

ScholarWorks@GSU

SenMinCom: Pervasive Distributed Dynamic Sensor Data Mining for Effective Commerce

Authors	Hiremath, Naveen
Citation	Hiremath, Naveen (2008). "SenMinCom: Pervasive Distributed Dynamic Sensor Data Mining for Effective Commerce." Thesis, Georgia State University. https://doi.org/10.57709/1059400
DOI	https://doi.org/10.57709/1059400
Rights	<p>I hereby certify that, if appropriate, I have obtained and attached hereto a written permission statement from the owner(s) of each third party copyrighted matter to be included in my thesis, dissertation, or project report, allowing distribution as specified below. I certify that the version I submitted is the same as that approved by my advisory committee. I hereby grant to Georgia State University or its agents the non-exclusive license to archive and make accessible, under the conditions specified below, my thesis, dissertation, or project report in whole or in part in all forms of media, now or hereafter known. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also retain the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report.</p>
Download date	2026-04-15 21:59:43
Link to Item	https://hdl.handle.net/20.500.14694/4160

**SENMINCOM: PERVASIVE DISTRIBUTED DYNAMIC SENSOR DATA
MINING FOR EFFECTIVE COMMERCE**

by

NAVEEN HIREMATH

Under the Direction of Dr. Yanqing Zhang

ABSTRACT

In last few years, the use of wireless sensor networks and cell phones has become ubiquitous; fusing these technologies in the field of business will open up new possibilities. To fill this lacuna, I propose a novel idea where the combination of these will facilitate companies to receive feedback on their products and services. System's unobtrusive sensors will not only collect shopping, mobile usage data from consumers but will also make effective use of this information to increase revenue, cut costs, etc.; the use of mobile agent based data mining allows analyzing the data from different dimensions, categorizing it on factors such as product positioning, promotion of goods, etc. as in the case of a shopping store. Additionally, because of the dynamic mining system the companies get on-the-scene recommendation of products rather than off-the-scene. In this thesis, a novel distributed pervasive mining system is proposed to get dynamic shopping information and mobile device usage of the customers.

INDEX WORDS: Wireless Sensor Network, Dynamic Data mining, Distributed Pervasive mining, Mobile agents

**SENMINCOM: PERVASIVE DISTRIBUTED DYNAMIC SENSOR DATA
MINING FOR EFFECTIVE COMMERCE**

by

NAVEEN HIREMATH

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
in the College of Arts and Sciences
Georgia State University

2008

Copyright by
Naveen Hiremath
2008

**SENMINCOM: PERVASIVE DISTRIBUTED DYNAMIC SENSOR DATA
MINING FOR EFFECTIVE COMMERCE**

by

NAVEEN HIREMATH

Committee Chair: Dr. Yanqing Zhang

Committee: Dr. Rajshekhar Sunderraman

Dr. Ying Zhu

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2008

DEDICATION

To my friends, for their unconditional love and support

ACKNOWLEDGEMENTS

I would like to send special thanks to my thesis advisor Dr. Zhang for his help, direction and support throughout the duration of the thesis. Without his enthusiasm the thesis would not have been in the current stage.

I am very thankful to my thesis committee members Dr. Sunderraman and Dr. Zhu for their invaluable time spent in reviewing my thesis.

I especially want to express my sincere gratitude to Dr. Sunderraman for all his help, advice and support during my graduate studies. I am very grateful to him.

I would like to thank all my friends especially the #128ians for providing support and the occasional welcome distraction while I worked on this thesis.

Special thanks to my dad, mom, and Nandeesh who have patiently been my motivation and moral support throughout my education and career.

Lastly, I would like to thank my late grandfather with whose patience, knowledge, and guidance I have achieved success and have come up so far. Wish you were here to see this day!

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF FIGURES	viii
1. INTRODUCTION	1
1.1. Motivation	1
1.2. Overview of Related Works	3
1.3. Contribution	7
1.4. Organization	8
2. BACKGROUND	9
2.1. Wireless Sensor Network	9
2.2. Mobile Agents	10
2.3. Data Mining	12
2.3.1. Knowledge Discovery in Databases(KDD)	13
2.3.1.1 The KDD Process	14
2.3.2. Concepts	16
2.3.2.1.Data Warehousing	17
2.3.2.2.Predictive vs. Descriptive Data Mining	17
2.3.3. Data Mining Strategies	18
2.3.3.1.Supervised Learning	18
2.3.3.2.Unsupervised Learning	19
2.3.3.3.Market Basket Analysis	19

2.3.3.4. Customer Segmentation	20
2.3.4. Synopsis of Approach	20
3. SYSTEM ARCHITECTURE	22
3.1. Shopper's Model	27
3.2. Mobile Device Usage Model	28
3.3. System Algorithms	29
4. SYSTEM SIMULATION	33
4.1. Shopping Scenario	33
4.1.1. Observations	34
4.2. Mobile Device Usage Scenario	38
4.2.1. Observations	39
5. CONCLUSION & FUTURE WORKS	46
5.1. Conclusion	46
5.2. Future Works	47
REFERENCES	50

LIST OF FIGURES

Figure 1: The KDD Process	15
Figure 2: System Architecture for SenMinCom	23
Figure 3: Schematic representation of a Distributed Dynamic Mining System	25
Figure 4: Popularity of aisles	36
Figure 5: Popularity of aisles at different time intervals	37
Figure 6: Lower v/s higher purchasing value	37
Figure 7: Customers checkout v/s shopping share	38
Figure 8: Shopping v/s checkout pattern of a customer	38
Figure 9: Mobile Devices sighted at GSU Plaza	40
Figure 10: Mobile Devices sighted at GSU Student Center	40
Figure 11: Most Popular Mobile Devices sighted at GSU	41
Figure 12: Google Maps TM plot of Mobile Devices sighted at GSU Plaza	42
Figure 13: Google Maps TM plot of Mobile Devices sighted at GSU Student Center	42
Figure 14: Google Maps TM plot of most popular Mobile Devices sighted at GSU	43
Figure 15: Volume usage of LG Mobile Devices sighted at GSU	44
Figure 16: Volume usage of Motorola Mobile Devices sighted at GSU	44
Figure 17: Volume usage of Motorola Mobile Devices sighted at GSU	45
Figure 18: Volume usage of New Mobile Devices sighted at GSU	46

1. INTRODUCTION

In the last few years, wireless sensor networks have attracted a significant attention among users, researchers, vendors and the like. Due to recent advances in technology, the development and deployment of low-cost, low-power, multifunctional tiny sensor nodes operating in an unattended mode to sense real-world events like traffic monitoring, habitat monitoring, fire detection, etc. has become rapid. Additionally, the use of cell phone or mobile device has become ubiquitous, and mobile commerce is speeding. With dropping prices and increasing utility, it is almost a foregone conclusion that not too far into the future, all will have a cell phone, quite possibly right into their clothing. Cell phones are not just communications devices sparking new modalities of interaction between people; they are also particularly useful computers that fit in your pocket, are always with you, and nearly always on providing entertainment, distance education, proactive service management, etc. Several real-world applications are being designed and developed that are taking advantage of these new pervasive technologies. However, the vast potential of sensor networks and cellular phones for commerce has not been explored completely.

1.1 Motivation

In the current market scenario, a customer's continuing business with a store can no longer be guaranteed. Hence, the stores have to understand their customers better, and should quickly respond to their needs. For example, consider the case of a customer who is visiting a store. He goes around the aisles looking for products but the items that he finally checks out may not

include all the items that he had shown interest in - the user not buying an item hinge on several factors. There is no such dynamic real time system that can automatically track these factors (at least some) and use it for the benefit of the store and the consumer. Additionally, the time frame in which these responses need to be made should be very less. For example, customers receive directed product advertisements and discount offers to their email id which in most case doesn't get their attention or gets regarded as spam; in that case the whole purpose of luring the customer goes futile. On the other hand, Pareto Principle indicates the major part of revenue for a company comes from a small fraction of customers. That is what so called 80/20 rule. Hence, it is no longer advisable to wait until the signs of customer dissatisfaction are obvious before action must be taken. To succeed, stores must be proactive and anticipate what a customer needs.

On a similar basis, recent advancements in development of short range (Bluetooth) and wide area (GPRS, UMTS) wireless technologies have made possible development of new generation of innovative applications. However, these applications are limited to the realm of social networks [35], epidemic modeling [32], tsunami warning [33], etc. that tend to capitalize on location information. The SpaceMe service from GypSii [34] for instance will show users where friends and other members are in real time [35]. Accordingly, why can't we have a service for the cellular phone companies to monitor their product popularity in the market?

This thesis proposes an effective way through which a cell phone can be used for business i.e. a dynamic mining system for shopping stores where the power of wireless sensor networks and computational intelligence techniques like data mining will provide a means to know the customers shopping pattern real-time unlike the current methods that discover the shopping patterns non-dynamically or offline. Additionally, during checkout the current shopping pattern of a customer can be compared with past purchase pattern in order to facilitate

evaluation and adjustment of promotional plans (aisles or products that the customer visited to the products eventually bought). This distinguishing key feature will help the store to know about a customer's interest and can offer him the product at a discounted price, before a customer exits the store; contrast this to the existing methods where the customer tend to get discount offers or coupons offline [52]. A second usage of the proposed system will be to find out the mobile usage model. Instead of companies relying on statistical agencies to release their market shares every quarter or so, now they can have real time data ready to use; data about their market share, which will lead to taking faster decisions on how and where to market their models. The cell phone will be wirelessly detected via Bluetooth by the Bluetooth enabled sensors. Bluetooth has been chosen because of its pervasiveness, availability on today's cell phones and relatively low power consumption. However, any other short-range wireless communication protocol can be used. For a given location the service will get the model the consumer is using and thus helps the company to find active cell phone users on a local area basis.

1.2 Overview of Related Works

The use of sensor networks and associated cell phone have been proposed and being applied in various contexts. The following description is by no means exhaustive.

Habitat and environmental monitoring (e.g. volcanic eruptions [5]) [1-7] provide the use of sensor nodes which will be of enormous potential (e.g. economical) benefits for scientific communities and society as a whole, over the traditional invasive methods of monitoring. [8] describes a scenario of how sensor nodes can be used to create a smart environment (self-organizing, adaptive systems). The sensor nodes can be embedded into appliances, etc., where

they communicate with each other and the room server. NTT DoCoMo's wellness phoneTM took things to new heights for the physique-obsessed; like a sensor system that Nike Inc. and Apple Inc. developed for the iPod NanoTM, the handset keeps track of runs, letting users set targets and record time, distance and calories burned; while listening to music through head phones. It also features a mini body fat calculator and a sensor at the top of the phone takes pulse from your fingertip [46]. There has been no mention of how sensors can be deployed and effectively used for getting customers' shopping pattern dynamically.

A sequence is a set of ordered elements. Times series is a sequence data measured strictly against the time dimension. In general cases of sequence data where time is not necessarily a useful static reference dimension, the time series analysis methods, such as the ARMA method, are not particularly useful for sequence data analysis. According to the associate distance measure (ADM) method, a sequence is a vector and the difference between two sequences is measured by the Euclidean distance between the two vectors. In sequence alignment method (SAM), a sequence is a set of elements arranged in a certain order. Compared with the ADM method, the SAM method takes the order of the elements into account in measuring the distance between the sequences. Also, the time measure in the sequence is missing in the SAM method. In summary, there have been several popular techniques for sequence data analysis that use a single time sensitive variable but has their limitations in general business sequence pattern analysis. Market basket analysis aims to discover interesting relationships between retail products in order to help retailers in identifying cross-sale opportunities. Although there have been applications of SAM and ADM methods for mining Web navigation sequence patterns [17] and customer segmentation [18], SAM and ADM don't represent the time measure explicitly.

The results of the above methods can be used for effective offline analysis of the business; a method for analyzing the real time shopping behavior of a store has not yet been dealt.

A simple approach to the analysis of sensor network data is the use of a centralized architecture where a central server maintains a database of readings from all the sensors. The whole analysis effort is localized in the server, whose mission is to extract the high-level information expected to be returned by the monitoring system. The limitation of this approach is strikingly evident: the number of messages sent as well as the number of variables of the data mining task is too large to be managed efficiently. It has been suggested in literature that alternative architectures are to be preferred in applications where neighboring sensors are likely to have correlated readings [30]. This is the case of aggregating systems which according to the definition of [19] are systems where the data obtained from different source nodes can be aggregated before being transmitted along the network. Once data gets aggregated mining can be employed at the aggregator node. In order for the mined results to be relevant, mining has to take place on dynamic data rather than on a snapshot of a data and mobile agents make this task noteworthy [20]. Mobile agents are found to be particularly useful for data fusion tasks in distributed wireless sensor networks [23]. The use of a mobile agent will lead to a better utilization of sensor nodes, and network bandwidth along with providing effective results from the live data being captured.

Rasid et al. present in [36] a telemedicine processor for mobile phones. It allows for transmission of biomedical data to remote locations such as hospitals. However, their work is limited to 1:n relationship (one doctor serves many users). The investigation of building disease spreading models using sensory data gathered by mobile phones from a large group of participants and their approach goes beyond the 1:n relationship to a n:n (many users sever many

users) approach [32]. Another example is of a Tsunami warning system which can register an earthquake and can relay warning messages to the people of the affected area in time [33]. Thus thousands of people can get the warning on cell phones shortly after an earthquake is detected. On the other hand mobile social networking services like [37, 35] make it easier for people to share photos, send invitations, or conduct polls among friends via mobile phone. These services are free and they don't require any special software to be installed. Some services share information about a user's location and help them find friends in their local area [38]. Bluetooth has been used in the real-world applications like ATM SpA and INTELLIBUS, providing wireless diagnostic and preventive maintenance support for the ATM bus fleet [40], for automatic and wireless solution of minimizing the impact on the patient's every-day activities [41, 42] and by Schwan's for route sales drivers throughout the U.S. to wirelessly record sales, issue receipts, and track inventory [43]. HealthGear consists of a set of non-invasive physiological sensors wirelessly connected via Bluetooth to a cell phone which stores, transmits and analyzes the real-time physiological data, and presents it to the user in an intelligible way [44]. These above discussed approaches discuss the usage of sensors, cell phones, cellular/wireless network for the well being, productivity or increasing the social relationships of the people. Advances in sensor technology and personal mobile devices are enabling the development of integrated dynamic sensor system that act as a useful tool. However, there is no such system or approach that deals with commerce or which proposes to get real time mobile cellular usage for companies.

1.3 Contribution

More customers, more products, more competitors, and less time to react means that understanding your customers is now much harder to do. Customers want things that meet their exact needs, not things that sort-of-fit. Successful stores need to react to each and every one of these demands in a timely fashion. Today's marketing strategy has been changed from mass marketing to data-driven one-to-one marketing. One-to-one strategy values customer individual preferences and it works on the base of tailoring marketing to each customer.

This thesis aims at building a system that handles the following multiple criteria that are useful for a store to maintain its competitiveness and customer satisfaction.

- The Right Offer
- To the Right Person
- At the Right Time
- Through the Right Channel [26]

Currently there is no such system that can get the real time product statistics of the shoppers' in a store. Hence, the major contributions towards the realization of this system are

- Develop an architecture to handle the communication between sensor network and cell phone
- An efficient mobile agent based mode of communication
- A novel Distributed Dynamic Data mining method that
 - Discovers shopper's behavior and
 - Mobile cell phone usage of different companies or models

On one hand the thesis presents a method for deploying sensors and how to achieve distributed mining using the concept of mobile agents, analyzing the value of a shopper, and a method of measuring profitability of customers by building a “one-on-one” relationship with each and every customer; gives purchase recommendations, discounts, etc. on the items that a customer evinced his interest while shopping, and thus wishes to prevent customer attrition. While on the other hand it makes use of the existing architecture to build an innovative application that gets real time usage of cell phones for an area. Real time usage of the mobile models will allow a company to streamline their advertisements for that region and help in opening up new business opportunities for the company like engaging consumers on multiple platforms, including the possibility of partnership with local music labels to launch co-branded portals targeted at local music communities.

1.4 Organization

The outline of the thesis is as follows: In chapter 1, other than an introduction to the thesis motivation for building the current system, contributions of the system with novel applications, and other related work are discussed. Then chapter 2 goes on with providing background information on data mining concepts and strategies, properties, advantages for the use of mobile agents, and wireless sensor networks, etc. In chapter 3, a detailed explanation about the proposed system along with its architecture, algorithms, etc. is made. It also mentions a second new application that can be built on the proposed architecture. Chapter 4 deals with System simulation of the shopping system and cellular phone usage. Finally, conclusions and ideas on how to extend the current work is explored in Chapter 5.

2. BACKGROUND

In this chapter, I will give a short overview of wireless sensor network, different approaches of gathering data in a sensor network, the advantages of using a mobile agent for data gathering, data mining; the ingredients that mould this thesis

2.1 Wireless Sensor Network

The Wireless Sensor Network (WSN) is an emerging technology that may greatly facilitate human life by providing ubiquitous sensing, computing, and communication capability, through which people can more closely interact with the environment. WSN is a network of sensor and other supporting nodes [27]. They are quintessentially event-based systems that consist of one or more “sinks” which subscribe to specific data streams by expressing interests or queries. A sensor node can be in four states of communication: transmit, receive, idle, or sleep and two states of monitoring: idle and active. A sensor network’s lifetime can be improved by allowing some sensors to sleep while other sensors are covering the area/target of interest. Other power saving techniques include power controlling by adjusting the sensing/transmitting range of sensor, using the energy efficient routing and data gathering techniques, reducing the amount of data transmitted and avoiding useless activity. The sensors in the network act as “sources” which detect environmental events and push relevant data to the appropriate subscriber sinks. Their tasks tend from receiving information from other nodes to carrying out local computations and transmitting partially processed or raw data from node to node. Their development was originally motivated by military applications such as battlefield surveillance. However, the applications of

WSN have become omnipresent. Most energy-efficient proposals are based on the traditional client/server computing model, where each sensor node sends its sensory data to a back-end processing center or a sink node. Because the link bandwidth of a wireless sensor network is typically much lower than that of a wired network, a sensor network's data traffic may exceed the network capacity, introduces latency along with an increase in power consumption. In addition, the sensors remain passive devices. Hence, it is necessary to have minimum amount of network communication, and maximum of local computation.

2.2 Mobile Agents

Several efforts have been proposed for achieving reliable, flexible, and efficient data collection in WSN's. However, none of the approaches addresses completely all the goals of reliability, flexibility, and efficiency. Multipath approaches focus on reliability. Query propagation approaches provide flexibility but lack reliability. Mobile agent approaches offer reliability and flexibility.

Multipath approaches provide higher reliability than single-path data routing approaches with similar latency. They achieve a higher degree of reliability by having multiple paths for the same source-destination pair. The main difference between multipath approaches is whether packets are redundantly forwarded through all multiple paths simultaneously or only forwarded along non-primary paths if the primary path fails. Another main difference is whether the multiple paths are completely disjoint.

SQL-like queries tend to make a sensor network a large distributed database. This brings more flexibility in terms of re-using the already deployed WSN for different purposes. However, the solution tends to propagate the SQL queries by the use of routing tree rooted at the collection

point i.e. sink. This is a fundamental weakness, since unreachable nodes can isolate an entire subtree.

To solve the problem of overwhelming data traffic [23], reliability and flexibility the mobile agents are used in distributed sensor network. By transmitting the software code, called mobile agent (MA), to sensor nodes, the large amount of sensory data can be reduced or transformed into small data by eliminating redundancy.

Mobile agents are self-aware and autonomous programs with the capability of moving and cloning themselves to other locations. Some of their characteristics are:

- Agents are **autonomous**, that is they act on behalf of the user
- Agents contain some level of **intelligence**, from fixed rules to **learning** engines that allow them to adapt to changes in the environment
- Agents do not only act **reactively**, but sometimes also **proactively**
- Agents have **social ability**, that is they communicate with the user, the system, and other agents as required
- Agents may also **cooperate** with other agents to carry out more complex tasks than they themselves can handle
- Agents may **move** from one system to another to access remote resources or even to meet other agents

Apart from the intrinsic flexibility and load distribution capabilities, mobile agents provide a high degree of robustness. If an agent cannot migrate to a node, the agent detects it and adapts its route so the failure can be bypassed; this avoids data loss. Moreover, a mobile agent can perform in-path aggregation of the data being sensed at the mobile agent's path. For particular

applications, mobile agents can save up to 90 percent of the data transfer time by avoiding the raw data being transferred [23].

The thesis proposes a general approach for visiting all aggregator nodes – for collection purposes, etc. based on mobile agents. The reliability of this approach will be comparable to multipath approaches and its flexibility to query propagation approaches. Moreover, the time it takes to cover the entire large WSN will be drastically reduced, and the energy spent on the collection is efficiently used.

2.3 Data Mining

Data mining, the extraction of hidden predictive information from large databases, derives its name from the similarities between searching for valuable business information in a large database – for example, finding linked products in gigabytes of store scanner data – and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business opportunities by providing these capabilities:

- **Automated prediction of trends and behaviors**

Data mining automates the process of finding predictive information in large databases. A typical example of a predictive problem is targeted marketing. Data mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings.

- **Automated discovery of previously unknown patterns**

Data mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together.

The gained knowledge is either represented as a model or generalization of the mined data. Many different data mining techniques have been developed. A great number of these techniques descend from classical statistics; nevertheless there are newer approaches that use artificial intelligence approaches. This makes prediction of trends possible which would not have been feasible with the traditional statistical methods. In addition, the data needed to conduct data mining is widely available in the age of the Internet and e-commerce.

2.3.1 Knowledge Discovery in Databases (KDD)

Frawley et al. [28] state that “Knowledge discovery is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data.” In order to get this information, we try to find patterns in the given data set. To know if a pattern is valuable, the assessment of its interestingness and certainty is crucial. Patterns that are interesting and certain enough according to the user's measures are called knowledge. The output of a program that discovers such useful patterns is called discovered knowledge.

According to [28], KDD exhibits four main characteristics:

- **High-Level Language:** The discovered knowledge is represented in a language that does not necessarily have to be directly used by humans, but its expression should be comprehensible.

- **Accuracy:** The measure of certainty implies whether the discovered patterns portray the contents of a database properly or not.
- **Interestingness:** Discovered knowledge is considered interesting if it fulfills the predefined biases. By denoting a pattern interesting, we mean that it is novel, potentially useful and the discovery process is nontrivial.
- **Efficiency:** Even for large Datasets, the running time of the algorithm is acceptable and predictable.

Data and patterns are defined in [29]: “Here, data is a set of facts and pattern is an expression in some language describing a subset of the data or a model applicable to that subset. Patterns should be understandable, if not immediately then after some post-processing.” The so-called KDD Process consists of several steps that are in place to achieve the defined goals for knowledge discovery.

2.3.1.1 The KDD Process

The KDD Process is interactive and iterative and requires decisions made by the user. [30] proposes nine basic steps (see Figure 1).

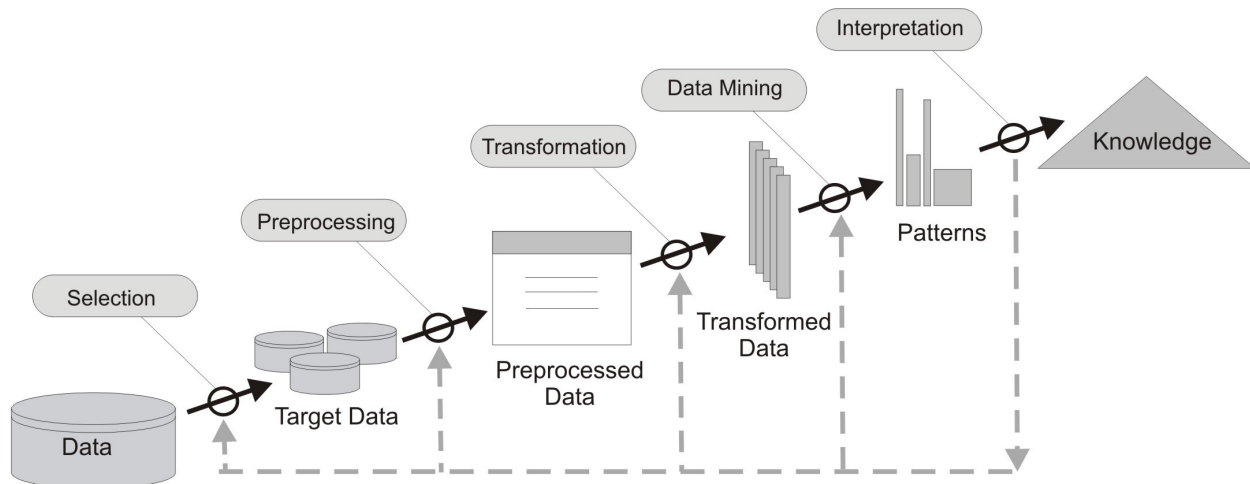


Figure 1: The KDD Process [30]

1. **Data Understanding:** learning the application domain for prior knowledge and goals of the application.
2. **Creating a target data set:** selecting the subset of the data on which the data mining will be performed.
3. **Data cleaning and preprocessing:** removing noise or outliers, developing strategies for handling missing data.
4. **Data reduction:** reduce dimensionality of the data set in order to get rid of data that is unnecessary for completing the mining task and thereby keep the computing time low.
5. **Selecting the data mining method:** the most important task here is to find the method that will best suit the completion of the KDD goals.
6. **Choosing the data mining algorithm:** there are many different data mining algorithms. Deciding on an efficient one to search for patterns in data is critical and includes decisions about appropriate models and parameters.

7. **Data mining:** applying the previously chosen algorithm to the data set and searching for interesting patterns in a particular representational form.
8. Interpreting mined patterns includes the visualization of mined patterns and a possible return to any of the steps 1-7 if the results are unsatisfactory.
9. **Consolidating discovered knowledge:** documenting the results and incorporating them into another system.

A distinction between KDD and data mining is emphasized in [29]. While KDD refers to the overall process of discovering useful knowledge from data, data mining is “a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data.”

2.3.2 Concepts

The last decade brought a huge advance in database technology which leads to a huge amount of data being collected. Hence, we are facing a great chance to make use of this data by extracting previously unknown patterns. Parallel processing constitutes an important technique to realize large-scale data mining applications because they handle a huge amount of data and therefore involve a lot of computation [31].

A main goal of data mining is to provide business with information in order to make predictions for future use. For this reason, data mining emerged in the 80's and made great progress in the 90's. It still is a research field of great interest. As a consequence, many different data mining concepts have been developed.

2.3.2.1 Data Warehousing

Organizations build data warehouses by integrating their different operational databases. Data warehousing is the process of centralized data management and retrieval [31]. The term is relatively new although the concept has been around for years. It represents a vision of installing a central repository of all organizational data relevant to data mining. The goal is to ensure easy access by the user and to allow quick access for data analysis. Data warehousing techniques are already being applied in many companies. Data mining provides the software and techniques to acquire useful information from the data warehouse. The data warehouse provides the company with a memory whereas data mining provides the company with intelligence.

The most important benefit of a data warehouse is to allow data mining in very large databases at a high level of performance and manageability. It integrates operational databases that might be divergent and thereby hard to access allowing much more efficient data mining. Two steps have to be considered for making a data warehouse valuable:

1. Integrate the internal and external data into the data warehouse with all the data needed for mining.
2. Organize and present the information in ways that assists complex decision making.

2.3.2.2 Predictive vs. Descriptive Data Mining

The two high-level goals in data mining can be defined as prediction and description. Prediction uses some fields or variables in the database to predict future values of other interesting variables. Description sets its focus on making the data comprehensible and interpretable. The boundaries between those two goals are not sharp, because some predictive models can be

descriptive meaning that they are understandable and vice versa. Nevertheless, the distinction can help understanding the overall mining goal.

2.3.3 Data Mining Strategies

According to the goal we want to achieve with data mining, there are several data mining strategies to choose from. These strategies can be broadly classified in supervised learning, unsupervised learning, and market basket analysis. Supervised learning is mainly used for prediction. Several input variables are used to build models which predict a specified output variable. Supervised learning methods either allow only one single or several output attributes. Unsupervised learning does not have any output variable but rather tries to find structures in the data by grouping the instances into different classes. The designation of market basket analysis is to find regularities in data in order to explore customer behavior. Results can help retailers design promotions or recommendations. This chapter will provide a closer look at these strategies.

2.3.3.1 Supervised Learning

Supervised learning is used in almost any domain, mainly for the purpose of prediction. It could also be called classification or inductive learning when used in association with machine learning. The goal is to create a function out of a given set of historical training data. This function generates the desired output; for example it is possible to predict whether a customer will buy a certain product or not. To be able to compute the function, we need enough training data to make an accurate prediction. Historically collected data with information about customers

who either bought or did not buy the product after a promotion will enable us to find out which potential customers will react on a promotion campaign. The data about a customer's reaction to the campaign serves as an output variable.

2.3.3.2 Unsupervised Learning

Unlike in supervised learning, we do not want to predict a specific output here, but rather discover unknown structures that exist within a data set. The technique used for unsupervised learning is clustering. This technique orders the instances of a data set into groups with similar attributes and values. These groups of items are called clusters. It is important to notice that instances of one single cluster are similar to each other, whereas instances of different clusters are very diverse from each other. By clusters we mean subsets of the overall data set that is being mined. Clusters are created in the mining process without a priori knowledge of cluster attributes.

2.3.3.3 Market Basket Analysis

Market Basket Analysis could be put under the domain of unsupervised learning, but in fact it is often treated in literature as a parallel topic. The goal is to discover interesting relationships between retail products in order to help retailers in identifying cross-sale opportunities. Association rule mining is the most common approach for performing this task. Such algorithms mainly deal with discovering these items which are frequently purchased together. The name is derived from a person walking through a supermarket throwing all the things to buy in a shopping cart. This "market basket" is then analyzed.

2.3.3.4 Customer Segmentation

Customer segmentation is the practice of dividing a customer base into groups of individuals that are similar in specific ways relevant to marketing, such as age, gender, interests, spending habits, and so on. Segmentation allows companies to target groups effectively, and allocate marketing resources to best effect. One of the key purposes of customer segmentation is customer retention.

Traditional segmentation focuses on identifying customer groups based on demographics and attributes such as attitude and psychological profiles. Value based segmentation on the other hand looks at groups of customers in terms of revenue they generate and the costs of establishing and maintaining relationships with them.

Customer Segmentation includes deciding what data will be collected and how it will be gathered; collecting data and integrating data from various sources; developing methods of data analysis for segmentation; establishing effective communication among relevant business units (such as marketing and customer service) about the segmentation; and implementing applications to effectively deal with the data and respond to the information it provides.

2.3.4 Synopsis of Approach

Segmentation of customers that started with market behavior, lifestyle, and socio-demographics moved to purchase histories. However, data mining techniques beyond the existing methods must be developed to analyze the sequence data where patterns are the major concern and time is a critical factor. Also, data mining has limitations in which it considers the data to be static in all aspects. In traditional approach, data mining algorithms are applied to the data at a single location when data is collected in distributed way; this means that data be transferred to single

central place to enable conventional algorithms to be applied. This approach is costly in terms of communication, storage at central site and during data mining, the algorithms assume the data will remain static even though new data could arrive or parts of the mining data might no longer be valid. Additionally, the algorithm assumes that the data are constantly available and it can read as many times as it needs.

Rather than using sequence alignment methods to segment customers based on their purchase histories, a new method is built to capture shopping-frequency and variety-seeking information that will have important applications for differentiation and market positioning strategies and lead to segmentation of shoppers' with results available dynamically rather from the traditional receipt data. Additionally, by using mobile agent the proposed system involves dynamic data mining approach that performs an in-path aggregation of the data from all the nodes as and when data arrives; instead of waiting for a complete snapshot of the data [52]. Similarly, one more innovative use of the proposed approach is being shown here that will get us to know the real time usage statistics of cellular phone models which is expected to be of an immense potential to the mobile company providers in terms of marketing.

3. SYSTEM ARCHITECTURE

As seen in previous chapters, sensor networks help the physical world to interact more closely. Grouping thousands of sensors together will revolutionize information gathering. For example, a disaster detector may be set up so that temperatures of a forest can be monitored to prevent small harmless brush fires from becoming monstrous infernos. Presently, sensor networks are extensively used to monitor the environment. There is a need for novel applications.

A novel tiered architecture for gathering dynamic shopping statistics has been developed. The lowest level consists of a wireless sensor network at the store which consist of sensor nodes embedded in products or can be fixed in aisles; the deployed sensors can be visualized like a matrix. When a customer is in the proximity of the deployed sensor(s) they sense the data from the customer's sensor enabled cell phone; the shopper is not required to take out his cell phone out of the pocket. A mobile agent dispatched from a sink visits the sensors present in its itinerary, mines the data, and carries the locally computed mined data back to the available database server or sink for generating a collective shopping model in real time. This information is further used with the shopping history of the customers to get their shopping patterns. Furthermore, the sink can update the information to a regional centralized database server that houses all the shopping pattern of customers' across the stores of a region – the central server provides a flexible way to get a holistic view of customers' shopping pattern. Additionally, since the assumption is that shoppers' cell phone will be registered and as the shopping pattern is determined by the centralized server it can forward this data to a mobile information server which aids in relaying data to customers' cell phone for advertisement, proactive service

the average sensed temperature needs to be transmitted at the end of the day. When applying in-node aggregation not all sensed values need to be transmitted; only an application-specific aggregation of all the sensed values needs to be transmitted from the nodes. On the other hand, suppose an application is interested in the minimum temperature of all sensed temperatures. In which case in-network aggregation is applied that sends the temperature of a particular node to the collection point.

Hence, *in-node* aggregation is applied at the aggregator node that will act as a common point from where local distributed mining takes place. Given the semantics of the application, *pure-collection* agent collects the data of all the aggregator nodes (collection point) and that cannot be in-network aggregated. Unlike the case of *pure-aggregation* agent, a *pure-collection* agent carries more data the more aggregator nodes it visits, i.e. the size of its code is fixed, whereas the size of the data varies every hop, and largely depends on the number of aggregator nodes and the data the agent gathers from it; there will be aisles in a shop that will have the highest number of customer visits to those that have the least or none. Regardless of whether the application-specific data changes, the network-exploration data always decreases with each hop. This is because as it traverses its path, it can forget the portion of the path that has already been fulfilled. As the result arrives at the sink the retail analysts can use this data to determine the real time shopping pattern of the customers. Wireless Sensor Communication is depicted in figure 3. This way of organizing sensors in clusters reduces the communication effort [21].

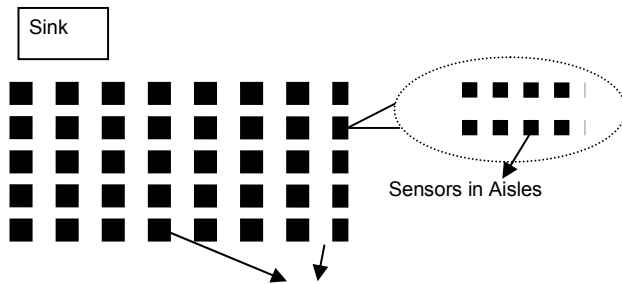


Figure 3: Schematic representation of a Distributed Dynamic Mining System

Traditionally, the raw data from the aggregator node is sent to the sink. However, this approach tends to consume a lot of network bandwidth, especially when the link bandwidth of wireless sensor network is very less [19]. Hence the agent's task is to carry only the result set [20]. Also, since the environment is closed and is not exposed to harsh extremities of nature, the working, longevity of sensors is not a problem, except interference from other in-store signals. [3, 6, 9, 10] delve into the approaches for implementing wireless sensor network architecture.

The sensors after collecting the product name, product id, device id, cell phone number, etc., from the customer's cell phone will transmit this to an aggregator node. Further, the data from each aggregator has to be filtered and moved around all the aggregators to collect the shopping pattern information [52]. This task can be easily accomplished by the use of a mobile agent. The applications and advantages of mobile agents are discussed in [22]. The mobile agent aggregates individual sensed data when it visits each aggregator node and hence does not need to have the overhead of constructing tree based dissemination protocols used in clustering or aggregation. An additional use of the agent is the dimensionality reduction of the data mining task thus improving the accuracy [25]. Also, mining data at a central location is costly in terms of communication and storage [23]. As a result, the approach discussed here deals with distributed mining on all the aggregator sensor nodes and in effect will make sensor an active

device rather being passive [52]. A single agent can collect data from all the aggregators or the task can be divided among multiple agents which collect data simultaneously, refer *future extensions* section for details.

The use of mobile phones to pay for goods electronically without taking the cell phone out from the pocket is being researched by NTT DoCoMo [45]. Similarly, with the help of a wireless sensor enabled device, store analysts can routinely go around the store to collect real time sales pattern. The advantage of this approach over the current method(s) is that the analyst/shopkeeper is not required to point the device towards the product – has to only walk around the aisles, and all the information will be automatically collected.

Another novel applicability of the proposed architecture as described in fig. 2 is in the field of real time tracking of a cell phones' market share by the use of Bluetooth wireless technology which can leverage more than just transferring files, photos, etc. from a mobile device to another Bluetooth enabled devices. Here too, the lowest level consists of a sensor network at a pre-determined spot like train station, shopping mall, colleges, etc. For example, installing Bluetooth sensors in schools, colleges and universities will aid a company in determining about the cell phone model usage among the school going populace who are likely to be in the age group of 18-28. As a result this will lead to real time directed mobile phone promotions or discounts depending on the students' crave for the models instead of relying on the traditional time consuming way of collecting usage statistics from students and then basing the marketing strategies. Consequently, to increase consumer base, a company can follow the same strategy to install Bluetooth enabled sensors in train stations, parks, shopping malls, etc.

As a person comes in network's proximity with the visible/discovery mode on it will detect the Bluetooth signal of the user and retrieve the device's Mac id (a unique Hex code), and

sometimes an editable broadcast name, as specified by the user. The sensor nodes transmit this information along with the date and time of sighting to an aggregator node. A mobile agent dispatched from a sink visits the aggregator and gets the filtered data back to the sink. While visiting other aggregator nodes if an agent detects the same Mac id again then the sighting time and the number of sightings gets updated. The forwarding of data to the sink can be done periodically or on a request to the sink or a remote database of the company. Since data is first stored and then sent at certain period of time it is rather easy to establish an error free communication channel that can preserve all relevant characteristics.

On a regular basis, this will lead us to the information of how many models have been found for a region. Maintaining a sightings field for each model will help us determine how many times a model has been found in a region, period since the model is in use, and can help in determining if the model's user is local to that region or is roaming, i.e. foreign to a area. In the same sense, tracking the first and last sighting of a model can help the mobile providers to know about their models longevity or the models that are in vague.

Since anyone can do a passive scan and only Bluetooth MAC addresses are tracked and not their personal information (phone number, sms history, etc.), there should not be any privacy problems arising.

3.1 Shopper's Model

I categorize the shoppers into the following categories:

- Random shoppers have no strong intention to purchase something, and just wander among aisles a.k.a. window shoppers

- Rational shoppers visiting a store, know clearly what they need a.k.a prompt shoppers
- Recurrent or regular customers are customers who visit the store often. They can be further divided into
 - Customers with higher purchasing power
 - Customers with lower purchasing power

Therefore, modeling customer groups and performing multi-level promotions are efficient marketing strategies in better serving the shoppers, and in customer retention.

3.2 Mobile Device Usage Model

I categorize the mobile usage into some of the below mentioned categories:

- Popular cellular phone cravings
 - Brand popularity where the people are attracted or loyal towards a company
 - For a cell phone company, popularity of a given model or total volume of their models
- Cellular phone usage among an age group
 - Educational period is a stage among the age group of 18-28, generally students attending schools, colleges, and universities.
 - Working period, among the age group 28-60

This way of modeling of cell phone consumers will help a company to quickly and effectively channelize their promotions in real time.

3.3 System Algorithms

Purchasing sequences of a store are major business intelligence sources of customers' purchasing behaviours, and can be used for marketing purchases. A DDMS is not typical time series because time is a spontaneous reference factor here. On the other hand, the time aspect is important to understand the shoppers' behaviours and ought to be modelled in the analysis.

A DDMS is a set of transactions $\langle T, t \rangle$ where 'T' is a purchase or product information event and 't' represents the time and date of the occurrence of T. On the same lines of [51] the formal expression for DDMS is

$DDMS_M = \{ \langle T_1, t_1 \rangle \dots \langle T_n, t_n \rangle \}$ where M is the recorded Mac Id of the customer's cell phone. Using a formal description

$$DDMS :: \emptyset \mid \langle T, t \rangle \mid \{ \langle P, t \rangle DDMS \}$$

For example, using symbols and numbers to encode the event of occurrences, a purchase sequence of a shopper is $DDMS_{01:34:A3:09:00:01} = \{ \langle \text{Fruits}, 05:30 \ 05/10/07 \rangle, \langle \text{Cereals}, 07:45 \ 01/11/06 \rangle \}$, and similarly for cell phone tracking is $DDMS_{01:34:C3:09:00:01} = \{ \langle \text{GSU Plaza}, 05:30 \ 05/10/07 \rangle, \langle \text{GSU SC}, 07:45 \ 05/10/07 \rangle \}$

The system will have a rule base from which the segmentation of the customers is achieved. Before the start of data mining, a miner designs a set of seed DDMS. Each rule is

considered to be an ideal or desired shopper with unique purchasing patterns. Some of the patterns for ideal customers are discussed below.

- Recurrent Customer with higher purchasing power

$R_1 = \{ \langle T, t \rangle, \langle T+a, t+b \rangle, \langle T+a, t+2b \rangle \}$ where “a” and “b” are pre-determined increments of T and t respectively.

- Recurrent Customer with lower purchasing power

$R_2 = \{ \langle T, t \rangle, \langle T+c+a, t+b \rangle, \langle T+c+a, t+3b \rangle \}$ where “a”, “b” and “c” are pre-determined increments of T and t respectively.

- Window Shopper

$R_3 = \{ \langle T, t \rangle, \langle T+a, t+4b \rangle, \langle T+a, t+6b \rangle \}$ where “a” and “b” are pre-determined increments of T and t respectively.

- Brand popularity

$R_4 = \{ \langle T, t \rangle, \langle T+a, t+b \rangle, \langle T+a, t+2b \rangle \}$ where “a” is the pre-determined variable and “b” is the automatic increment for a given T and t respectively.

Similarly, we can have patterns for rational shoppers, shoppers who respond to promotion offers, etc. These parameters will be highly dependent on the specific application. Setting the value of

these parameters will affect effectively the resultant segment sizes. The rules updating process is important and can itself be regarded as a data mining task. The thesis discusses on analyzing real time shoppers behaviour against a given set of rules. The purchasing patterns of customers are random. Hence, an algorithm that can match a pattern with a rule is provided below.

The mining algorithm is discussed as follows:

1. Define Rules and corresponding parameters for each Rule
2. for shopper (Mac Id) $m=1$ to M
 1. Identify $DDMS_m$ from set of DDMS
 2. for segment $r=1$ to R
 1. Select Rule r from the set of Rules, R
 2. If ($Rule_r \subseteq DDMS_m$)
 1. Add Shopper m to group r
 3. The R groups of shoppers are the segments

The above described algorithm will be the mining part of the agent code. The agent does mining on each cluster node and gets only the result set or summary structure. Mobile agents can save up to 90 percent of the data transfer time by avoiding the raw data being transferred [23].

In the applications as discussed above, the mobile agent will be a *pure-collection* agent; a mobile agent carries more data the more nodes it visits; the agent visits only the cluster nodes. The weight of an agent depends on the size of its code and on the size of the data it carries. The size of its code is fixed, whereas the size of the data varies every hop, and largely depends on what the application does upon visiting each node.

The mobile agent migration algorithm is discussed below:

```

If(thisNode == firstAggregator)
    MA migrates toward firstAggregator
Else if( (thisNode == nextAggregator) && (nextAggregator != lastAggregator) )
    MA collects sensed raw data and does local mining
    Set nextAggregator in the MA packet
    MA migrates towards next aggregator
Else if(thisNode == lastAggregator)
    MA collects sensed data
    MA migrates back to sink

```

When the mobile agent arrives at an aggregator node, it looks at the identifier of the current node to decide whether or not it has arrived at the destination source. If not, the mobile agent continues migrating towards the specific source. Otherwise, it operates as follows:

1. Collect the locally mined data
2. Delete the identifier of current target aggregator from aggregatorList
3. Choose nextAggregator as the next destination. If the current node is lastAggregator, the mobile agent will return to the sink.

4. SYSTEM SIMULATION

In order to gain more insight of the proposed system a simulation of a wireless sensor enabled network and shoppers going around the aisles with a sensor enabled cell phone was done. The aisles were depicted as a two dimensional matrix for each aggregator. The aggregator nodes that collected the data and where an agent did the filtering were treated as computer terminals. Each of these computer hosts had data supplied from a two dimensional size aisle. For simplicity the randomly generated sensor data was modeled on a Bluetooth signal, i.e. <Mac Id, Model Name>.

Bluetooth wireless technology is a short-range communications system intended to replace the cables connecting portable and/or fixed electronic devices. The key features of Bluetooth wireless technology are robustness, low power, and low cost. It operates in the unlicensed 2.4 GHz spectrum and can operate over a distance of 10 meters or 100 meters. The main feature of this technology is the absence of line-of-sight positioning of connected devices and the cost of these chips is under \$3 [39]. A comparison with other present/futuristic, popular/non-popular wireless technologies can also be found in [39].

4.1 Shopping Scenario

Consider the scenario of a book store company where the books have been categorized in different aisles. Each aisle has a set of sensors which in turn send data to an aggregator node present for a pair/set of aisles. To model the users moving among the aisles the data was generated randomly for these aisles. As and when customers move around the aisle and look out for books, the information <Mac Id, Model Name, Shopper Name, Aisle Number, Products,

Checkout> is sent to an aggregator node from where it is then collected by a visiting mobile agent in a pre-determined manner. As a result, customers shopping and checkout patterns gets dynamically monitored in the store.

4.1.1 Observations

The criteria to find the popularity of aisles by filtering out the products lifted from an aisle are depicted in figure 4. Forty eight books have been picked from Aisle VII while on the other hand only twenty seven books from aisle VI, which reveals that the previous aisle is more popular than the latter one. This real time information can be used to maximize selling – by restacking the books from both the aisles.

Figure 5 gives us information on the amount of books picked from an aisle at different time intervals with the data supplied from two separate simulations. On a similar basis we can get to know the popularity of aisles for different hours of a day, days in a week, weekends, etc. This information will be important for a retail to strategically place or promote the books – after determining the time based popularity of aisles.

Figure 6 gives us customers' scenario according to their purchasing power. Shoppers were filtered depending on the products they lifted from the aisles. Here for a threshold value of forty products they have been categorized into low and high purchasing value customers. This categorization will lead a store to effectively streamline their advertising.

Figure 7 depicts the number of books picked to the books checked out by the customers. Among the shoppers that were generated for the aisles one more selection was made out for checkout. Analyzing this with the previously stored data will aid in determining the customer's shopping information on a one-on-one basis. After studying the customers shopping history the

customers can be offered discount or promotion offers on the books that they did not checkout. On one hand the result reveals us about the prompt shoppers Chirayu & Mounica, and on the other hand prompt shoppers Alice and Monica.

Most of the times, selections are made on impulse but this impulse might not end up in owning the item. This will no more be valid because at checkout, a customer can be offered discounts on the book(s) that he had evinced interest but didn't end up buying. For example, we can see that a customer didn't check out any books from aisles I & X revealing the likely book preferences of the customer, figure 8 shows the shopping share to the checkout share for a customer along with the shoppers aisle movement pattern.

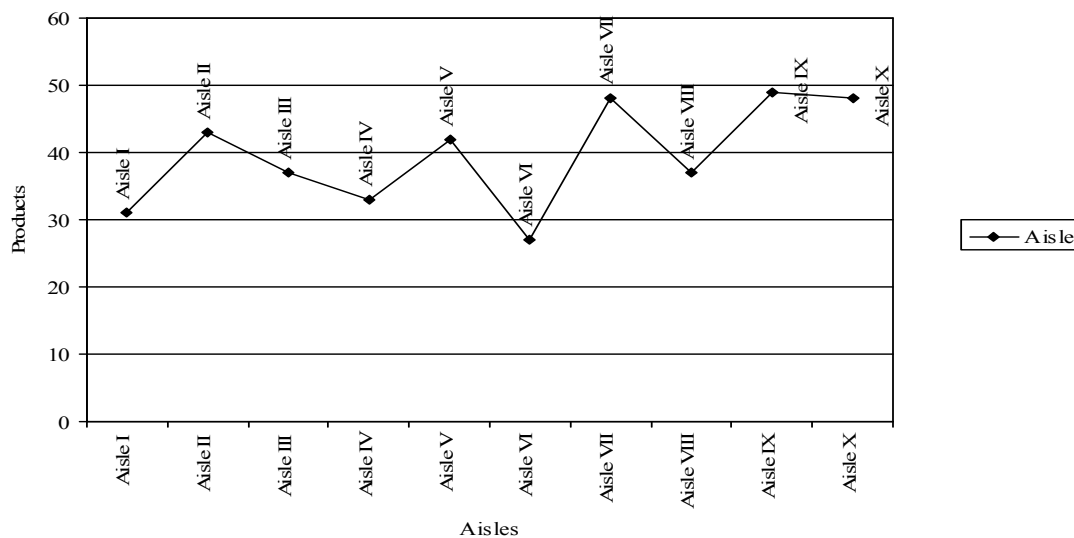


Figure 4: Popularity of aisles

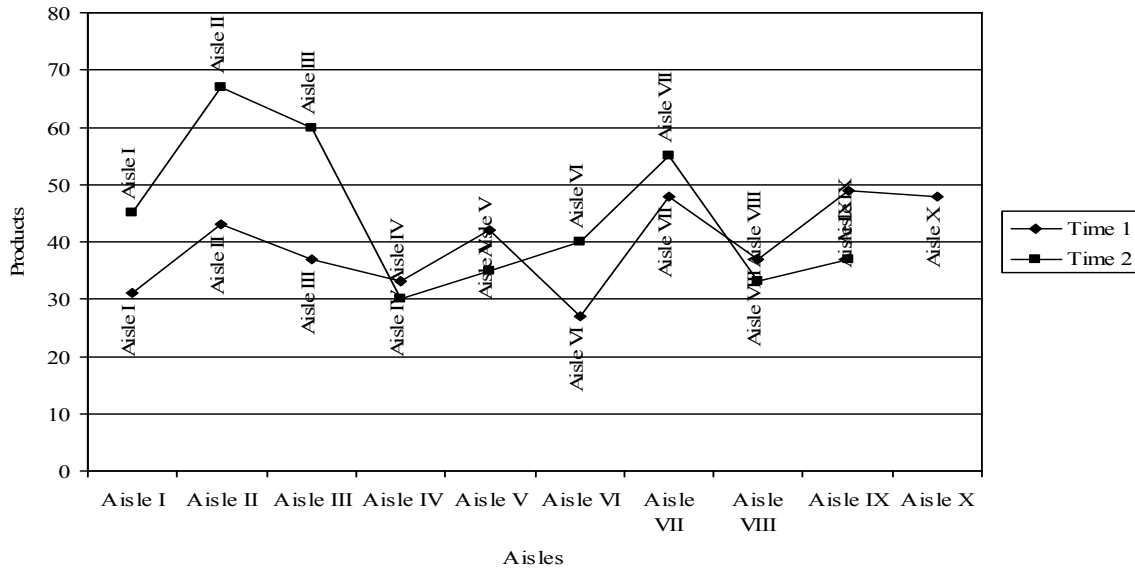


Figure 5: Popularity of aisles at different time intervals

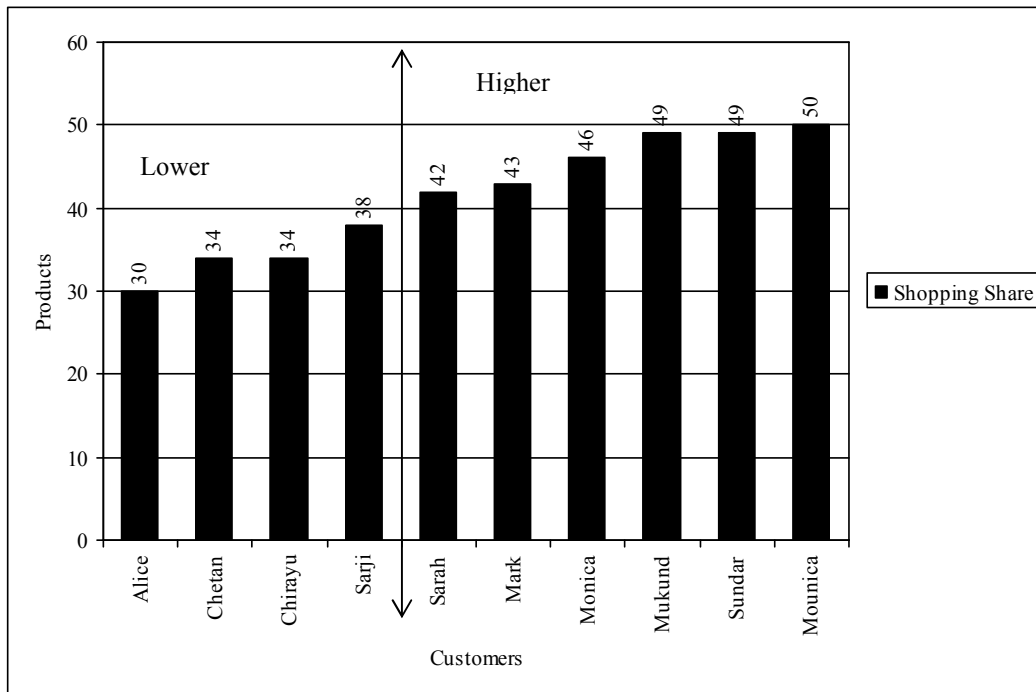


Figure 6: Lower v/s higher purchasing value

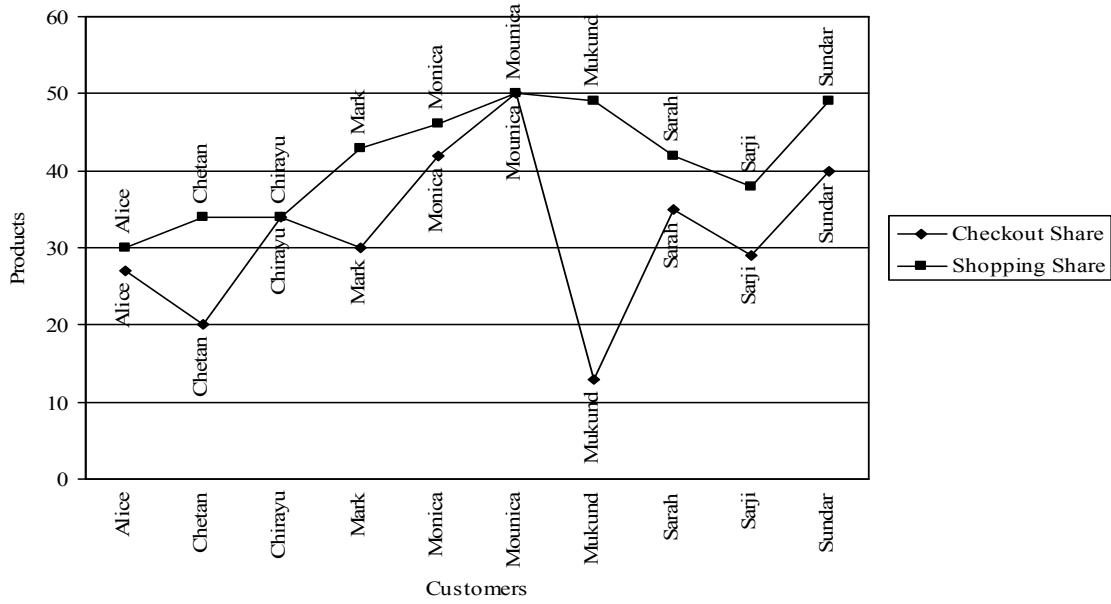


Figure 7: Customers checkout v/s shopping share

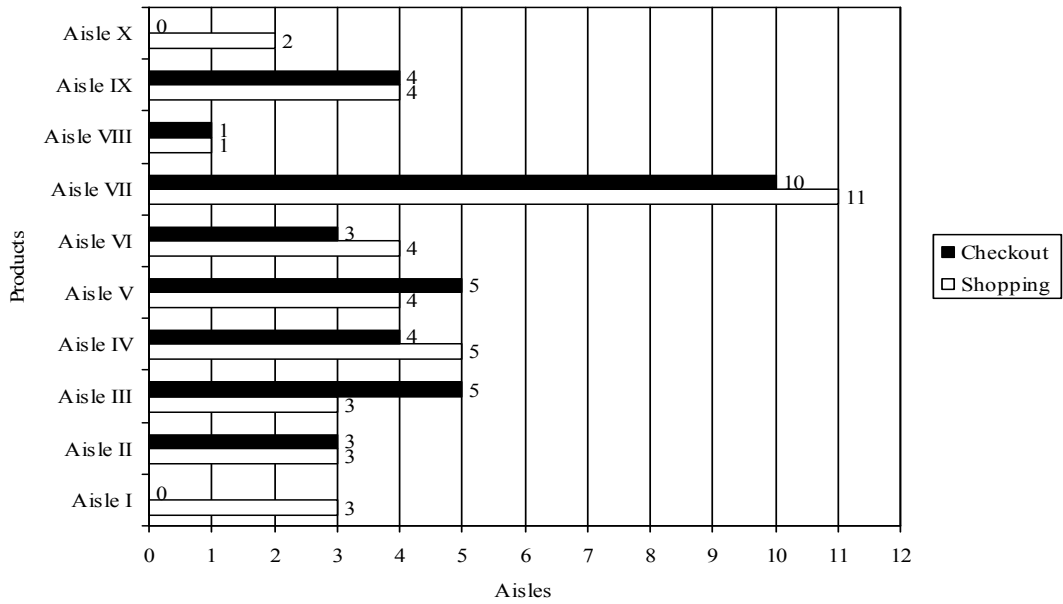


Figure 8: Shopping v/s checkout pattern of a customer

Depending upon retail's requirements much more useful information can be mined to effectively understand a customer's shopping trait dynamically, i.e. before a customer exits a store.

4.2 Mobile Device Usage Scenario

Consider the scenario of Georgia State University (GSU) campus where a mobile device company wants to collect information about their cell phone penetration among the graduate and undergraduate students so that accordingly they can base their marketing policies. Instead of conducting a manual market survey in GSU they rather decide to install Bluetooth enabled sensors in campus's plaza and student center. As a result when students turn on their visible/discovery mode and are in the vicinity of the Bluetooth sensors their unique hexadecimal Mac id gets tracked along with their sighted time. This way the company gets real-time device usage scenario. Moreover, the data can also reveal which models are famous among all the models currently in market. A side advantage of real time data gathering is that the companies can get to know how the new models are gaining market share, almost non-stop.

Consider a sub scenario where AppleTM is releasing a new model of its new iPhone for the market. According to a market study the sales are expected to rise drastically. Additionally, the company's strategy is to stop manufacturing the current model and make the new model their main stake. But in order to do this they should get a tab on how the users are switching from their current iPhone model to the new one; they need to have a real time statistic, where the old model usage is going down and the new one climbing up. Instead of relying on the traditional often delayed market study description the company can now make use of the real time statistics and can make a quick decision of how to streamline their promotions for the new market release.

4.2.1 Observations

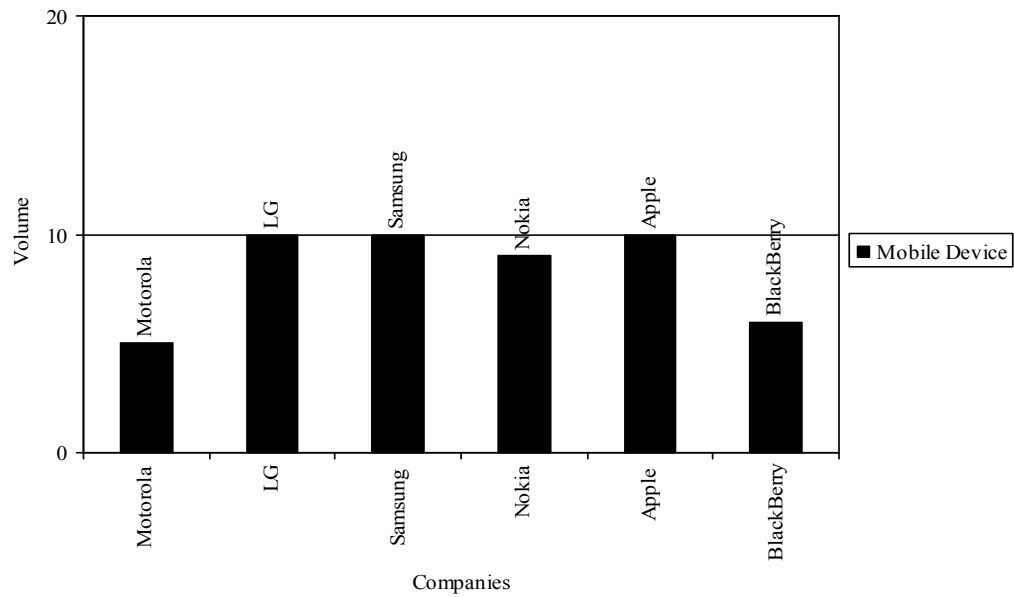


Figure 9: Mobile Devices sighted at GSU Plaza

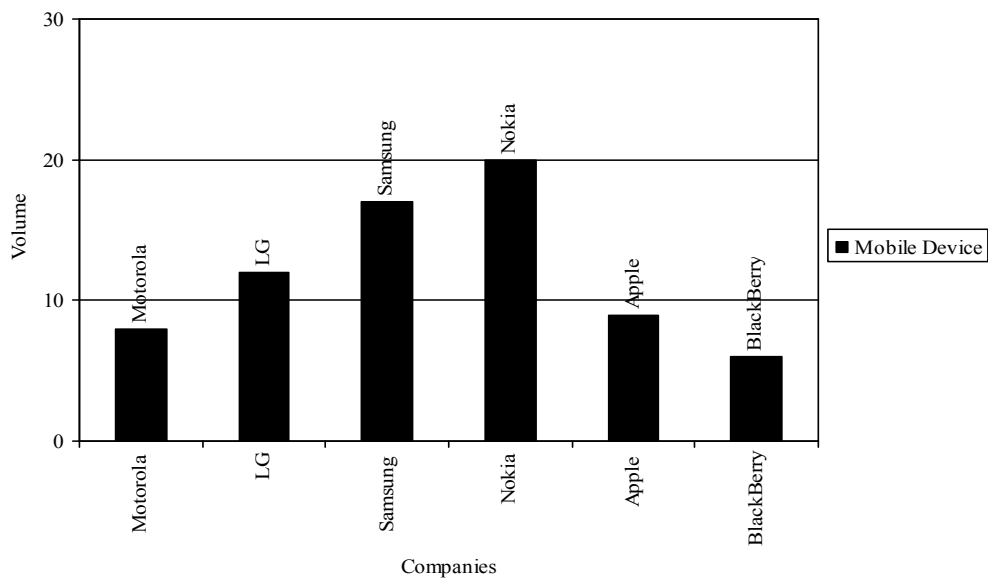


Figure 10: Mobile Devices sighted at GSU Student Center

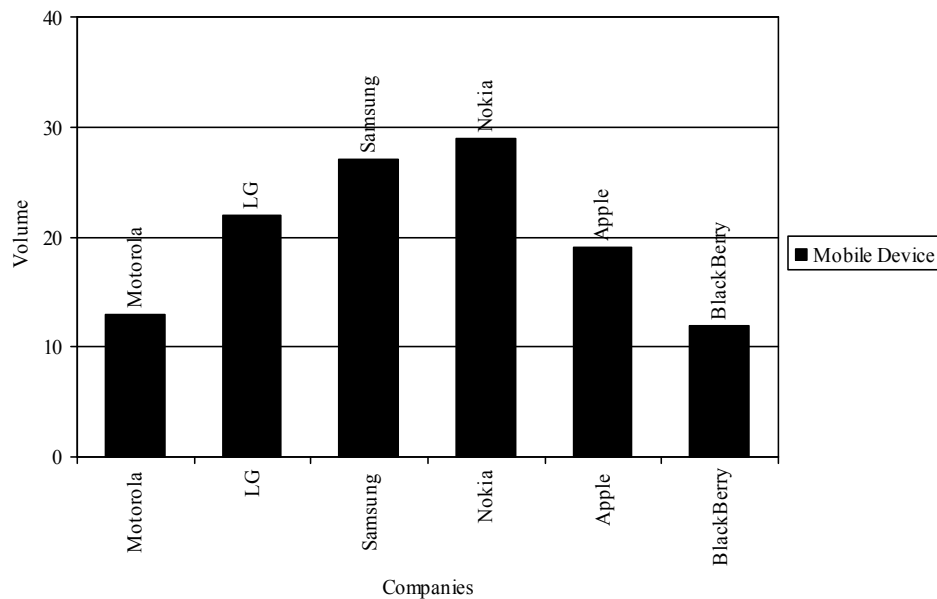


Figure 11: Most Popular Mobile Devices sighted at GSU

Figure 9 and figure 10 depict the scenario of the mobile devices sighted by the Bluetooth enabled sensors at the GSU student plaza and GSU student center. Since the Mac ids will be unique for a device and can be associated with a company the above results can be gathered. As a result, figure 11 depicts the total mobile device usage scenario of university students, on a company basis. Hence, the dynamically generated results help a company of how to market their models and increase their popularity among students. Figures 12 to 14 show the same dynamic results as obtained above from the perspective of Google maps TM and Google charts TM. The results depicts that Nokia has the highest brand following in GSU next only to Samsung, etc.

Note: Motorola (Motto), Samsung (Sam), Nokia (Nok), Apple (App), BlackBerry (BIBr)

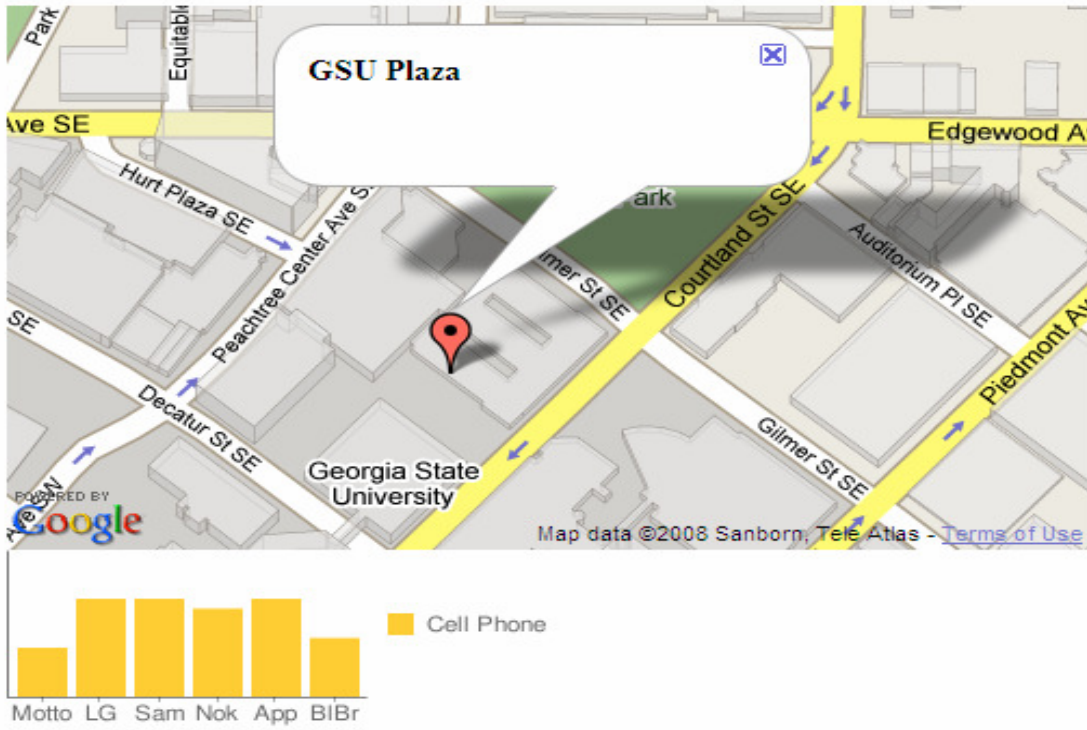


Figure 12: Google Maps™ plot of Mobile Devices sighted at GSU Plaza

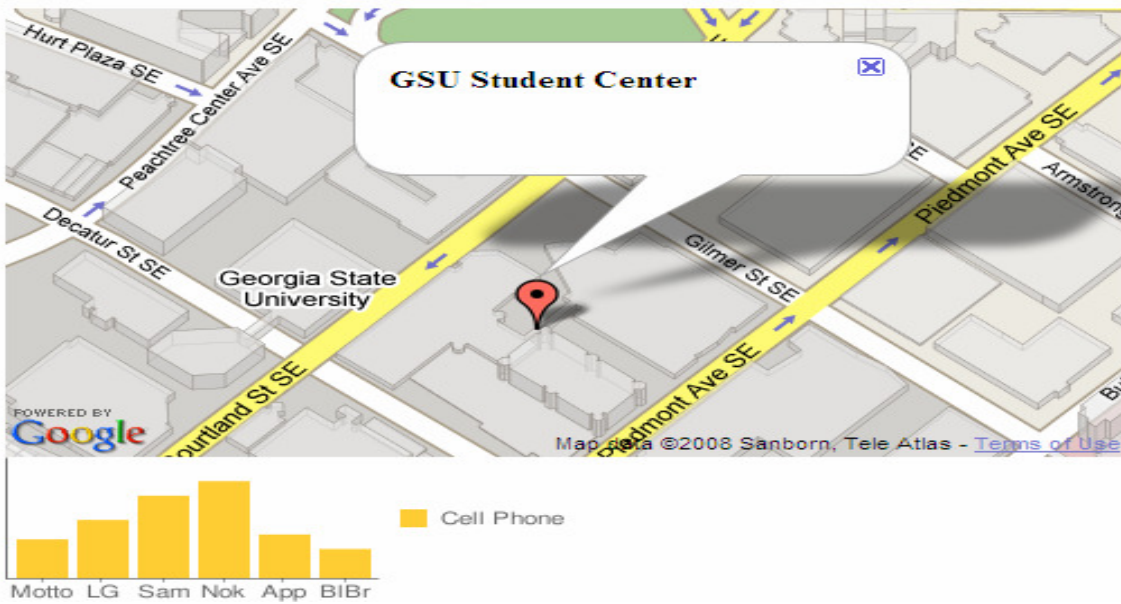


Figure 13: Google Maps™ plot of Mobile Devices sighted at GSU Student Center

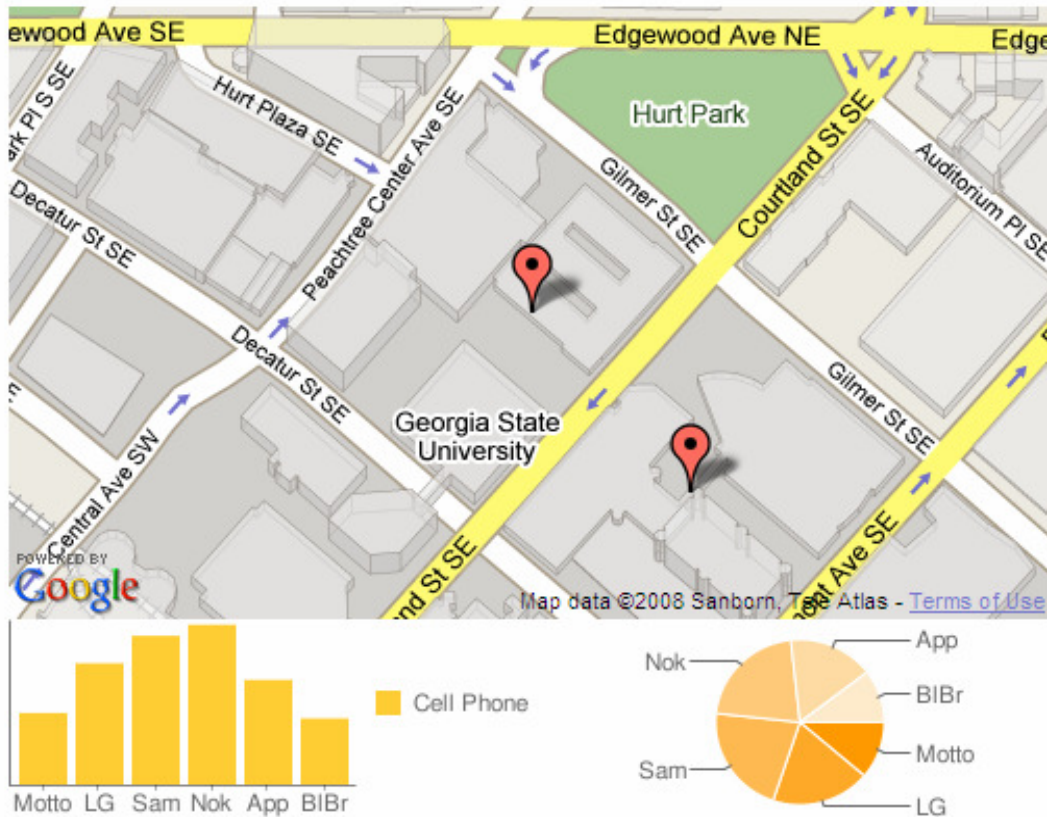


Figure 14: Google Maps™ plot of most popular Mobile Devices sighted at GSU

The popularity graphs depict the brand consideration among the students of GSU. Instead, if a company wants to find out how their different models are prominent among the students so that it would help in gauging the feature usage and preferences of its customers, the system has to get this information and present the various shares of the models. Occasionally, the gathering of this data will also shed information on the popular mobile carrier at place.

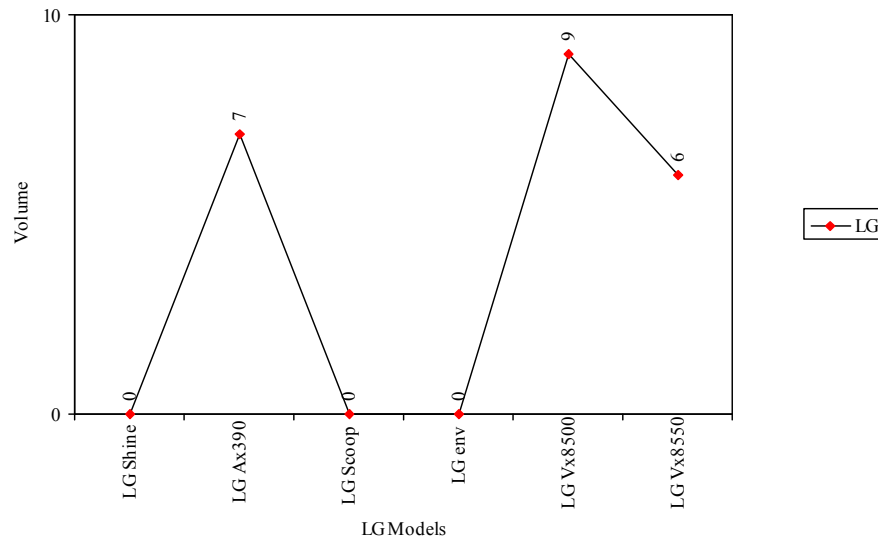


Figure 15: Volume usage of LG Mobile Devices sighted at GSU

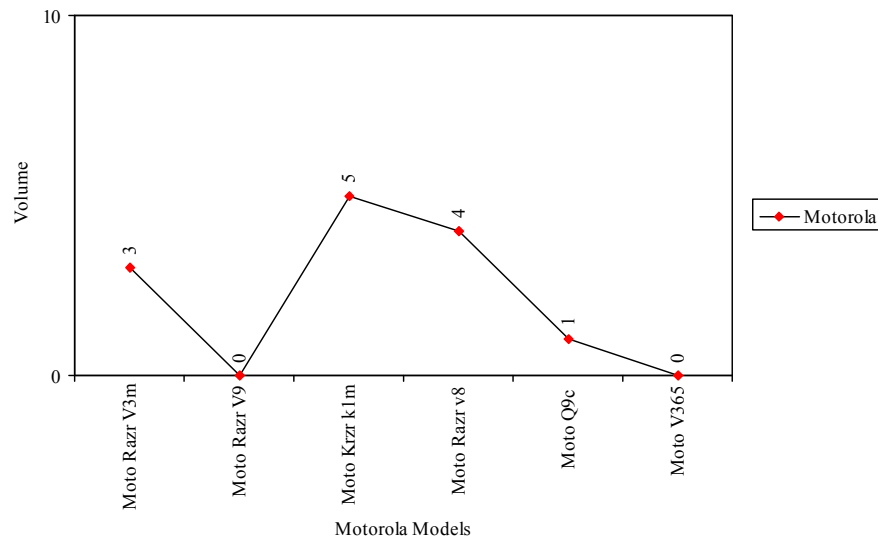


Figure 16: Volume usage of Motorola Mobile Devices sighted at GSU

Figure 15 reveal that only LG Ax390, LG Vx8500 and LG Vx8550 have been sighted among the students. This reveals a pattern that the other models are not very popular among the students either because of the cost, intriguing features of the devices, etc. On the other hand various models of Motorola can be spotted, except two, as seen from figure 16. This shows us that although Motorola is not the popular brand among the students at GSU (from figure 11) it still enjoys volume domination amongst LG and Motorola. On a same scale the results can be obtained for various other brands and other key industry performance metrics can be tracked in real-time.

Figure 17 shows how a new release is getting response among the students. As seen from the graph in May there were no sales of the product but in June the sale picked up and is still climbing. This will provide the company an insight of their new product and will assist in making a marketing strategy. Thus, the usage statistics of the newly released apple iPhone3G can be tracked in-time and will be interesting to see the growth of iPhone3G among the student community.

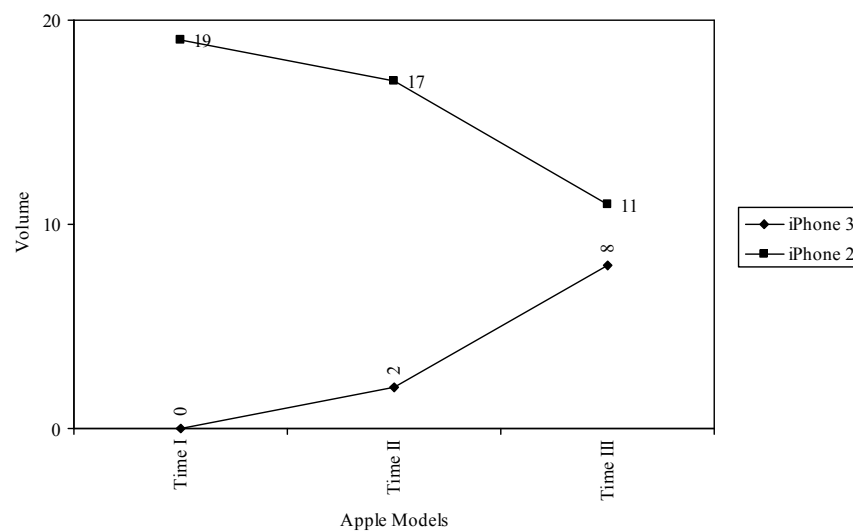


Figure 17: Volume usage of Apple Mobile Devices sighted at GSU

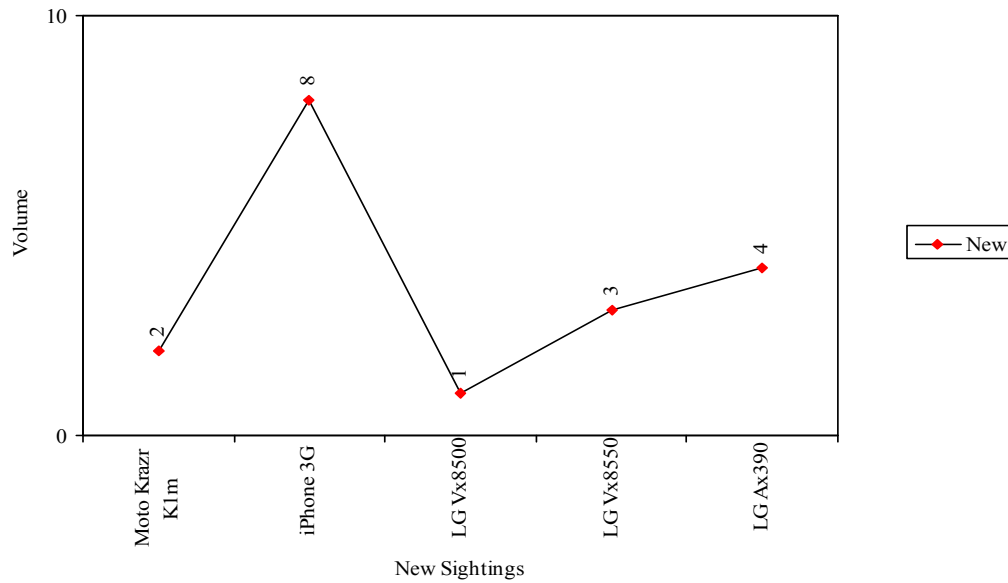


Figure 18: Volume usage of New Mobile Devices sighted at GSU

If a company wants to find out what models are new in GSU either to cross-check how effective their promotion campaign is or to know the students' current preferences figure 18 helps in revealing that information thus leading a company to know the real-time market trend for cellular phones.

5. CONCLUSION & FUTURE WORKS

While waiting in line at a store, you might hear the beep, beep, beep of the store scanners, reading the bar codes on the grocery items, ringing up on the register, and storing the data on servers located at the store headquarters. Each beep indicates a new “observation” in the information being collected about the shopping habits; identifying and reaching out to the most profitable and promising customer is the first vital step for one-to-one marketing. Likewise devices of users with Bluetooth wireless technology enabled will not only be wirelessly connected to a headset, transferring files, photos etc. but will be doing more than they realize; will be helping cellular phone companies to get current market usage data of their brands.

5.1 Conclusion

By better understanding the relationships between customers and products, customers can be easily guided through the purchasing process; discounts, promotion offers can be made to groups who will be the most receptive to the offering. Having a system that can do this in a limited time frame can make the difference between a lost sale and a satisfied new customer. Wireless sensor networks are inherently large and resource constrained. In addition, sensor applications are time-sensitive. Thus data collection on wireless sensor network has to be time and energy aware. In this thesis, a novel solution is proposed where the limited use of sensors can be extended to the vast field of retail, apart from the existing traffic monitoring, habitat monitoring, etc. It is a real-time pervasive system that collects, computes, and analyzes data in an unobtrusive manner. More specifically, it effectively combines the real time purchase and

purchase history to identify customers' shopping pattern. Information of this kind can be useful for retailers to target customers with more personalized marketing actions; be it shopping or mobile device usage scenario. Additionally, it is modeled such that the computation is moved to data rather than the data moving to computation (data-centric), and provides meaningful results to even non-experts. The real-time mobile device usage application adds on to the advantages of the proposed architecture and thus provides a service that will go a long way in finding out the market penetration, and lends its hand in forecasting a company's marketing model for consumers, falling from different regions.

In summary, the thesis provides a novel methodology of analyzing the business and helps to effectively draw dynamic correlations between customers and the products they purchase, and can be complementary to existing decision support systems [52].

5.2 Future Works

Currently, the task of mining and aggregating the mined results is given to a single agent. However, this can be divided among multiple mobile agents that can simultaneously mine the aggregator nodes and get the results back to sink. This feature can lead to a speedy retrieval of the collected data. Additionally, every mobile agent can make its own local decisions to adapt its given route to topology changes. In the event of a node failure, if an agent cannot successfully migrate from the i^{th} node of its proposed route, to the $(i+1)^{\text{th}}$, it incrementally tries to migrate to the j^{th} node such that $j > (i+1)$.

Since the thesis assumes the use of a sensor enabled cell phone this feature can be further explored by giving assistance for a particular product. The user can request product assistance which is sensed by the closest proximal sensor and is submitted to sink. Additionally, the

customers traversing pattern can be studied by storing that information in the user's cell phone, and when he checks out, all this information will be retrieved. With regard to the agent communication; datagram (UDP) based communication can be used as it costs less energy compared to streaming communication (TCP); the unreliability factor of UDP can be taken care of during implementation.

Depending on the customer's shopping pattern targeted product promotion offers can be sent especially when a customer is in a store or in a particular aisle and had evinced an interest in that product before. The product promotions can be either through "push" or "pull" method [11]. The SenMinCom architecture can be extended to provide support for location based services as proposed in [12].

A social network is described as a social structure made of nodes (which are generally individuals or organizations) that are tied by one or more specific types of relations, such as financial exchange, friendship, trade, hate, web links, etc. Currently there are two basic types of mobile social networks. The first is companies that partner with wireless phone carriers to distribute their communities via the default start pages on mobile phone browsers, e.g. Jumbuck [47], AirG [48], etc. The second type are companies that don't have such carrier relationships (also known as "off deck") and rely on other methods to attract users, e.g. Next2friends [49], Bluepulse [50], and many more.

On a similar basis the time is ripe for a third type among the cell phone providers. This can be on how their various models are performing in the market for a particular region or area. Consequently, an incremental holistic view of the mobile market share can be built in real time. An off shot of this service will be acquiring the cell phone market share information of the rival companies. Furthermore, this can be offered as a service to the consumers or to the companies

itself to let them know their cell phone or particular model popularity. In order to gain a larger share of the marketing pie the companies will eventually switch to competitive marketing; in the run making consumers happy, and the concept holds good for any blue tooth enabled device e.g. Apple iPod, Microsoft zune player, etc.

In future, the proposed architecture can be extended to support dynamic personalized sensors fusion system that monitors abnormal behaviors and health conditions (fatigue, heart attack) of a person or a driver, and that makes a real time warning to the driver and relevant people (relatives, nearby doctor, etc). Additionally, by the novel use of soft computing methodologies health predictions can be added which help in spotting dangerous symptoms before-hand. Additionally, the system can be used upon to build value added service for any domain.

REFERENCES

- [1] Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D., Anderson, J., “Wireless Sensor Networks for Habitat Monitoring”, Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, 2002, pp.88-97.
- [2] Warrior, J., “Smart Sensor Networks of the Future”, Sensors Magazine, March 1997.
- [3] Pottie, G.J., Kaiser, W.J., “Wireless Integrated Network Sensors”, Communications of the ACM, vol. 43, no. 5, pp.551-558, May 2000.
- [4] Cerpa, A., Elson, J., Estrin, D., Girod, L., Hamilton M., Zhao, J., “Habitat monitoring: Application driver for wireless communications technology”, 2001 ACM SIGCOMM Workshop on data Communications in Latin America and the Caribbean, Costa Rica, April 2001.
- [5] Werner-Allen, G., Johnson, J., Ruiz, M., Lees, J., Welsh, M., “Monitoring volcanic eruptions with a wireless sensor network”, Wireless Sensor Networks, 2005. Proceedings of the second European Workshop, 2005, pp.108-120.
- [6] Intel Research Sensor Network Operation, http://intel.com/research/exploratory/wireless_sensors.htm.
- [7] Shih, E., Cho, S., Ickes, N., Min, R., Sinha, A., Wang, A., Chandrakasan, A., “Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks”, Proceedings of ACM MobiCom’01, Rome, Italy, July 2001, pp.271-286.
- [8] Herring, C., Kaplan, S., “Component-based software systems for smart environments, IEEE Personal Communications, October 2000, pp. 60-61.
- [9] Kahn, J.M., Katz, R.H., Pister, K.S.J., “Next Century Challenges: Mobile Networking for Smart Dust”, Proc. ACM MobiCom ‘99, Washington DC, 1999, pp.271-278.

- [10] Bennett, F., Clarke, D., Evans, J., Hopper, A., Jones, A., Leask, D., "Piconet: Embedded Mobile Networking", IEEE Personal Communications, 4(5), October 1997.
- [11] Varshney, U., Vetter, R., "Framework, Applications, and Networking Support for M-commerce", ACM/Kluwer Journal on Mobile Network and Applications (MONET), June 2002.
- [12] Varshney, U., Vetter, R., Kalakota, R., "Mobile Commerce: A New Frontier", IEEE Computer, 2000, 22(10), pp.32-38.
- [13] Subramanian, L., and Katz, H.R., "An architecture for building self-configurable systems", In IEEE/ACM Workshop on Mobile Ad Hoc Networking and Computing (MobiHOC 2000), 2000.
- [14] G.E.P. Box and G.M.Jenkin, "Time Series Analysis: Forecasting and Control, San Fransico, CA: Holden-Day, 1970.
- [15] B. Everitt, "Cluster Analysis", New York, NY:Halstead Press, 1980.
- [16] D. Sankoff, and J.B. Kruskal, "Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison, Reading, MA: Addison-Wesley, 1983.
- [17] B.Hay, G.Wets, and K. Anhoof, "Mining navigation patterns using a sequence alignment method", Knowledge and Information Systems, vol.6, pp. 150-163, 2004.
- [18] C.H.Jon, H.J.P. Timmermans, and P.T.L. Popkowski-Leszczyc, "Identifying purchase-history sensitive shopper segments using scanner panel data and sequence alignment methods", Journal of Retailing and Consumer Services, vol.10, pp. 135-144, 2003.
- [19] Chen M., Kwon, T., Choi, Y., "Data Dissemination based on Mobile Agent in Wireless Sensor Networks", Proceedings of the IEEE Conference on Local Computer Networks 30th anniversary (LCN '05).

- [20] Ong, K., Zhang, Z., Ng, W., Lim, E., “Agents and Stream Data Mining: A New Perspective”, IEEE Intelligent Systems, June 2005.
- [21] S Goel and T. Imielinski, “Prediction-based monitoring in sensor networks: Taking lessons from mpeg”, ACM Computer Communication Review, 31(5), 2001.
- [22] Lange, B.D., Oshima, M., “Seven Good Reasons for Mobile Agents”, Communications of the ACM, March 1999, Vol. 42, No.3.
- [23] Iyengar, H.Qi. S., and Chakrabarty, K.,”Multi-resolution data integration using mobile agents in distributed sensor networks”, IEEE Trans. Syst., Man, Cybern. C, vol.31, no.3, pp. 383-391, Aug. 2001.
- [24] Goel, S., and Imielinski, T., “Prediction-based monitoring in sensor networks: Taking lessons from mpeg”, ACM Computer Communication Review, 31(5), 2001.
- [25] Bontempi, G., Borgne, Y., “An Adaptive Modular Approach to the mining of Sensor Network Data”, 2005 SIAM International Conference on Data Mining, April 2005.
- [26] Thearling, K., Data Mining and Customer Relationships,
<http://www.thearling.com/text/whexcerpt/whexcerpt.htm>
- [27] Akyildiz et al., 2002, Wireless sensor networks: a survey. Computer Networks. v38 i4. 393-422.
- [28] Frawley, William J., Piatetsky-Shapiro, G., Matheus, Christopher J., Knowledge Discovery in Databases: An Overview, AAAI/MIT Press, 1992
- [29] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Knowledge Discovery and Data Mining: Towards a Unifying Framework, Proceeding of the second International Conference on Knowledge Discovery and Data Mining, pages 82-88, 1996

- [30] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., The KDD Process for extracting Useful Knowledge from Volumes of Data. Communications of the ACM, Volume 39, Issue 11 pages 27-34,1996
- [31] Sumathi, S., Sivanandam, S.N., Introduction to Data Mining and its Applications, Springer, 2006
- [32] Kohler, M., Trifa, V., Epidemic Modeling using Mobile Phones, ACM SenSys 2007, The 5th ACM Conference on Embedded Network Sensor Systems, Nov 2007
- [33] MedDay has Breakthrough solution for Tsunami Warning System based on disease detection and management system, <http://www.hoise.com/vmw/05/articles/vmw/LV-VM-02-05-16.html>
- [34] GyPSii Webtop, <http://www.gypsii.com/>
- [35] Social Networking moves to the cell phone, http://www.nytimes.com/2008/03/06/technology/06wireless.html?_r=1&oref=slogin
- [36] Rasid, M.F.A. and Woodward B., Bluetooth telemedicine processor for multichannel biomedical signal transmission via mobile cellular networks, IEEE Transactions on Information Technology in Biomedicine, 9(1):35-43, 2005
- [37] Social Network Zingku, http://www.infoworld.com/article/07/09/28/Google-buys-Zingku-mobile-social-networking-service_1.html
- [38] Dodgeball.com Mobile Social Software, <http://www.dodgeball.com/home>
- [39] Bluetooth.com, Compare with Other Technologies, <http://www.bluetooth.com/Bluetooth/Technology/Works/Compare/>

- [40] Bluetooth in Public Transportation Bus Fleet of Milan, Italy, <http://www.bluetooth.com/NR/rdonlyres/235EC2BC-6A4F-4806-85ED-DD1666FEF43B/0/atm.pdf>
- [41] Norwegian Centre for Telemedicine, <http://www.bluetooth.com/NR/rdonlyres/BA25A6D5-15F0-437F-84C1-AABA58818BFD/0/diabetes.pdf>
- [42] Bluegiga Technologies Oy, http://www.bluetooth.com/NR/rdonlyres/1E2A2409-7AE1-4282-ACE4-62B652112FFC/0/medicalcasestudy_Email_version.pdf
- [43] Schwan's, <http://www.bluetooth.com/NR/rdonlyres/826F390E-C82E-43A4-9810-B1D7291A275D/0/schwans.pdf>
- [44] Oliver, N., Flores-Mangas, F., HealthGear: A Real-time Wearable System for Monitoring and Analyzing Physiological Signals, Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN '06), 2006
- [45] NTT DoCoMo Newsletter, Mobility, Adding the Human Touch to Communication, http://www.nttdocomo.com/binary/about/mobility_doc_15.pdf
- [46] New Cell phone doubles as personal trainer and shrink, <http://tech.yahoo.com/blogs/null/50133>
- [47] Jumbuck Entertainment – Mobile Community, <http://www.jumbuck.com/home/default.php>
- [48] AirG, <http://corp.airg.com/>
- [49] Next2Friends, <http://www.next2friends.com/>
- [50] Bluepulse- the free mobile social messenger, <http://www.bluepulse.com/learnmore.php>
- [51] Wang, H., Wang, S., “A Purchasing Sequences Data Mining Method for Customer Segmentation”, IEEE International Conference on Service Operations and Logistics, and Informatics , SOLI' 06, June 2006

- [52] Hiremath, N., Zhang, Y., “SenMinCom: Pervasive Distributed Dynamic Sensor Data Mining for Effective Commerce”, Proceedings of 2008 IEEE International Conference on Granular Computing (GrC 2008), Aug 2008