and negative relationship rules. So, the existing framework consumes more time to proceed in which the proposed ILDR generate the rules based on the confidence threshold value with the integral logic. Compared to a multi-level hierarchy for mining association rules, the proposed ILDR based rule induction mining consumes less time to process the rules using integral logic and obtained a high quality of knowledge with the short span of time.

**Memory consumption:** The memory consumption rate is the rate at which memory is consumed for storing the rules in the Integral Logic Derivative Rule (ILDR).

No. of items	Execution time (seconds)	
	Proposed ILDR	Existing LBPDMP
500	100	200
1000	140	280
1500	180	360
2000	235	470
2500	260	530
3000	320	640
3500	360	720
4000	410	800
4500	450	890
5000	500	1000
5500	560	1150

**Table 5.3 No. of Rules vs. Memory Consumption**

Fig 5.3 described the memory consumption of rules from the database. In the proposed logic based pattern discovery using Integral Logical Derivative Rule (ILDR) scheme, it generates the rule easily and helps to identify the data attributes rapidly with lesser memory consumption. The data attributes fetching from the dataset consumed less memory compared to an existing LBPDM. Comparison result of ILDR with an existing Logic based pattern discovery using the multi-level propositional logic based on memory consumption variance, measured in terms of KiloByte (KB).



**Fig 5.3 No. of rules vs. Memory consumption**

The performance graph of the proposed Integral Logical Derivative Rule (ILDR) scheme for memory consumption rate is shown in the fig 4.3. The integral approach will identify and apply the rules between the integral logic derivations; this approach uses the aggregation operator and makes the lesser consumption of memory when compared to the LBPDM technique. The variance in the memory consumption would be 15-20% low in the proposed Integral Logical Derivative Rule (ILDR) scheme.

Finally, it is being observed that the proposed ILDR eradicates the time consuming process to generate the rules and its mixtures to decide suitable confidence threshold. The proposed Integral Logic Derivative Rules determines the normal threshold supported on surveillance of data set. It presents a balanced reinforcement to the detection procedure of patterns by consuming less running time.

## 6.CONCLUSION

In this work, we efficiently performed the logic based pattern discovery using the Integral Logical Derivative Rule (ILDR). It is efficiently analyzed to produce the rule based system. These rules comprise item sets that are commonly and uncommonly processed in a set of transaction records. We extract well-organized subsets of features using the distinctive analysis and the Integral Logical Derivative Rule. The proposed distinctive analysis is based on the similarity and the dissimilarity of one sample compared to the defect classes. The Integral Logical Derivative Rule is part of the aggregation techniques based on Neuro-fuzzy integrals. Neuro- Fuzzy integrals and Integral Logical Derivative Rule together used as a fusion operator, and it have been successfully used as fusion operators in various applications. Thus the Logical based pattern discovery produces the confidential based support threshold. An analytical and empirical result shows the 30 – 40 % lesser execution time with

the efficient integral based pattern discovery of our proposed scheme.

## 7. REFERENCES

- [1] Man Lung Yiu., Ira Assent., Christian S. Jensen., and Panos Kalnis., "Outsourced Similarity Search on Metric Data Assets" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 2, 2012
- [2] J.Malar Vizhi., and Dr. T.Bhuvaneswari., "Data Quality Measurement With Threshold Using Genetic Algorithm International Journal of Engineering Research and Applications (IJERA), 2012
- [3] Alex Tze Hiang Sim et. Al., 'Logic based pattern discovery', IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 6, JUNE 2010
- [4] Nittaya., kittisak kerdprasop., "The Discovery of Frequent Patterns with Logic and Constraint Programming," Recent Researches in Computational Techniques, ISBN: 978-1-61804-011-4, 2010
- [5] Ning Zhong., Yuefeng Li., Sheng-Tang Wu., "Effective Pattern Discovery for Text Mining," IEEE Transactions on Knowledge and Data Engineering, 2012
- [6] Donghun Lee., Sang K. Cha., Arthur H. Lee., "A Performance Anomaly Detection and Analysis Framework for DBMS Development," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 8, AUGUST 2012
- [7] Mishra, D., Satapathy, S.K. , "Fuzzy pattern tree approach for mining frequent patterns from gene expression data," International Journal on Computer Applications, Volume: 2, 2011
- [8] Kyriacos E. Pavlou., and Richard T. Snodgrass., "The Tiled Bitmap Forensic Analysis Algorithm," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, APRIL 2010
- [9] Weifeng Su., Jiying Wang., and Frederick H. Lochovsky., "Record Matching over Query Results from Multiple Web Databases," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 4, 2010
- [10] K.Venkateswara Rao., A.Govardhan., and K.V.Chalapathi Rao., "Spatiotemporal Data Mining:issues, tasks and Applications," International Journal of Computer Science & Engineering Survey (IJCSES) Vol.3, No.1, 2012
- [11] Yuzhe Tang., Shuigeng Zhou., and Jianliang Xu., "LIGHT: A Query-Efficient Yet Low-Maintenance Indexing Scheme over DHTs," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 22, NO. 1, 2010
- [12] Feng Qian., Qinming He., Kevin Chiew., Jiangfeng He., "Spatial co-location pattern discovery without thresholds," Knowledge and Information Systems, Volume 33, Issue 2, pp 419-445, November 2012
- [13] Jyoti Jadhav., Lata Ragha., Vijay Katkar., "Incremental Frequent Pattern Mining," International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-6, August 2012