International Journal of Computer and Communication Technology

Volume 5 | Issue 2 Article 8

April 2014

Database and Data Mining in Social Networking

S Muktar Yakub Saheb
NIMS University, Rajasthan, India,, md_mukhtar@rediffmail.com

Mahesh S. Darak
NIMS University, Rajasthan, India, maheshd_kumar@yahoo.com

Pravin More
C.S. NES SCN, Nanded, pravinmore@gmail.com

Database and Data Mining in Social Networking

S Muktar Yakub Saheb[1], Mahesh S. Darak[2] & Pravin More[3]

[1&2]NIMS University, Rajasthan, India, [3]C.S. NES SCN, Nanded
E-mail : md_mukhtar@rediffmail.com[1], maheshd_kumar@yahoo.com[2]

Abstract **-** Today's data driven world exploiting the latest trends of database and its allied technologies like Data Warehouse and Data Mining. Data Mining in recent years emerged as one of the most efficient database technique proved to be very reliable almost in every organisation enabling to find previously unknown hidden data patterns for the benefit of organisation. At the same time it is imposing serious problems concerned to data privacy and its potential misuse.

Key words **-** *co-relations, decision trees Inference, DMKD, anonymisation, perturbation, augmentation.*

I. INTRODUCTION

- What is social networking?

In both professional and personal life, human beings naturally form groups based on affinities and expertise. We gravitate to others with whom we share interests. Most of us belong to real world networks that formed organically. Not surprisingly, these networks rapidly migrated to the online world.

Online social networking has been around in various forms for nearly a decade, and has begun to achieve wide notice in the past few years.

Online social networks take many forms, and are created for many reasons. Despite their differences, online social networks do, however, commonly exhibit a number of the following concepts.

Profiles – Each member in a network has an online profile that serves as the individual's identity in the network. In the professional context, profiles often contain information regarding the individual's experience, education, interests and affiliations as well as information about the individual's skills and resources.

Connections – Online social networks typically enable individuals to make connections with others in the network. In some cases, these connections are implicit, and derived from past actions (such as sending an email to another member of the network).

In other cases, the connections are explicit, and are set up and created by the members themselves.

Deceptively simple, online social networks contain great power. They change the online space from one of static web pages and stale marketing messages to a live, vibrant network of connected individuals who share their abilities, expertise and interests.

II. SOCIAL NETWORK RATIONAL

The information revolution has given birth to new economies structured around flows of data, information, and knowledge. In parallel, social networks have grown stronger as forms of organization of human activity. Social networks are nodes of individuals, groups, organizations, and related systems that tie in one or more types of interdependencies: these include shared values, visions, and ideas; social contacts; kinship; conflict; financial exchanges; trade; joint membership in organizations; and group participation in events, among numerous other aspects of human relationships. Indeed, it sometimes appears as though networked organizations competing with all other forms of organization certainly, they outpace vertical, rigid, command-and-control bureaucracies. When they succeed, social networks influence larger social processes by accessing human, social, natural, physical, and financial capital, as well as the information and knowledge content of these. (In development work, they can impact policies, strategies, programs, and projects including their design, implementation, and results and the partnerships that often underpin these.).

To date, however, we are still far from being able to construe their public and organizational power in ways

that can harness their potential. Understanding when, why, and how they function best is important.

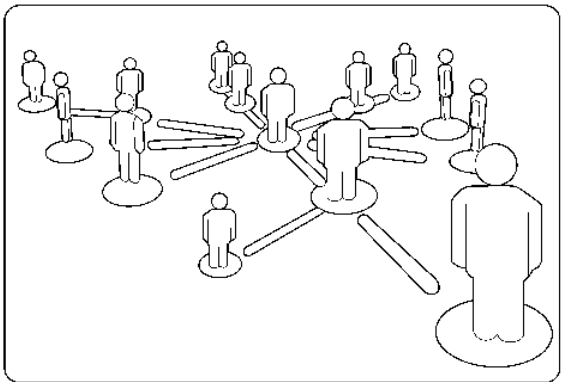

Fig. I : Social Network

The defining feature of social network analysis is its focus on the structure of relationships, ranging from casual acquaintance to close bonds. Social network analysis assumes that relationships are important. It maps and measures formal and informal relationships to understand what facilitates or impedes the knowledge flows that bind interacting units, viz., who knows whom, and who shares what information and knowledge with whom by what communication media (e.g., data and information, voice, or video communications) because these relationships are not usually readily discernible, social network analysis is somewhat akin to an "organizational x-ray".

Social network analysis is a method with increasing application in the social sciences and has been applied in areas as diverse as psychology, health, business organization, and electronic communications. More recently, interest has grown in analysis of leadership networks to sustain and strengthen their relationships within and across groups, organizations, and related systems.

III. BENEFITS

We use people to find content, but we also use content to find people. If they are understood better relationships and knowledge flows can be measured, monitored, and evaluated, perhaps (for instance) to enhance organizational performance. The results of a social network analysis might be used to:

- Identify the individuals, teams, and units who play central roles.
- Discern information breakdowns, bottlenecks, structural holes, as well as isolated individuals, teams, and units.
- Make out opportunities to accelerate knowledge flows across functional and organizational boundaries.
- Strengthen the efficiency and effectiveness of existing, formal communication channels.
- Raise awareness of and reflection on the importance of informal networks and ways to enhance their organizational performance.
- Leverage peer support.
- Improve innovation and learning.
- Refine strategies.

Development work, for one, is more often than not about social relationships. Hence, the social network representation of a development assistance project or program would enable attention to be quickly focused (to whatever level of complexity is required) on who is influencing whom (both directly and indirectly).

IV. PROCESS

Typically, social network analysis relies on questionnaires and interviews to gather information about the relationships within a defined group. The responses gathered are then mapped.

This data gathering and analysis process provides baseline information against which one can then prioritize and plan interventions to improve knowledge flows, which may entail recasting social connections.

Notwithstanding the more complex processes followed by some, which can entail sifting through surfeits of information with increasingly powerful social network analysis software, social network analysis encourages at heart participative and interpretative approaches to the description and analysis of social networks, preferably with a focus on the simplest and most useful basics. Key stages of the basic process will typically require practitioners to:

- Identify the network of individuals, teams, and units to be analyzed.
- Gather background information, for example by interviewing senior managers and key staff to understand specific needs and issues.
- Define the objective and clarify the scope of the analysis, and agree on the reporting required.
- Formulate hypotheses and questions.
- Develop the survey methodology
- Design the questionnaire, keeping questions short and straight to the point. (Both open-ended and closed questions can be used.)10

- Survey the individuals, teams, and units in the network to identify the relationships and knowledge flows between them.
- Use a social network analysis tool to visually map out the network.
- Review the map and the problems and opportunities highlighted using interviews and/or workshops.
- Design and implement actions to bring about desired changes.
- Map the network again after a suitable period of time. (Social network analysis can also serve as an evaluation tool.)

V. SOCIAL NETWORK DATA

On one hand, there really isn't anything about social network data that is all that unusual. Net workers do use a specialized language for describing the structure and contents of the sets of observations that they use. But, network data can also be described and understood using the ideas and concepts of more familiar methods, like cross-sectional survey research.

On the other hand, the data sets that net workers develop usually end up looking quite different from the conventional rectangular data array so familiar to survey researchers and statistical analysts. The differences are quite important because they lead us to look at our data in a different way and even lead us to think differently about how to apply statistics.

"Conventional" sociological data consists of a rectangular array of measurements. The rows of the array are the cases, or subjects, or observations. The columns consist of scores (quantitative or qualitative) on attributes, or variables, or measures.

Data Mining :

Data Mining is one of the advantages of database technology and it is a sophisticated statistical analysis of data, most often predictive modeling.

Thus in simple words Data Mining is *"Searching through large amounts of data for correlations, sequences, and trends."*

"Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about data".

Data mining technology provides a user- oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the administrators to improve the quality of service.

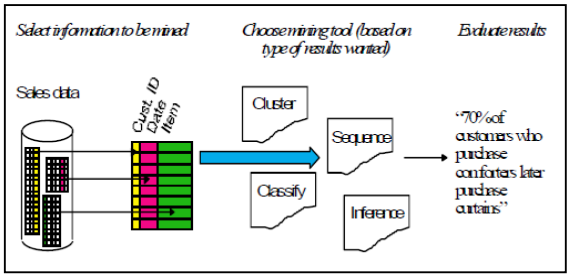

Fig. II : Working of Data Mining

Many times data mining is required to be allowed in public domain and it is a well known fact that any entity in public domain is very much vulnerable for security breaches. Like recently IPL decided to make the team owners information public to bring more transparency in on-going Twenty20-2010 matches.

The ultimate purpose of security policy in data mining is to achieve Data confidentiality, data integrity & data availability.

For example, publicly accessible corporate telephone books can decrease the need for telephone operators. Sharing need not be completely public making inventory information available to suppliers can help a retail operation to avoid shortages, and can lower the supplier's cost (thus allowing the retailer to negotiate a better price).

These two advantages, when combined, can become a disadvantage. For example, mining a corporate directory to determine staffing of a particular project (and changes in staffing) could help a competitor to determine product rollout dates, allowing preemptive marketing campaigns. Mining retailer inventory data could allow a supplier to determine sales and supplies of competing products, leading to pricing and marketing strategies aimed at reducing the competition.

VI. DATA MINING AND SECURITY

Data mining is the process of posing a series of appropriate queries to extract information from large quantities of data in the database. Data mining techniques can be applied to handle problems in database security. On the other hand, data mining techniques can also be employed to cause security problems. This position paper reviews both aspects.

Data mining techniques include those based on rough sets, inductive logic programming, machine learning, and neural networks, among others. Essentially one arrives at some hypothesis, which is the information extracted, from examples and patterns observed. These patterns are observed from posing a

series of queries; each query may depend on the response obtained to the previous queries posed.

Data mining techniques have applications in intrusion detection and auditing databases. In the case of auditing, the data to be mined is the large quantity of audit data. One may apply data mining tools to detect abnormal patterns. For example, suppose an employee makes an excessive number of trips to a particular country and this fact is known by posing some queries. The next query to pose is whether the employee has associations with certain people from that country. If the answer is positive, then the employee's behavior is flagged.

While the previous example shows how data mining tools can be used to detect abnormal behavior, the next example shows how data mining tools can be applied to cause security problems. Consider a user who has the ability to apply data mining tools. This user can pose various queries and infer sensitive hypothesis. That is, the inference problem occurs via data mining. There are various ways to handle this problem. One approach is as follows. Given a database and a particular data mining tool, apply the tool to see if sensitive information can be deduced from the unclassified information legitimately obtained. If so, then there is an inference problem. There are some issues with this approach. One is that we are applying only one tool. In reality, the user may have several tools available to him. Furthermore, it is impossible to cover all ways that the inference problem could occur.

Another approach is to build an inference controller which acts during run-time. As the user applies data mining tools, the inference controller will analyze the queries posed by the user and the answers, and determines the types of responses that should be released to the user for each query. The issues involved in building such an inference controller have to be determined.

VII. DATA MINING AND INFERENCE

The problem of data privacy in Data Mining founds its seeds in Inference. Inference is the process of posing queries and deducing unauthorized information from the legitimate responses received. For example, the names and salaries of individuals may be unclassified individually, but while taken together they are classified. This means that one could retrieve names and employee numbers, and then later retrieve the salaries and employee numbers, and make the associations between names and salaries.

The problem that occurs through this inference is called the inference problem. In the early 1970s, much of the work on the inference problem was on statistical databases. Organizations such as the census bureau were interested in this problem. However, in the mid 1970s and then in the 1980s, the United States Department of Defense started an active research program on multilevel secure databases, and research on the inference problem was conducted as part of this effort. The pioneers included Morgenstern, Thuraisingham, and Hinke In particular, it was shown that the general inference problem was unsolvable by Thuraisingham, and then approaches were developed to handle various types of inferences. These approaches included those based on security constraints as well as those based on conceptual structures. These approaches handled the inference problem during database design, query, and update.

VIII. DATA PRIVACY CONCERNS

First of all, with the World Wide Web, there is now an abundance of information about individuals that one can obtain within seconds. This information could be obtained through mining or just from information retrieval. Therefore, one needs to enforce controls on databases and data mining tools. This is a very difficult problem especially with respect to data mining. In summary, one needs to develop techniques to prevent users from mining and extracting information from the data whether they are on the web or on servers.

That is, we have portrayed data mining as a technology that is critical for analysts and other users so that they can get the right information at the right time. Furthermore, they can also extract patterns previously unknown. This is all true. However, we do not want the information to be used in an incorrect manner. For example, based on information about a person’s salary or income, he or she or his or her dears and nears may get trapped in dangerous situation like kidnap and ransom.

In many cases these denials may not be legitimate. Therefore, information providers have to be very careful in what they release. Also, data mining researchers have to ensure that privacy aspects are addressed. Next, let us examine the social aspects. In most cultures, privacy of the individuals is important. However, there are certain cultures where it is impossible to ensure privacy. These could be related to political or technological issues or the fact that people have been brought up believing that privacy is not critical. There are places where people divulge their salaries without thinking twice about it, but in many countries, salaries are very private and sensitive. It is not easy to change cultures overnight, and in many cases you do not want to change them, as preserving cultures is important. So what overall effect does this have on data mining and privacy? We do not have an answer to this yet as we are only beginning to

look into it. We are however beginning to realize that perhaps we do have many of the technological solutions for handling privacy.

That is, many of the technologies we have proposed for Information security in general and secrecy and confidentiality in particular could be applied for privacy. However we have to now focus on the social aspects. That is, we need the involvement of social scientists to work with computer scientists on privacy and data mining.

Next, let us examine the political and legal aspects. We include policies and procedures under this. What sort of secrecy/privacy controls should one enforce for the web? Should these secrecy/privacy polices be mandated or should they be discretionary? What are the consequences of violating the secrecy/privacy polices? Who should be administering these policies as well as managing and implementing them? How is data mining on the web impacted? Can one control how data is mined on the web? Once we have made technological advances on data mining, can we then enforce secrecy/privacy controls on the data mining tools? How is information transferred between countries?

We have, however, begun discussions. Note that some of the issues we have discussed are related to privacy and data mining, and some others are related to just privacy in general.

We have raised some interesting questions on privacy issues and data mining as well as privacy in general. As mentioned earlier, data mining is a threat to privacy. The challenge is on protecting the privacy but at the same time not losing all the great benefits of data mining. At the 1998 knowledge discovery in database conference in New York City, there was an interesting panel on the privacy issues for web mining. Much of the focus at that panel was on legal issues. It appears that the data mining as well as the information security communities are now conducting research on privacy. Furthermore, social scientists are also now interested in privacy in the new information technology era.

IX. PRIVACY AND ACCURACY

A. Privacy

Data mining itself is not ethically problematic. The ethical and legal dilemmas arise when mining is executed over data of a personal nature. Perhaps the most immediately apparent of these is the invasion of privacy. Complete privacy is not an inherent part of any society because participation in a society necessitates communication and negotiation, which renders absolute privacy unattainable. Hence, individual members of a society develop an independent and unique perception of their own privacy. This being the case, privacy exists within a society only because it exists as a perception of the society's members. This perception is crucial as it partly determines whether, and to what extent, a person's privacy has been violated.

An individual can maintain their privacy by limiting their accessibility to others. In some contexts, this is best achieved by restricting the availability of their personal information. If a person considers the type and amount of information known about them to be inappropriate, then they perceive their privacy to be at risk. Thus, privacy can be violated when information concerning an individual is obtained, used, or disseminated, especially if this occurs without their knowledge or consent.

Huge volumes of detailed personal data are regularly collected and analyzed by marketing applications (Berry and Linoff, 1997; Khaw and Lee, 1995), in which individuals may be unaware of the behind-the-scenes use of data, are now well documented (John, 1999; Klang, 2004). However, privacy advocates face opposition in their push for legislation restricting the secondary use of personal data, since analyzing such data brings collective benefit in many contexts (Gordon and Williams, 1997). DMKD has been instrumental in many scientific areas such as biological and climate-change research and is also being used in other domains where privacy issues are relegated in the light of perceptions of a common good. These include human genome research ((Tavani, 2004)), combating tax evasion and aiding in criminal investigations (Berry and Linoff, 1997) and in medicine (Roddick et al., 2003).

As privacy is a matter of individual perception, an infallible and universal solution to this dichotomy is infeasible. However, there are measures that can be undertaken to enhance privacy protection. Commonly, an individual must adopt a proactive and assertive attitude in order to maintain their privacy, usually having to initiate communication with the holders of their data to apply any restrictions they consider appropriate. For the most part, individuals are unaware of the extent of the personal information stored by governments and private corporations. It is only when things go wrong that individuals exercise their rights to obtain this information and seek to excise or correct it.

B. Accuracy

Mining applications involve vast amounts of data, which are likely to have originated from diverse, possibly external, sources. Thus the quality of the data cannot be assured. Moreover, although data pre-processing is undertaken before the execution of a mining application to improve data quality, people conduct transactions in an unpredictable manner, which can cause personal data to expire rapidly. When mining

is executed over expired data inaccurate patterns are more likely to be revealed.

Likewise, there is a great likelihood of errors caused by mining over poor quality data. This increases the threat to the data subject and the costs associated with the identification and correction of the inaccuracies. The fact that data are often collected and analyzed without a reconceived hypothesis shows that the data analysis used in DMKD are more likely to be exploratory (as opposed to the confirmatory analysis exemplified by many statistical techniques).

This immediately implies that results from DMKD algorithms require further confirmation and/or validation. There is a serious danger of inaccuracies that cannot be attributed to the algorithms per se, but to their exploratory nature.

This has caused some debate amongst the DMKD community itself. Freitas (2000) has argued that mining association rules is a deterministic problem that is directly dependent on the input set of transactions and thus association rules are inappropriate for prediction, as would be the case of learning classifiers. However, most uses of association rule mining are for extrapolation to the future, rather than descriptions of the past.

The sharing of corporate data may be cost-efficient and beneficial to organizations in a relationship but allowing full access to a database for mining may have detrimental results. The adequacy of traditional database security controls are suspect because of the nature of inference.

Private and confidential information can be inferred from public information.

The following measures have thus been suggested to prevent unauthorized mining:

- Limiting access to the data - By controlling access to the data and preventing users from obtaining a sufficient amount of data, consequent mining will result in low confidence levels. This also includes query restriction, which attempts to detect when compromise is possible through the combination of queries (Miller and Seberry, 1989).
- Anonymizaton - Any identifying attributes are removed from the source dataset. A variation on this can be a filter applied to the rule set to suppress rules containing identifying attributes.
- Dynamic Sampling - Reducing the size of the available data set by selecting a different set of source tuples for each query.
- Authority control and cryptographic techniques - Such techniques effectively hide data from unauthorized access but allow inappropriate use by authorized (or naive) users (Pinkas, 2002).
- Data perturbation - Altering the data, by forcing aggregation or slightly altering data values, useful mining may be prevented while still enabling the planned use of the data.

  Agrawal and Srikant (2000) explored the feasibility of privacy-preservation by using techniques to perturb sensitive values in data.
- Data swapping - Attribute values are interchanged in a way that maintains the results of statistical queries (Evfimievski et al., 2002).
- The elimination of unnecessary groupings - By assigning unique identifiers randomly; they serve only as unique identifiers. This prevents meaningful groupings based on these identifiers yet does not detract from their intended purpose.
- Data augmentation - By adding to the data in non-obvious ways, without altering their usefulness, reconstruction of original data can be prevented.
- Alerting - Labeling potentially sensitive attributes and attribute values and from this calculating an estimate of the sensitivity of a rule (Fule and Roddick, 2004).
- Auditing - The use of auditing does not enforce controls, but it may detect misuse so that appropriate action may be taken.

C. Legal Liability

When personal data have been collected it is generally decontextualised and separated from the individual, improving privacy but making misuse and mistakes more likely (Gammack and Goulding, 1999). Recently, there has been a trend to treat personal data as a resource and offer it for sale. Information is easy to copy and re-sell. The phrase data mining uses the metaphor of the exploitation of natural resources, further contributing to the perception of data as commodity.

Moreover, the question of whether it is appropriate in terms of human rights to trade in personal data has seen insufficient academic and legal debate. The negative consequences of such trade are similar to those of data mining: transgression of privacy and the negative impacts of Sinaccurate data. However, the repercussions of inaccurate data are more serious for organizations trading in personal data, as the possibility of legal liability is introduced. There is the potential for those practicing data trade or data mining to make mistakes and as a consequence lose heavily in the courts.

Compensation may be ordered against any organization that is found to have harmed (or failed to prevent harm to) an individual to whom it owed a duty of care. Once liability (the tort of negligence) has been established, the plaintiff can claim financial compensation for any consequential losses caused by the negligent act (Samuelson, 1993). The extent and exact nature of the losses is, for the most part, unique to each plaintiff, but the boundaries of negligence are never closed. A mining exercise might erroneously declare an individual a poor credit risk, and decisions may be made prejudicial to that individual on that basis.

In some cases, algorithms may classify correctly, but such classification could be on the basis of controversial (ie. ethically sensitive) attributes such as sex, race, religion or sexual orientation. This could run counter to Anti-Discrimination legislation. In some cases, such as artificial neural networks, nearest neighbor classifiers, which do not make their knowledge explicit in rules, the use of controversial classification attributes may be hard to identify. Even with methods that make transparent their classification, such as decision trees, there is little to prevent a corporation using rules based on controversial attributes if that improves accuracy of the classification. Individuals who suffer denial of credit or employment on the basis of race, sex, ethnic background or other controversial attributes in a context where this is contrary to law are in a strong position to demonstrate harm only if they illustrate the artificial classifiers are using such attributes. The question is how they obtain access to the classifier results.

In the event that the person loses money or reputation as a result of this, courts may award damages. Moreover, since the potential for inaccuracies involved in the exercise is great, it is conceivable that the courts might apply a higher than usual standard of care in considering whether an organization has breached its duty to a plaintiff sufficiently to amount to negligence.

Another legal issue is whether organizations manipulating personal data can be considered capable of defaming a person whose data they have mined. It is quite conceivable that since data mining generates previously unknown information, the organization using the data mining tool can be considered the author of the information for the purposes of defamation law. Moreover, it can be argued that organizations trading in personal data are analogous to publishers, as they are issuing collections of data for sale and distribution. Hence, if the information is capable of being deemed defamatory by the courts, the data mining organizations are capable of being found liable for damages in this tort also. One difficulty is that the terms author and publisher have long been associated with text or music, not data. Note that census data also faces this challenge and other technologies are complicating the issue still further.

Consider aerial/satellite photography that can now achieve resolution to within a few metres and which can be freely purchased over the Internet. What is the resolution that makes such data be considered personal data? How can individuals living at identifiable houses decide if the aerial photo is to be used for a potential beneficial analysis, such as bush fire risk analyses of their property, or an analysis that could be considered defamatory or discriminatory?

Market analysts often see privacy concerns as unreasonable. Privacy is an obstacle to understanding customers and to supplying better suited products. Hundreds of millions of personal records are sold annually in the US by 200 superbureaux to direct marketers, private individuals, investigators, and government agencies (Laudon, 1996).

We are in urgent need of an extended interpretation of existing tort doctrine, or preferably a broadening of the boundaries of the current doctrines. Indeed, Samuelson (1993) warns that the engineered, technological nature of electronic information dissemination suggests a greater liability for its disseminators. Commonly, the conveyers of information are excused from liability if they are simply the carriers of the information from the publishers to the public -

A book store selling a book that carries defamatory material will be excused from liability that might rightly attach to the author and the publisher. It is quite possible that a mining exercise, particularly one that had mined inaccurate data, might be deemed by the courts to be an exercise in publishing, not just in dissemination.

X. POTENTIAL SOLUTIONS

A. Anonymisation of Data

One solution to the invasion of privacy problem is the anonymizaton of personal data. Data anonymizaton means *the removal of attribute values that would allow a third party to identify the individual.* This has the effect of providing some level of privacy protection for data subjects. However, this would render obsolete legitimate mining applications that are dependent on identifiable data subjects, and prevent many mining activities altogether. A suggested compromise is the empowerment of individuals to dictate the amount and type of personal data they consider appropriate for an organization to mine.

While anonymizaton of data is a step in the right direction, it is the weakest of the possible options. It is well known that additional information about an

individual can easily be used to obtain other attributes; an anonymous table of salaries and addresses, together with the knowledge of one attribute would be sufficient to determine the other. Moreover, grouping two sets of anonym information can result in disclosure. Identifier removal (such as name, address, phone number and social security number) can be insufficient to ensure privacy (Klosgen, 1995). Anonymizaton is a form of cell suppression, a technique applied on statistical databases. Indeed, the research agenda is still far from closed since most of the solutions proposed so far in the DMKD community (Piatetsky-Shapiro, 1995) are easily translated to previously suggested methods for statistical databases.

Data perturbation i.e. altering data, by forcing aggregation or slightly altering data values. Clifton and Marks (1996; 1999) indicated new and renewed threats to privacy from mining technology. Clifton's small samples method (Clifton, 1999) is unsatisfactory as the data is too small to make useful inferences, and individuals whose data is in the sample have no protection. Estivill-Castro and Brankovic (1999) indicated the potential of data perturbation methods which was subsequently adopted by Agrawal and Srikant (2000).

B. Inaccurate Data

The data quality issue is more difficult to resolve. Inaccurate data are undetected by the individual until he or she experiences some associated repercussion, such as a denial of credit, or the withholding of a payment. It is also usually undetected by the organization, which lacks the personal knowledge necessary for the exposure of inaccuracies. The adoption of data quality management strategies by the organization, coupled with the expedient correction of any inaccuracies reported by individuals and intermittent data cleansing may go some way to resolving the dilemma. Other solutions are apparent (for example, data matching) but they may have unsatisfactory implications for privacy protection.

C. Legal solutions

Legal regulation of applied technology is currently one of the more pressing needs facing policy-makers. But how does one approach the development of what is needed? Legislative reform may be unsuitable as it is an unwieldy tool in a rapidly expanding technical environment.

Common law change is probably a more appropriate and suitable mechanism of legal regulation as it can be creatively applied to novel situations. The disadvantage of the common law, however, is that there needs to be a number of precedent court cases upon which to build common law principles, and litigation in this age of mediated dispute resolution and in confidence settlements is becoming a rare phenomenon. Furthermore, the common law's record on subjects such as the protection of privacy and dealing with the legal implications of applied technologies is weak. This is a rapidly expanding social and business landscape, and the law is failing to keep pace. Public awareness of the possibility of legal suits alleging negligence and defamation may possibly have some prophylactic effect upon the potential transgressions of contemporary technology.

Another weakness of legal regulation is the fact that the jurisdictional boundaries that determine the limits of our legal system were drawn up in ignorance of technological developments that render these boundaries virtually irrelevant, given the international structure of many organizations and the burgeoning international presence of an individual on the Internet.

If an Australian were to conduct a transaction and provide data for a multinational organization registered in Europe via a web site physically situated in the United States, which legal system governs the transaction and its legal repercussions? This is a legal dilemma not lost on international lawyers, and is one that does not readily admit of a simple solution.

XI. CONCLUSION

Data Mining provides efficient way to implement and utilize database but it is also obvious that if this data and data mining tools in wrong hands may pose severe problems. To implement effective privacy protection when applying data mining, it is not sufficient to focus on Privacy Preserving Data Mining methods and algorithms. In addition to this the whole business or governmental process in which data mining is used has to be taken into account. To accommodate these problems various solutions are suggested like data anonymizaion, legal solution and inaccurate data.

ACKNOWLEDGMENT

We the authors of this paper express our deep gratitude towards the CMR College for organizing such event and allowing us to share the platform for sharing our knowledge and views. We are also looking forward that in future more such events at international level will be held giving an opportunity to many more budding and renowned researchers.

REFERENCES

[1] Frawley and Piatetsky-Shapiro, 1996. Knowledge iscovery in Databases: An Overview. The AAAI/MIT Press, Menlo Park, C.A. Glymour, C., D. Madigan, D. Pregidon and P.Smyth, 1996.

Statistical nference and data mining. Communication of the ACM, pp: 35-41.

[2] Shams, K. and M. Frashita, 2001. Data Warehousing Toward Knowledge anagement. Topics in Health Information Management, 21: 3.

[3] Pawlak, Z. (1990). Rough sets. Theoretical Aspects of Reasoning about ata, Kluwer Academic Publishers, 1992

[4] Lin, T. Y. (1994), "Anamoly Detection -- A Soft Computing Approach", Proceedings in the ACM SIGSAC New Security Paradigm Workshop, Aug 3-5, 1994,44-53. This paper reappeared in the Proceedings of 1994 National Computer Security Center Conference under the title "Fuzzy Patterns in data".

[5] Lin, T. Y. (1993), "Rough Patterns in Data-Rough Sets and Intrusion Detection Systems", Journal of Foundation of Computer Science and Decision Support, Vol.18, No. 3-4, 1993. pp. 225-241. The extended version of "Patterns in Data-Rough Sets and Foundation of Intrusion Detection

[6] Systems" presented at the First Invitational Workshop on Rough Sets, Poznan-Kiekrz, September 2-4. 1992.

[7] Agrawal, R. and Srikant, R. (2000), Privacy-preserving data mining, in 'ACM SIGMOD Conference on the Management of Data', ACM, Dallas, 439–450.

❑❑❑