OBJECT'S SURVEY FOR OPTIMAL RESULTS IN SPATIAL DATABASE FOR USERS QUERY

M.Sharmila,
Asst.Professor,
Department of Computer Applications,
Valliammal college for women's, Chennai, India.

Abstract : The main aim of this paper is to think about the problem of finding out the probabilistic data to allow such data to be stored in legacy systems that accepts only deterministic input. Pre-existing web applications are correspondingly related with legacy system. Automated data analysis and enrichment methods like entity resolution, data extraction, and speech processing systems generates Probabilistic data. The intention is to produce a deterministic representation of probabilistic data of the end application built on deterministic data such data that optimizes the quality. Methods are triggers and selection queries are two various data handling tasks are used to find out the deterministic problem. Traditionally the top-1 selection or thresholding is used to find determinization performance for many applications. As an alternative, we expand a query-aware strategy and demonstrate its advantage over existing solutions through a complete empirical evaluation over real and synthetic datasets.

Index Terms **- Branch and bound algorithm, determination, data quality, uncertain data, query workload.**

I. INTRODUCTION

Data mining is the means of extracting data from a dataset for users to use it for various purposes. The purpose of such data plays a significant role in keyword searching. In data mining searching is a common activity. Searching for spatial objects from a spatial database has recently sparked enthusiasm among researchers. These encourage developing methods to get spatial objects. Users need to store their data in various existing web applications with the advancement of cloud computing and the web-based applications. Signals processing, data analysis/enrichment methods are frequently used by the user to produce data and also these process are takes place before store in web application. For example recent Digital Single Lens Reflex (DSLR) cameras support vision analysis to create tags such as indoors/ outdoors, scenery, landscape/portrait, etc. These cameras also have microphones to speak out a desired sentence. This type of result should be processed using a speech recognizer to produce a set of tags to be associated with an image [1][2].The Image with the set of tags/sub words are streamed by using wireless connectivity to Web applications. For Example in the application Flickr it could be used. To swing such data into web applications introduces a new challenge. The main disadvantages are such type of automatically generated content may confuse the user and also generates objects with probabilistic attributes. For example vision analysis may lead to produce the result with tags [2], [3]. Like a same way Automatic Speech Recognizer (ASR) may producean N-best list or confusion network of utterances [4] . So such types of probabilistic data could be "determinized" before stored in legacy web applications. The problem of mapping probabilistic data into the corresponding deterministic representation should be referred by the determinization issues.

Several such approaches should be followed for a determinization issues. Two basic methods they are most probable value / all the possible values of the attribute with non-zero probability could be used respectively. It is called the Top-1 techniques. For instnce a speech recognition system that produce a single answer/tag for each utterance can be resulted as using a top-1 strategy and threshold τ *method* . All the attribute values with a probability higher than τ. they often lead to suboptimal results so those types of approaches are being doubter to the end application.A better approach among those is design customized determinization strategies they often lead to suboptimal results which select a determinate representation that should optimize the quality of the endapplication. Consider as an example some type of web application supports triggers/alerts on a content which generated automatically. Examples of those varities of end-application to contain publish/subscribe systems like Google Alert, wherein users specify their contribution in the form of keywords (e.g., "California earthquake") and imply over metadata (e.g., "knowledge is video"). Based on the contributions details Google Alert should find all matching data items to the user. About California Earthquake that is to be published on Youtube which is consider as a video. The video consists of large number of tags that should be obtain using each of two that is automated vision processing and/or information extraction methods applied over reproduced speech. Like "California": 0.9,"earthquake":0.6, "election": 0.7, while the true tags of the video could be "California" and "earthquake" are may generate tags with probabilities using these types of automated tools. The determinization process helpful to associate the video with relevant tags. The user who are really interested in the specific video such as "California Earthquake" is declared while the other is not discontented by irrelevant data.Thus, in the above example, the determinization process could be used to reduced metrics like false positives and false negatives that viewed

from a determinized representation of data.From cameras along with tags images are uploaded automatically in Flickr application that may produced depend on image analysis or speech annotation.

Based on a photo tags Flicker supports effective retrieval. User may be interested in select determinate representation that optimizes set-based quality metrics for example F-measure insteadof reducing the false positives/negatives rates in such a type of application. The difficulty of finding datasets with probabilistic attributes (usually automatically generated by data analyses/enrichment) is the issue. Our way accomplishment a workload of triggers/queries to select the top deterministic representation for two various types of applications– triggers on generated content and another effective retrieval. Moreover in the past the issue of Determinization has not been explored widely. Now the leading research is how to give effective deterministic answers to a query (e.g. query of conjunctive selection) among probabilistic database. For a Determinization problem solution can not be applied straightforwardly. In this paper we studied Probabilistic data: that is related to our project. They determine over a probabilistic data that is how to determine answers to a query. Likewise the best deterministic representation of data is to use existing end-applications that take deterministic input. The contest in the two problem settings gives a way to many various challenges. It describes problem of uncertain objects to be cleaned, in order to retrieve the quality of query answers. However, aim is to improve quality of single query, while our aim is to optimize quality of overall query workload [6].

II. RELATED WORKS

[2]A statistical modelling approach is used to overcome the challenge for researchers in computer vision and content-based image retrieval. To train a dictionary of many statistical models each representing a concept the categorized images are used. Images of any given concept are regarded as example of process that characterizes the concept. The appearance of the image based on the characterizing process is computed to analyse the association between an image and the textual description of a concept.

[5]The modern data processing methods like entity resolution, data cleaning, information extraction, and automated tagging always generate results consisting uncertainty object.along with a measure of probability for instead values,each uncertain attribute are frequently captured in the form of a set of multiple mutually exclusive value. If outputted in such a form the lay end-user, and some end-applications, cannot interpret the results. So, the risk is how to convey such results for user to make it easy.for example,the user should be interested in attribute-value selection and object selection queries.Here the problem is,on top of such a probabilistic representation of maximizing the quality selection queries.The standard and commonly used set-based quality metrics are used to measure the quality.So determine the problem and overcome with effcient approaches which gives high-quality answers.Over existing the advantage of our approach the comprehensive empirical evaluation techniques.

[7]The detection of duplicate tuples is an important task in data integration and data cleaning. There are many techniques exist to identify, the merging or elimination of duplicates tuples are difficult task that relies on ad-hoc and manual solutions. It proposes effective approaches that produce detailed query answering over duplicated data, in the clean database. Database contain duplicate store turn each answer with the probability that the answer is in the clean database. To the semantics of duplication our rewritten queries are sensitive and help a user understand which query answers are repeated in the clean database. The probabilities are produced in the way the semantics that we adopt is independent, but can effectively exploit them while answering query. Without external knowledge that depends on search data base tuple with a probability, proposed a method, using tuple summaries, it automates this task. The rewriting query does not introduce a significant overhead in query execution time. This task belongs to the Conquer project at the University of Toronto; it gives importance on the efficient management of inconsistent and dirty databases.

[6]The notion of a consensus world (or a consensus answer) which is a deterministic world (answer) which reduces the expected distance to the possible worlds (answers).A generalization of the well-studied inconsistent information aggregation problems (e.g. rank aggregation) to probabilistic databases can be seen as a problem.(Select-Project-Join) SPJ queries, Top-k ranking queries, group-by aggregate queries, and clustering are various types of queries existing whith same problem. we obtain polynomial time optimal or approximation algorithms for computing the consensus answers (or prove Non deterministic Polynomial time-NP hardness) for various distance metrics, and xor tree model, that significantly generalizes previous probabilistic database method such as x-tuples and block-independent disjoint models, and independent in interest. The problem is to find out a single representative answer to a query in probabilistic databases. Top-k query is used to overcome this problem with the results in semantic query answer.

[8]An important and challenging problem for many domains including, computational biology and computer vision, natural language understanding is approximate MAP inference in graphical models. Current state-of-theatre methodologies don't convex relaxations of these issues then provide weak running time guarantees. It develop an approximate inference algorithm for efficient and with stuffed theoretical guarantees. It results in required time -accurate solution of the convex relaxation, it generates via state of-the-art techniques.

[9] A high likelihood indicates a strong association. In our experimental implementation, the two-dimensional multiresolution hidden Markov models 2D MHMMs is implemented and tested the ALIP (Automatic Linguistic Indexing of Pictures) system on a photographic image database which consists of about 600 different concepts, each has about 40 training images. It resulted with indexing of photographic images.

III. METHODOLOGY

Fig (1) flow chart explains the methodology that has been developed for the user required results. It will describe the functions and features of our system and its requirements. The system generates image details from the dataset, and then it provides results with predefined customer ratings and minimum inter object distance.

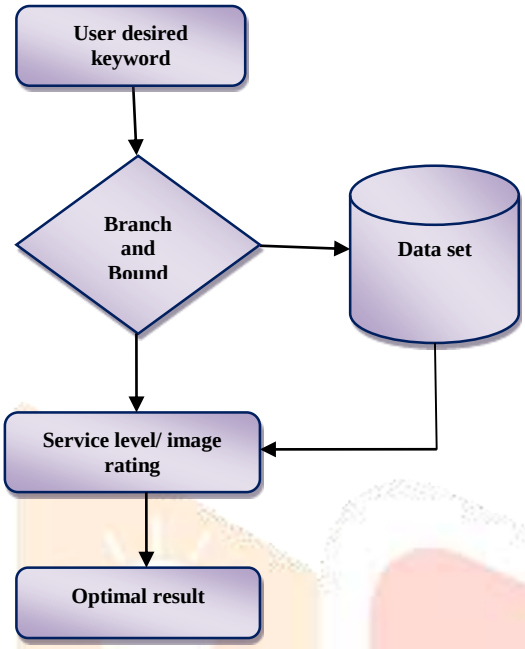

Fig (1).Methodology Flowchart

Input of keywords is being obtained first from the users. System (consists of spatial database) processes user provided input with branch and bound algorithm. This algorithm identifies a sum of independent objects which are nearest to the user required image along with matching of keywords with object associated texts. Hence output resulted from this process is efficient in terms of desired keywords as well as objects nearest to the desired image. Output will be delivered to the user as well as other system as applicable. Effective and efficient results can be obtained through a well-defined Output.

Initially, a query is entered by the user which is given to the system. The system consists of a spatial database and uses the Keyword- Nearest Neighbor Expansion algorithm. The algorithm finds a number of independent objects, in which every object is nearer to the location of the query and the keywords associated are related to the collection of query keywords [10] [11]. An output which represents the information as a graphical representation to the end user demands is efficient. The results of processing are conveyed to the users and to other system through outputs, in every system an efficient and smart output plan helps in improving the user's relationship with the system which further helps in better decision-making.

IV. ENHANCEMENT AND RESULT

A quality output presents the information clearly according to the requirements of the end user.

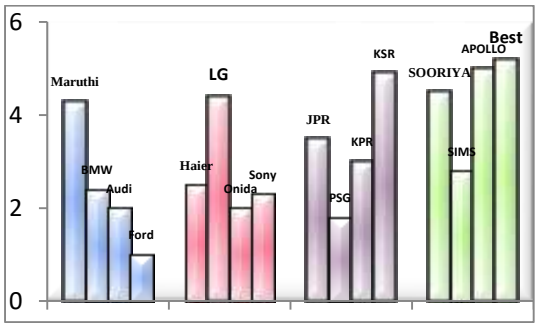

Fig (2).Result-Example

Through outputs, results of processing are communicated to the end users and to other system. For immediate need of information and also the hard copy output the output is designed. It is direct source and important information to the end user. Skilled and intelligence of output design develops the relationship of system to help end user decision-making. The following objectives should

accomplish for the informative output. Present information regarding past, current and future activities status. Fig (2) shows the graphical method of result presented.

V. CONCLUSION

This survey is considered the problem of finding uncertain objects. Such data to be saved in pre-existing systems, which take only deterministic input. The aim is to produce a deterministic representation that generates the quality of answers to queries/triggers that execute over the deterministic data representation. We have proposed efficient algorithms that are orders of faster than the enumeration based optimal solution but gets almost the similar quality as the optimal solution. We plan to explore determinization method in the context of applications as a future work, wherein users are also interested in retrieving objects in a graphical way.

REFERENCES

[1] D. V. Kalashnikov, S. Mehrotra, J. Xu, and N. Venkatasubramanian, "Asemantics-based approach for speech annotation of images," IEEE Trans.Knowl. Data Eng., vol. 23, no. 9, pp. 1373–1387, Sept. 2011.

[2] J. Li and J. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," IEEE Trans. Pattern Anal. Mach. Intell.,vol. 25, no. 9, pp. 1075–1088, Sept. 2003.

[3] C. Wangand, F. Jing, L. Zhang, and H. Zhang, "Image annotation refinement using random walk with restarts," in Proc. 14th Annu. ACM Int. Conf. Multimedia, New York, NY, USA, 2006.

[4] B. Minescu, G. Damnati, F. Bechet, and R. de Mori, "Conditional use of word lattices, confusion networks and 1-best string hypotheses in a sequential interpretation strategy," in Proc. ICASSP, 2007.

[5] R. Nuray-Turan, D. V. Kalashnikov, S. Mehrotra, and Y. Yu, "Attribute and object selection queries on objects with probabilistic attributes," ACM Trans. Database Syst., vol. 37, no. 1, Article 3,Feb. 2012.

[6] J. Li and A. Deshpande, "Consensus answers for queries over probabilistic databases," in Proc. 28th ACM SIGMOD-SIGACTSIGART Symp. PODS, New York, NY, USA, 2009.

[7] R. Cheng, J. Chen, and X. Xie, "Cleaning uncertain data with quality guarantees," in Proc. VLDB, Auckland, New Zealand, 2008.

[8] V. Jojic, S. Gould, and D. Koller, "Accelerated dual decomposition forMAP inference," in Proc. 27th ICML, Haifa, Israel, 2010.

[9] D. Sontag, D. K. Choe, and Y. Li, "Efficiently searching for frustrated cycles in map inference," in Proc. 28th Conf. UAI,2012.

[10] Yufei Tao Cheng Sheng, "Fast Nearest Neighbor Search with Keywords" IEEE IEEE Trans Knowl. Data Eng., Nov 29, 2014 .

[11] N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles.In Proc. of ACM Management of Data (SIGMOD), pages 322–331, 1990.

[12] P. Andritsos, A. Fuxman, and R. J. Miller, "Clean answers over dirty databases: A probabilistic approach," in Proc. 22nd ICDE,Washington,

[13] D. Carmel et al., "Static index pruning for information retrieval systems," in Proc. 24th Annu. Int. ACM SIGIR, New Orleans, LA, USA, 2001.

[14] J. Li and M. Sun, "Scalable term selection for text categorization," in Proc. EMNLP-CoNLL, Prague, Czech Republic, 2007.

[15] X. Li, Y. Wang, and A. Acero, "Learning query intent from regularized click graphs," in Proc. 31st Annu. Int. ACM SIGIR,Singapore, 2008.

[16] E. Pitler and K. Church, "Using word-sense disambiguation methods to classify web queries by intent," in Proc. Conf. EMNLP, Singapore, 2009.

[17] A. Ashkan, C. L. Clarke, E. Agichtein, and Q. Guo, "Classifying and characterizing query intent," in Proc. 31th ECIR, Toulouse, France, 2009.

[18] J. Teevan, S. Dumais, and D. Liebling, "To personalize or not to personalize: Modeling queries with variation in user intent," in Proc. SIGIR, Singapore, 2008.

[19] I. Bordino, C. Castillo, D. Donato, and A. Gionis, "Query similarity by projecting the query-flow graph," in Proc. 33rd Int. ACM SIGIR, Geneva, Switzerland, 2010.