

Investigation Study on Vantage Point Indexing Search in Multiple Data Pattern Mining

^{1,|||}Dr R.S. Vetrivel, ^{|||}J. Niyoperpasiline, ^{|||}P.Jeyanthirani

^{|||}Professor, Computer Science, Subramanya College of Arts & Science, Palani.

^{|||}Research Scholar, Computer Science, Subramanya College of Arts & Science, Palani

^{|||}Assistant Professor, Computer Science, Subramanya College of Arts & Science, Palani

Abstract

Vantage Point Indexing Search in Multiple Data Pattern Mining maintains the multiple users and produces enhanced search efficiency in internet. Vantage Point Indexing support nearest neighbor points and improve the search efficiency as per user interest. Multiple data pattern handles multiple databases and filters the data by using pattern mining approach. It creates considerable relation rules between sets of items from database tables containing data with optimal taxonomic structures.

Multiple databases have combination of authoritative information that is unmatched in online service. Multiple databases on various domains contain materials from hundreds of tables. But, there is no particular domain for presenting all the user information along with their relevant sources. Probably require a search of more than one database and possible to search each database one-by-one but required increased memory space and time.

Initially, Pattern Mining is performed for ensuring a given sequence of pattern with the occurrence of some other pattern elements. The pattern recognition filter the data with pattern mining either in the form of sequences or tree structures. Next, Sequential Pattern Mining technique is developed for discovery statistically relevant patterns between data in a sequence. The data values are discrete and related with time series mining, but frequently measured a different activity. Sequential pattern mining is a structured data mining. A vantage-point is a relating point where the data is separated based on the selection of vantage-point position. The separation of data point is carried out and they are searched through vantage point for providing nearer neighbor vantage point. Finally, Vantage Point Indexed search queries are generated in relational query processors and Indexed algorithm. They produce the results on indexed query search optimized multiple data pattern mining. The two substantial query operators based on modification, index-join algorithm is implemented and the operators are non-blocking that are integrated into pipelined execution. Therefore, a vantage point heuristics combine original join operators in query processors and implement in a prototype database.

A vantage-point tree determines the nearest neighbor of a point with the development of search algorithm. A vantage point and a threshold distance is major consideration in the point interest for searching the nearest sub tree. Vantage Point Index search is distance-based index structures for parallel queries on huge metric spaces using reference points (vantage points). The separation of relational data space into spherical region is processed with the help of vantage point in hierarchical manner. The multi-vantage point tree structure uses more than one vantage point to separate the space into spherical cuts at each level.

Feasibility and effectiveness of the proposed approach are tested using real online data search implementation on the social networking space and also using following Metrics. The evaluation is carried out in terms of True Positive Rate, Search Time and Search efficiency.

Key words

Pattern Mining, Sequential Pattern Mining, Multiple Data Pattern Mining and Vantage Point Indexing Search.

I. Introduction

A. Data Mining

Data mining is the process of extracting the hidden predictive information from large databases to provide on essential information. Data Mining is the removal of unknown analytical information from large records. It also helps to locate the hidden patterns, predictive information to use the specialists with solution outside their expectations. The aim of data mining is to remove the knowledge from dataset in human-understandable structures. It is described as a method of extraction and analysis of patterns, relationships and information from huge databases.

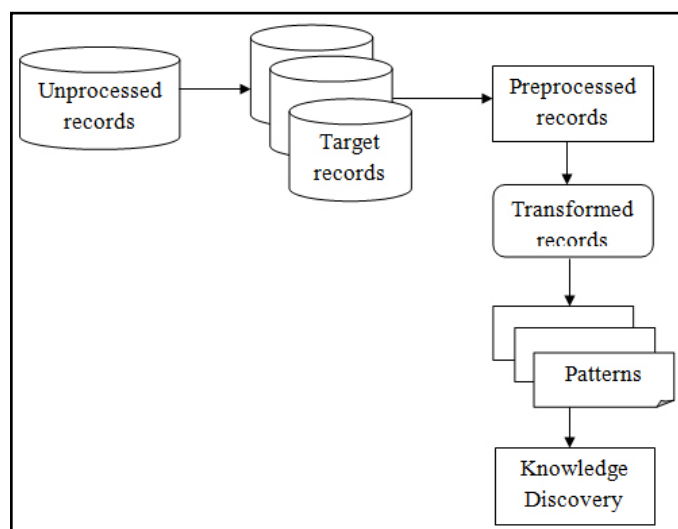


Fig. 1.1 : Process of extracting the hidden predictive information

The data mining process is called as Knowledge Discovery in

databases (KDD). Mining process consists of three phases namely, pre-processing, data mining and post-processing for the extraction of data patterns. Initially, pre-processing phases is developed that involves data cleaning, integration, selection and transformation. Next, data mining approach is carried out for creating the hidden knowledge. Finally, a post-processing computes the mining result along with user's needs and domain knowledge.

Data mining is the method of determining the connections or patterns with in large relational records. Data Mining represents a method planned to examine large amount of data gathered. It is the process of identifying the data from many views and summarizing it into useful information. Data mining software is a kind of systematic tools for analyzing the data. Data collected from different areas like marketing, health, communication are employed in data mining. . Data mining techniques are used in very large interesting organizations and data investigations. Most of the data mining approaches exploit classification related methods for identification of useful information from continuous data streams.

1. Database

A database is an organized for the collection of data and the data are characteristically prepared to model relevant features that support processes requiring the information. Database management systems are specially designed applications that interact with the user, other applications. The database itself captures and analyzes the data in relational database. A general-purpose database management system is a software system designed to allow the definition, creation, querying, update, and administration of databases.

2. Relational Database Management System

A relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model. Relational databases have often replaced legacy hierarchical databases and network databases because they are easier to understand and use. However, relational databases have been challenged by object databases, which were introduced in an attempt to address the object-relational impedance mismatch in relational database, and XML databases.

B. Association Rule Mining

Association rule mining is the considerable task and essential techniques of data mining to remove the attractive connections, frequent patterns, connections or casual structures connecting the sets of items in the transaction databases or data repositories. Telecommunication networks, market, risk management and inventory control are some of area that operates using association rules. Many association mining techniques and algorithms are designed. Association rule mining is the significant data mining to remove the attractive connections, frequent patterns, associations or casual structures linking the set of items in the transaction databases or other data repositories.

Association rule mining is designed with predefined minimum support and confidence from a specified database. The data in the item sets are located with larger space than a specified threshold in the database. The type of item sets is called as frequent or large item sets. Initially, item sets create the data from transactional database and creates the frequent association rules from the frequent items. The low minimum support is set to collect the least item set that increases the computational result and its complexity. Tracing is significant for domain applications as it disclose fundamental

and hidden information. The information is utilized by users as a guideline or method for prevention or improvement of the particular processes in specific domain.

C. Vantage Point Indexing Search

A vantage-point separate data in a metric space by choosing a position in the space and dividing the data points into two divisions. The division of data points is nearer to the vantage point than a threshold and those that are not nearer to the vantage point. The nearest vantage target from a given point is close to the distance, they identify the neighbor points on the user data. Therefore, all the data points are considered for determining the vantage point with better efficiency.

1. Advantages of Vantage Point

The indexing data is directly constructed based on the distance as a substitute of inferring multidimensional points for domain. Thus, it avoids pre-processing stage. Updating a VP tree is relatively easy compared to the fast-map approach. For fast maps, after inserting or deleting data, there will come a time when fast-map will have to rescan itself. That takes up too much time and it is unclear to know when the rescanning will start. Distance based methods are more flexible and it is able to index objects that are represented as feature vectors of a fixed number of dimensions.

II. Literature Survey

1. Figurative thought, colour categorization and vantage construal in scientific language

An effective tool presents the description for colour categorization and explanation in [1] with scientific language and provides more concretely in marine biology. The scientists rely heavily on colour to conceptualise and lexicalize sea organisms are based on corpus data through figurative thought. Both metaphor and metonymy play a fundamental responsibility in categorization and explanation of color. The analysis of the terms explains the different vantages to assign sea organisms by means of conventional (e.g. hue, saturation, and brightness) and unconventional (illumination) colour dimensions. They provide the combinations that involving colour and an uncommon source for metaphorical extension, such as texture. Colour categorization through metaphor is also shown to give rise to both inter- and intra lingual differences at the cognitive and lexical level.

The location of vantage point tree is used to analysis the corpus data with the assumptions and claims from cognitive linguistics. These corpus data include general principles from cognitive psychology on colour description. Finally, they provides the colour categorization gives approaching into concept formation in scientific language through metaphor.

2. Disparate vantage points: Race, gender, county context, and attitudes about harsh punishments in the US

This paper [2] uses data from the General Social Survey, the FBI Uniform Crime Reporting Statistics, and other sources to detect the relationship using vantage point. They identify the relationship among the difference assumption, competition, sexual category, county arrest rates and attitudes about punishment. After the identification, they supported for the claim and the country situation may moderate race/gender groups view punitiveness. Some of the support for the claim is important for conceptualization of

“punitiveness”. These findings identify associations rather than causal relationships, and they point to a need for more research to unpack context. They may provide with shape and attitudes about punishment by influencing our relationship to the criminal justice system.

3. Geolocation Databases of Internet Hosts based on Machine Learning

Geolocation database method [3] is based on machine learning approach for improving the accuracy and robustness of Geolocation. The optimization of geolocation in internet hosts is developed with both routing policy and machine learning. The proposed method performs the database with internet hosts after the stage of optimization approach in geolocation method. The constraint based geolocation method and topology based geolocation methods are implemented for the construction of optimized objects.

Ground truth node based confirmation provides various internet datacenters for collection of ground nodes to improve accuracy of geolocation database. They consist of correct information about geographic location of the ground truth nodes. Cross validation method produces limited ground truth nodes for enhancing the accuracy.

4. Internet exchange points with software-defined networking

Inter-domain routing [4] modification is heavy for Internet Exchange Points that are economically reducing path latencies and exchanging over the large amount of data. Traffic engineering controls the traffic flow of data in network operation. Internet exchange point algorithm is designed to calculate optimal routing paths and specific features of routing scheme. Automatic traffic control leads to lesser human error by processing liberated resource. Distributed distance of services provides attacks on internet with different software-defined networking (SDN).

Initially, black holing is considered for detecting the malicious traffic occurred nodes with continues set nodes. Next, IP address for internet exchange point is applied to prevent towards dummy medium access control. SDN- enabled black holing that specify fine grained operation and process for reducing the risk of undefined configuration. Application programming interface is implemented after the detection of attacks and monitor the flow of internet data.

5. Latency based Any cast Geolocation

A protocol-agnostic technique [5] was proposed for anycast replicas discovery and geolocation. In proposed approach, city level geolocalization of anycast instance is achieved with various feature identification. The geolocalization achieves anycast instance known as vantage point and determine optimization and classification problem. Latency is used an input for optimization problems and channel side information is referred for classification problem.

Geographic location is provided with the measurement of latency according to increasing the size of vantage points. The decision on geolocalization is considered with different number of instance that transfers the original data. A lightweight iGreedy with protocol agnostic anycast replicas provides details and geolocation and a suite of datasets and software is presented to facilitate the development, validation new techniques.

6. Text Mining for Adverse Drug Events

Text mining consists of both structured and unstructured data in network and they use the computational process. With the help of computational process, a large amount of information is extracted from unstructured text. Hence, Adverse Drug Event was presented in [6] for improving the pharmacovigilance that controls the underutilized data sources. There are many data sources for providing pharmacovigilance in free-text formats. The applications of text mining for drug events with several data sources are such as social media, web search, biomedical and labeling. Pharmacovigilance is presented for the analysis of trials and unstructured report that are performed by the domain experts. Therefore, the requirement of drug related information and effective efforts with an accurate knowledge is defined as the adverse drug events.

7. Car dependent practices from a sequence pattern mining

A car dependent practice [7] identifies the car confidence in transport research by using micro-social, macro approach and a meso level approach. A micro-social considerate the attribute of individuals whereas macro approach provides attribute of societies or local areas. At the theoretical stage, the social practices are designed to put ahead the perception of ‘car dependent practices’. At the experimental level, it shows the application of sequence pattern mining techniques to allow the identification of car and mobility intensive activities based on time usage.

The dynamic relationships involving car dependent practices and car ownership is investigated using proposed approaches. The commitments towards the certain activities are achieved by relating the ownership with dominant factors. The relationships between the macro-, micro- and meso-levels of car dependence are dynamic and self-reinforcing interactions.

8. Mining Spatial-Temporal Patterns and Structural Sparsity

A data-driven-based robust human motion de-noising approach [8] were designed for removing noise and outliers. The proposed approach is developed with spatial-temporal patterns mining and the structural sparsity embedded in motion data. Initially, a fine-grained partlet model is used as feature representations by replacing the frequently used entire pose. Feature representation develops the abundant local body element position and movement similarities. Next, a multiple compact and representative motion dictionaries develop a robust dictionary learning algorithm from the training data in parallel. Finally, a robust structured sparse coding is created with distribution information and the temporal smoothness of human motion.

9. Research Analysis on Big Data

Big data is differentiating by its size and complexity in capturing, managing and processing within a reasonable period of time. National science foundation and national science foundation of china develops a key requirement for offers a project terms. Natural Language Processing (NLP) is applied by extracting the phases with support of text mining. Initially, Big Data performs removal of common or basic instance and fuzzy word matching technique is used for combining the terms. Here, fuzzy words are combined with similar structures based on pattern commonality. They remove both extreme words and occurrence on single records. Finally, low frequency phrases are combined with high

frequency phrases in the same records.

The multiple disciplinary programs are developed and extended with support agency for understanding the improvement of research in a rapidly emerging field. Big Data is predictably considered as a part of information sciences and it is processed based on the mining potential information. The information's are obtained either in the form of structured, semi-structured and/or unstructured data. Big Data also has a very close relationship with information and computer technologies, including data collection, storage, processing, and analysis/ visualization

10. Mining Temporal Patterns in Time Interval-Based Data

End point representation and end time representation are developed using mining temporal patterns [10]. The proposed representation simplifies the processing of complex relationships between occurrence intervals. Temporal pattern, occurrence-probabilistic temporal pattern and duration-probabilistic temporal pattern are the different types of interval-based patterns on the proposed representations. These representations are designed with the help of Temporal Pattern Miner (TPMiner) and Probabilistic Temporal Pattern Miner (P-TPMiner) algorithm.

The temporal pattern exposes the relationship between the intervals for a time interval-based database to help users recognize with the actual activities. Next, the occurrence probabilistic temporal pattern expresses the correlation among intervals with the addition of information on the time of the occurrence. It indicates the occurrence probability, enabling users to predict potential activities. Finally, the duration-probabilistic temporal pattern expresses the correlation among intervals and the distribution of different interval lengths that gives users discovered patterns.

11. Human Behavioral Pattern Mining from Multivariate Temporal Data

The frequent behavioral pattern detection approach [1] is generic and not dependent on a single source of information and the scalable approach from multiple sensor information. The scalable algorithm is introduced to identify patterns of human daily behaviors and the patterns are extracted from multivariate temporal data is collected using smart phones. The sensor is developed on devices for identifying the frequent behavioral patterns with a temporal granularity. The patterns attained from the sensor are used for both end-users and other users who provide services based on this information. Therefore, the data on the device is analyzed by the users to control the data, thus it removes the network cost and privacy on the data.

12. A Measurement Experimentation Platform at the Internet's Edge

Dasu, a measurement at the internet edge [12] was addressed with the help of key algorithm and using a larger scale. The edge network distinguishes the complexity increase and sufficient resource is established to provide a host platform vantage points. The development of Dasu approach illustrate the occurrences of node's that collectively offer broad network coverage, suitably high accessibility and fine grained synchronization. The single perception Dasu function carries to internet measurement.

A distributed collection of customers and a set of management services are arranged to form a Dasu function. Broadband characteristic and network management is provided with coverage area and it is given as Dasu's data collections. The frequency and

duration measurement is included in configuration settings during initialization process. Then, co-ordination service is assigned on measured task for submitting the updates and retrieve limits for different task.

13. Micro aggregation technique for large numerical data volumes

Micro aggregation technique [13] shows the reduced computational needs with two steps sorting the dataset with regard to vantage point and set of k-nearest neighbor's searches. Based on k-nearest neighbor's search technique, an efficient quality preserving micro aggregation algorithm is proposed. Data structures for nearest neighbor computation can be approximately classified in pivot-based or clustering-oriented. Pivot-based structures classify points according to their distance to one or several reference points and it also called vantage or pivot point. Similarly, Clustering-oriented method can be further divided into flat or tree-like structures.

Vantage point micro aggregation process uses a vantage point to reduce the amount of computational effort. It is commonly used for kNN search problems. Hence, the obtained distance points are positioned to the global centroid.

14. Multi vantage Point Systems on the Tele operation of Spacecraft Docking

The ThirdEye system [14] was developed to maintain the complex relative navigation task in the human operator of final approach and docking. A tele operation system enhances a robotic system with the sensitivity, cognition, and pronouncement ability of a human operator. This system provides additional accomplishment and more flexible tele robotic system. A flexible camera vantage point is presented for the operation that is situated in the relevant space approximately in the region between the chaser and target spacecraft.

The vantage point is attained with the help of primary and secondary camera views. A spontaneous graphical user interface is included with an attitude head-up display and a trajectory prediction display. Final approach and docking with the presentation of the tele operation system estimate the effects of ThirdEye system. The operator task load is increased during high-difficulty circumstances due to the dual task of tele operating the camera arm and the spacecraft in tandem. This leads to a minor increase in failure rate in these scenarios.

15. Mining human mobility patterns

The mining human mobility patterns [15] were developed for recognize the people trajectories from geo-referenced posts. Initially, geo-tagged data is collected and detect the data location using clustering approach. Next, Sequential pattern mining is used to identify the locations and frequent travel routes for the collection of data. Finally, spatial temporal features are extracted with data movement and duration at each location from all the frequent patterns.

The relevant semantic locations from geo-tagged posts are detected by proposing an efficient approach. This proposed approach identifies semantics during supervised approach that is based on spatial-temporal features. The mobility pattern mining is prepared for expressing the route where the data's are located. A fine grain analysis of unpredictable and irregular information is performed from geo-tagged tweets.

III. Investigation Study On Vantage Point Indexing Search In Multiple Data Pattern Mining

Vantage Point Indexing Search in pattern mining support multiple database for quicker search in relational databases. They provide higher ranking queries and improve the search efficiency as per user interest. The multiple data patterns on various domains that contain unmatched patterns in online service. Data mining approach search the relationship between internet users and the association rule determines frequently used patterns. Clustering and classification process is involved in data mining for separating the group of search and extracting the data patterns. The multiple databases are searched at once to ensure broadest coverage of relevant material and remove duplicate patterns across databases.

Relational query processors and Indexed algorithm generates vantage point indexed join queries that produces a result of individual order on a user-specified function. During the operation of client query search, indexed query search produces optimized multi-join results. Index-join algorithm and non-blocking operators are the basic query operators based on the variation of join flow and it is integrated into pipelined execution.

Multiple data pattern mining has the collection of authoritative information among multiple databases that are unparalleled in online service. Search similarity is based on query distance between relational data points and referencing vantage points. Indexing technique for vantage point is given by either discrete distance function or continuous distance function. The function which returns a set of value is known as discrete distance function and infinite set of values is known as continuous distance function.

1. System Architecture

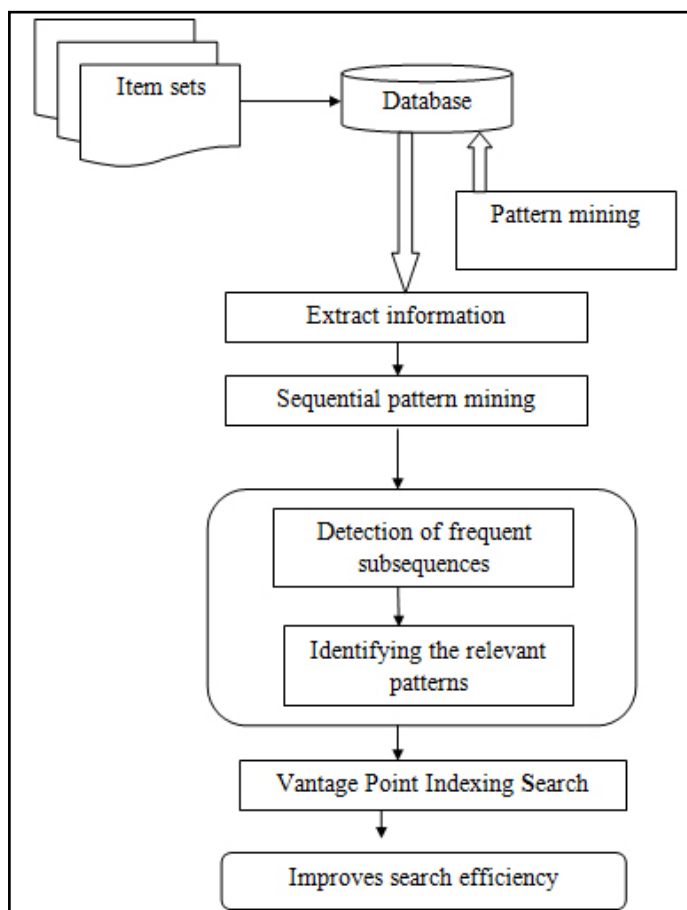


Fig. 3.1 : Architecture diagram of Vantage Point Indexing Search in Multiple Data Pattern Mining

Above figure 3.1 explains about the system architecture of vantage point indexing search that shows improved search efficiency of user internet. Pattern mining algorithm checks the sequence and filters the data with patterns. After that, sequential pattern mining identifies the relevant patterns according to their data values. The multiple databases are searching through the vantage point for discovering the nearest neighbor point that provides better data collection in data mining. The process of Vantage Point Indexing Search in Multiple Data Pattern Mining is separated into following steps.

- a) Pattern Mining
- b) Sequential Pattern Mining
- c) Searching through Vantage Point
- d) Vantage Point Indexing Search

a) Pattern Mining

Frequent patterns mining is used for identifying the relationship among the item sets, subsequences, or substructures in a database. A data set is appeared with frequency that is less than a user-specified threshold. The Apriori method creates frequent pattern with all item sets against transaction database. Many variations of frequent pattern mining such as interesting patterns, negative patterns, constrained pattern mining, or compressed patterns are produced in pattern mining approach.

Clustering and classification is included in the application of frequent pattern mining approach. Pattern mining is efficient from the large database or non trivial extracts the useful information from large set of database. The Apriori algorithm based on association rule mining is most useful algorithm to discover frequent pattern.

b) Sequential Pattern Mining

Sequential pattern mining is given by a set of sequences and support threshold that finds the complete set of frequent subsequences. Sequential mining consists of large number of possible sequential patterns that are hidden in databases. A mining algorithm is developed to complete the set of patterns and satisfying the minimum support (frequency) threshold. The number of database is integrated with user-specific constraints to provide an efficient and scalable pattern on user data.

Sequential pattern mining is one the major data mining approach for the detection of frequent subsequences in a sequence database to identifying the relevant patterns. A major technique for sequential pattern mining is pattern growth. Some of the applications of sequential pattern mining weblog analysis, financial market prediction, bioinformatics, medical treatments, natural disasters, telephone calling patterns, DNA sequences and gene structures etc...

c) Searching through Vantage Point

A vantage-point consists of series process that provides separate data in a pattern mining with space by choosing a position in the user space and partitioning the data points. Here, the data points are separated into two divisions such that, points that are nearer to the vantage point than a threshold, and those points that are not. A vantage-point tree can be used to find the nearest neighbor of a point with the development of search algorithm. A threshold distance is considered from the vantage point and based on the distance nearest neighbor point is identified. Relational query process uses individual input data for producing multi-join results during client query search operation. Finally, a vantage point

is presented to integrate new join operators in practical query processors for identifying a significant number of relevant data patterns.

d) Vantage Point Indexing Search

Vantage Point indexing search results from an extensive variety of structured text databases from most on-line sources. The user data's can also be imported from Excel and Access. With the additions of some other data, databases frequently contain unwanted variation like names of companies or people and they are wanted to clean. Vantage Point provides tools to automatically and interactively clean and manipulate the used data. Hence, they enabling harmonize and customize results and produces the output results as clean and brief. Indexing search provides time saving and more efficient by cleaning the unwanted frequent data patterns. Vantage Point provides the ability analysis to interact with data at a very granular level and view its information from a variety of vantage points. From co occurrence and similarity mapping to natural language processing, they determine regular expression. Vantage Point gives a robust set of tools to explore indexing search results to facilitate turn information into actionable knowledge. Therefore, vantage point indexing search able to communicate data insight with others and produces the effective utilization of information. Finally, search efficiency on user internet is improved with better results.

Input : Number data patterns in relational database
Output: Improved search efficiency
Begin
Step 1: Perform pattern mining for identifying the data and filtered according to the patterns
Step 2: For each data patterns
Step 3: Perform Sequential pattern mining for representing the relevant patterns among multiple data patterns
Step 4: Searching the user data through Vantage Point
Step 5: Identify the nearest neighbor point
Step 5: Perform vantage point indexing search
Step 6: Provide improved search efficiency
Step 7: End

IV. Results and Discussion

The performance analysis is carried out in this paper with the metrics of True Positive Rate, Search Time and Search efficiency. The performance metric of Vantage Point Indexing Search in Multiple Data Pattern Mining method is evaluated and analyzes the values in java environment. Following metrics are used for experimental purposes.

- True Positive Rate
- Search Time
- Search efficiency

1. True Positive Rate

The true positive rate is defined as the measure of relevant patterns provided by the users on social internets. It avoids the un-matched patterns successively according to different number of data patterns that provides from internet. True positive rate is defined as the ratio of total number of data patterns to the relevant patterns from data users. It is measured in terms of percentage (%).

Table 4.1 : Tabulation of True Positive Rate (%)

Number of Data Patterns	True Positive Rate (%)	
	Existing SMKQ	Proposed VPIS
10	63.2	69.2
20	65.74	72.56
30	67.95	74.69
40	69.21	76.58
50	71.3	78.36

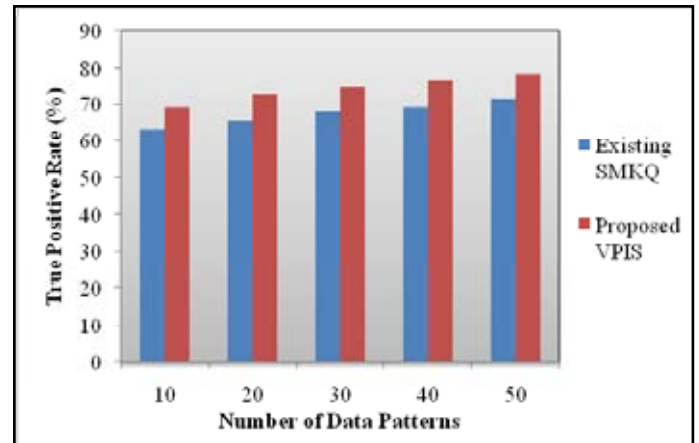


Fig. 4.1 : Measure of True Positive Rate (%)

Above figure demonstrate the measure of true positive rate with respect to different number of data patterns that ranges from 10 to 50. The figure shows the comparison made between existing Single and Multiple Keyword Querying (SMKQ) and Vantage Point Indexing Search (VPIS). When the number of data patterns are increased, true positive rate is also get increased. Therefore, Vantage Point Indexing Search approach achieves higher true positive rate by identifying the relevant patterns on internet. As a result, true positive rate is improved by 10% when compared to the existing Single and Multiple Keyword Querying approach.

2. Search Time

Search time is defined as the measure of time taken for identifying the relevant patterns among the multiple data patterns provided by data mining approach. It is measured in terms of milliseconds (ms). The lower the search time, the proposed approach is said to be more efficient.

Table 4.2 : Tabulation of Search Time (ms)

Number of Data Patterns	Search Time (ms)	
	Existing SMKQ	Proposed VPIS
10	1.23	0.84
20	1.56	1.18
30	1.85	1.54
40	2.23	1.87
50	2.54	2.24

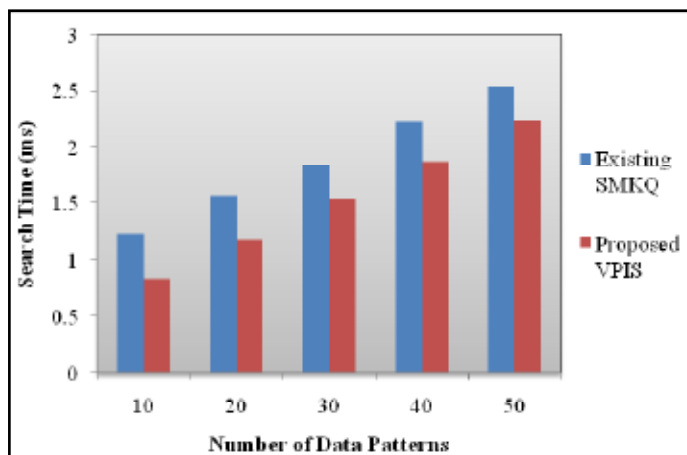


Fig. 4.2 : Measure of Search Time (ms)

Above figure shows the measure of search time with respect to different number of data patterns that ranges from 10 to 50. The figure shows the comparison made between existing Single and Multiple Keyword Querying (SMKQ) and Vantage Point Indexing Search (VPIS). When the number of data patterns are increased, search time is also get increased. Therefore, Vantage Point Indexing Search approach achieves minimum search time for identifying the relevant patterns with the help of nearest neighbor points. As a result, search time is reduced by 20% when compared to the existing Single and Multiple Keyword Querying approach.

3. Search Efficiency

The search efficiency is defined as the measure of relevant data patterns and proves to be effective on the query in multi-Relational Join database. The Vantage Point Indexing Search is used to improve the data patterns and achieve effective mining on data patterns. Search efficiency is measured in terms of percentage (%).

Table 4.3 : Tabulation of Search Efficiency (%)

Number of Data Patterns	Search Efficiency (%)	
	Existing SMKQ	Proposed VPIS
10	74.2	79.64
20	75.96	80.3
30	78.11	82.47
40	79.63	84.6
50	81.2	86.38

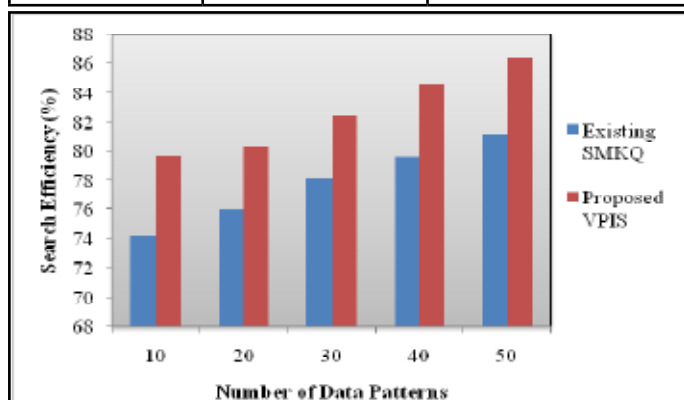


Fig. 4.3 : Measure of Search Efficiency (%)

Above figure shows the measure of search efficiency with respect to different number of data patterns that ranges from 10 to 50. The figure shows the comparison made between existing Single and Multiple Keyword Querying (SMKQ) and Vantage Point Indexing Search (VPIS). When the number of data patterns are increased, search efficiency is also get increased. Therefore, Vantage Point Indexing Search approach achieves minimum search efficiency for identifying the relevant patterns with the help of nearest neighbor points. As a result, search efficiency is reduced by 6% when compared to the existing Single and Multiple Keyword Querying approach.

V. Conclusion and Future Work

Vantage Point Indexing Search in Multiple Data Pattern Mining is developed for quicker search in relational databases. The multiple databases are searched to ensure broadest coverage of relevant material and remove duplicate patterns across databases. Initially, frequent pattern mining is used for identifying the relationship among the item sets in a database. Clustering and classification is included in the application of frequent pattern mining approach. Then, Sequential pattern mining approach detects the frequent subsequences in a sequence database for identifying the relevant patterns. Finally, Vantage Point indexing search provides higher ranking queries and improve the search efficiency as per user interest. The future work includes the some other web page based documents for improving the search efficiency through multiple database.

Reference

- [1] Jose Manuel Urema Gomez-Moreno, "Figurative thought, colour categorization and vantage construal in scientific language", Elsevier, Language Sciences, Volume 41, Part B, January 2014, Pages 197–211.
- [2] Erin Carll, "Disparate vantage points: Race, gender, county context, and attitudes about harsh punishments in the US", Social Science Research, Elsevier 2016, pages 1-17.
- [3] Ting Wang, Ke Xu, Junde Song, and Meina Song, "An Optimization Method for the Geolocation Databases of Internet Hosts Based on Machine Learning", Hindawi Publishing Corporation, Mathematical Problems in Engineering, Volume 2015, pages 17.
- [4] Marco Chiesa, Christoph Dietzel, Gianni Antichi, Marc Bruyère, Ignacio Castro, Mitch Gusat, Thomas King, Andrew W. Moore, Thanh Dang Nguyen, Philippe Owezarski, Steve Uhlig, and Marco Canini, "Inter-Domain Networking Innovation on Steroids: Empowering IXPs with SDN Capabilities" IEEE Communications Magazine, Volume: 54, Issue: 10, October 2016, Pages 102-108.
- [5] Danilo Cicalese, Diana Zeaiter Joumblatt, Dario Rossi, Marc-Olivier Buob, Jordan Auge and Timur Friedman, "Latency-Based Anycast Geolocation: Algorithms, Software and Datasets", IEEE Journal on Selected Areas in Communications, Volume: 34, Issue: 6, June 2016, Page(s): 1889 – 1903.
- [6] Rave Harpaz, Alison Callahan, Suzanne Tamang, Yen Low, David Odgers, Sam Finlayson, Kenneth Jung, Paea LePendu and Nigam H. Shah, "Text mining for adverse drug events: the promise, challenges, and state of the art", Springer 2014, Volume 37, Issue 10, Pages 777-790.
- [7] Giulio Mattioli, Jillian Anable and Katerina Vrotsou, "Car dependent practices: Findings from a sequence pattern

- mining study of UK time use data*”, *Transportation Research Part A: Policy and Practice*, Volume 89, Elsevier July 2016, Pages 56–72.
- [8] Yinfu Feng, Mingming Ji, Jun Xiao, Xiaosong Yang, Jian J. Zhang, Yueting Zhuang, and Xuelong Li, Fellow, IEEE, “Mining Spatial-Temporal Patterns and Structural Sparsity for Human Motion Data Denoising”, *IEEE Transactions on Cybernetics*, Volume 45, Issue 12, December 2015, Pages 2693-2706.
- [9] Ying Huang, Yi Zhang, Jan Youtie, Alan L. Porter and Xuefeng Wang, “How Does National Scientific Funding Support Emerging Interdisciplinary Research: A Comparison Study of Big Data Research in the US and China” *PLoS ONE* 2016, Volume 11, Issue 5, pages 1-20.
- [10] Yi-Cheng Chen, Wen-Chih Peng and Suh-Yin Lee, “Mining Temporal Patterns in Time Interval-Based Data”, *IEEE Transactions on Knowledge and Data Engineering*, Volume: 27, Issue: 12, December 2015, Page(s): 3318 – 3331.
- [11] Reza Rawassizadeh, Elaheh Momeni, Chelsea Dobbins, Joobin Gharibshah and Michael Pazzani, “Scalable Daily Human Behavioral Pattern Mining from Multivariate Temporal Data”, *IEEE Transactions on Knowledge and Data Engineering*, Volume. 28, Issue. 11, November 2016, Pages: 3098 3112.
- [12] Mario A. Sanchez, John S. Otto, Zachary S. Bischof, David R. Choffnes, Fabian E. Bustamante, Balachander Krishnamurthy and Walter Willinger, “A Measurement Experimentation Platform at the Internet’s Edge”, *IEEE/ACM Transactions on Networking*, Year: 2015, Volume: 23, Issue: 6, Pages: 1944 – 1958.
- [13] Marc Sole, Victor Muntés-Mulero and Jordi Nin, “Efficient micro aggregation techniques for large numerical data volumes”, *International Journal of Information Security - Special Issue: Supervisory control and data acquisition (SCADA) archive*, Volume 11 Issue 4, August 2012, Pages 253-267.
- [14] Markus Wilde, Zarrin K. Chua, and Andreas Fleischner, “Effects of Multivantage Point Systems on the Teleoperation of Spacecraft Docking”, *IEEE Transactions on Human-Machine Systems*, Volume. 44, Issue. 2, April 2014, Pages 200-210.
- [15] Carmela Comitoa, Deborah Falcone and Domenico Talia, “Mining human mobility patterns from social geo-tagged data”, *Pervasive and Mobile Computing*, Elsevier, Volume 33, December 2016, Pages 91–107.