

MI TECHNIQUES USING QUERY DETERMINIZATION

Bhavya N Javagal¹, Manhitha G², Ashwini G³, Kavya K C⁴

¹Assistant Professor, Computer Science Department, TJIT, Karnataka, India.

²Pursuing B.E, Computer Science Department, TJIT, Karnataka, India.

³Pursuing B.E, Computer Science Department, TJIT, Karnataka, India.

⁴Pursuing B.E, Computer Science Department, TJIT, Karnataka, India.

Abstract

The data should be in probabilistic form because the legacy system accepts only the probabilistic data for a given image by the user, legacy systems are the systems which will be used across the industries and cannot be replaced. So, the data has to be in the deterministic form to be understandable by the systems. The system has to generate the dataset. Through the generated dataset the real tags are termed as the Candour tags which are associated to images. Based on these query related tags will be generated and the deterministic output is produced. The main aim is to produce the deterministic showcase of an probabilistic data that effectively uses the application at the user end that is based on the deterministic data. Determinization problem can be solved in different contexts of data processing such as alerts and selection queries.

Keywords: Determinization, Probabilistic Data, Query, and Candour Tags.

1. INTRODUCTION

Through booming cloud computing and excessive use of web apps, the end- users tend to reserve and access the information which often be stored in pre existing application of web. Generally the data presented by the user will get automatically generated by processing signals and analyzing the data in prior to storing the data across the web.

For instance, cameras now-a-days support analysis of vision through which they generate tags like scenery cameras have microphones installed, so what users speaks that can be used to generate a associated tags. As internet is used across the legacy systems but the user is not very familiar with the legacy system so there is a need to determinize the user input. The basic strategies used are Top-1 and All techniques which selects the highest probable values. For example: For the speech recognition, a single answer can be generated that can be expressed using Top-1 approach. The second approach will use a threshold value and selects only the values higher than the selected threshold.

Let us consider an application that supports alerts where in the user enters the keyword on the search engine in the form of keywords which in turn predicates over the data already stored and the search engine will return the data based on user keywords. For example like user enters the key FLOODS in Arunachal Pradesh. The search engine will use automatic tools which produces tags with highest probability (Ex:

“FLOODS”:0.8 “Arunachal Pradesh”:0.6). The process of determinization should result in a video that contain proper tags such that, the user searching for a video should receive relevant data. In the above example the process of determinizing is maximizing the metrics such as false positive and false negative which is calculated from the data.

In this paper, we try to solve the problem associated with determinizing of data using probabilistic values. Our paper makes use of the alerts or queries to select from the deterministic representation of a given scenario. The determinization problem has not been countered properly in recent past. It is closely related [4][5][6] which explains how to answer a query over a probabilistic database.

Table no.1: Notations

notations	meanings
U	Vacillating objects
$X=\{x_1,x_2,\dots,x_x\}$	Set of uncertain/vacillating tags for U
$Y=\{y_1,y_2,\dots,y_x\}$	Probability that X_i is in the candour tags
$W=\{w_1,w_2,\dots,w_w\}$	Query of User
$Q=\{q_1,q_2,\dots,q_q\}$	Set of query terms of query W
R_w	Weight or frequency of query W
f_w^+	Cost of false positive for query W
f_w^-	Cost of false negative for query W
T_w	Set of objects that satisfy W based on the candour tags
M_w	Set of objects that satisfy W according to the algorithm

2. PRELIM

Let us look into the notations being used here. The notations are being summed up in the Table no 1: Notations . Illustration is shown with the use of a Fig.1.ex.



Fig 1. Example

2.1 Modules

Objects

An Vacillate object U are tagged through probabilistic traits. For example the result of an record linkage are tagged as probabilistic tags and are produced, formally object U is associated with

1. $X = \{x_1, x_2, \dots, x_n\}$ which are the group of Vacillate tags which are produced by some techniques.
2. $Y = \{y_1, y_2, \dots, y_n\}$ where each $y_i \in Y$ is the probability that corresponds uncertain tags $X_i \in X$ belongs to Candour tags T.
3. This object model is also called as label model. The group of Vacillate tags of U is $X = \{mountain, beach, horse\}$. The corresponding probabilistic group is $Y = \{0.35, 0.60, 0.70\}$. These tags are handled using XOR model.

Candour tags

The Candour tags T are the tags that are associate with object. Candour tags corresponding to the pronounced words in the speech annotated images. The Candour tags are not known to the methods. Hence they are not piece of vacillating objects.

Best Benchmark are defined by these annotations. In the above fig $T = \{horse, beach\}$ uncertainty is a result obtained by some methods, the group of vacillating tags U with the group of Candour tags T.

Divination

Divination process of an thing U is selected for divination representation which is reserved in the old systems which take in only probabilistic information.

Group of divination tags $M = \{X_1, X_2, \dots, X_n\}$. for example object in an Divination representation is $M = \{mountains, beach\}$, when the phrase is not crystal-clear we subscript "U" in variables X, T and M. Where M re-written as "X of object U"

Query of user

Query from the user is a group of queries $W = \{w_1, w_2, \dots, w_n\}$ that will be portrayed in the divination representation. We observe that each query $w \in W$ is an conjoin query.

1. $Q = \{q_1, q_2, \dots, q_n\}$ stores a set of terms associated with $W = q_1 \wedge q_2 \wedge q_3 \dots q_n$ for an deterministic representation M_0 of an thing U, the thing satisfies a conjunctives query W if Q subset of M_0 .

2. $R \in [0, 1]$ is represented into query mass that corresponds to related prevalence of W in workload.

T is subscripted for variable Q and R to represent the query called as R_w "Weight of query W".

Effective BB factor

BB factor technique is effective when compared to others. This approach discovers answers in the systematic enumeration so that the answer group with minimum cost is the first preference.

Here in this module the input is given as the set of all query W along with the objects U, where the output obtained is effectively gives the set of required answer set for the respective query along with the object present in that particular image.

3. RELATED WORKS

Image Annotation For the webpage and non web image there is textual information related to it Ex: URL, surrounding text. The stop words are removed (ex: and, the etc..) rest of words are given the score and the highest values are reserved for annotation of image. Results in the removing almost noisy words which is relevant to the image and the drawback is it does not concentrate more on the textual information of image give when the annotation is finalized then the original details is ignored.

Series of methods have been devised for fetching content based image, and automatic image by meaning which still remains a problem. Our motive is based on the single branch tree model. The branches of the tree of an image, termed as structured elements, are devised and grouped based on their low level features such as color, spatial location, coarseness and shape.

Query workload By the concept of the COLT (continuous online tuning) of the workload to the database system. It set the model for the incoming queries in workload and select the query which gives the best performance for the observed workload in the given space. Results in the continuously monitor of incoming queries and for each gives the resource to execute the query, there is a framework for self-tuning for the query. The drawback is it requires the plenty of CPU resources and cause serious slowdowns.

Query predictions Query predictions is the essential part of search engine. It cracks a sequence of same queries for improvise the search effectiveness. This proposed system focuses on new query prediction method which relays on themes and context. It measures same queries from the semantic degree and a new same query benchmark is offered. In the new query prediction method, theme and query context has been considered. People depend on the search engines but now, the use of the custom search engines for data finding is still hard to make the end-user happy.

Queries over probabilistic database Finding the best query answer using the probabilistic database is addressed here. For this purpose they have proposed the information of an full agreement answer, that minimizes the distance of the possible answer. It is generally seen as the unstable

4. CONCLUSION

The limelight is on the case of divination of vacillating objects are handled which are stored only in already existing systems as Flickr, Picasa etc which captures only the deterministic inputs. Here the main aim is to produce the deterministic demo of the best answers to the queries that are implemented over the deterministic data representation. If algorithms are planned effectively then they can be faster to retrieve the details based on the required optimal solution which are obtained as same as the required solution. As the future work, we can look into exploring the divination methods in the application vibe where the end-user will be interested to accumulation of problems that is present in the probabilistic database. Various types of queries such as select-project-join query, top-k query, and group by query etc are considered. For calculation of list metrics polynomial optimal time or approximate algorithm are used to compute full agreement are most of the results are for generating probabilistic database model called as XOR model, which generalizes probabilistic database models such as x-tuples and block independent.

Cleaning uncertain data The inappropriate data are distributive in application like general data collection and integration , sensor monitoring location based application etc. To obtain answer with statistical agreement, applications used probabilistic database can be used to store inappropriate data and querying facilities. The limited amount of resources is used to clean the database. In order to achieve the best improvement in the quality of query answers the group of vacillating objects is neat. Explore the retrieval of the objects in the grade or ranked system or on the percentage basis.

REFERENCES

- [1] S. Bhatia, D. Majumdar, and P. Mitra, "Query suggestions in the absence of query logs," in Proc. 34th Int. ACM SIGIR, Beijing, China, 2011.
- [2] D. V. Kalashnikov, S. Mehrotra, J. Xu, and N. Venkatasubramanian, "A semantics-based approach for speech annotation of images," IEEE Trans. Knowl. Data

Eng., vol. 23, no. 9, pp. 1373– 1387, Sept. 2011.

[3] 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.

[4] Jie Xu, Dmitri V. Kalashnikov, and Sharad Mehrotra, Member, IEEE "Query Aware Determinization of Uncertain Objects" IEEE Transactions on Knowledge and Data Engineering (Volume:27 , Issue: 1)207 – 221 04 November 2013

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

[6] K. Schnaitter, S. Abiteboul, T. Milo, and N. Polyzotis, "On-line index selection for shifting workloads," in Proc. IEEE 23rd Int. Conf. Data Eng. Workshop, Istanbul, Turkey, 2007.

[7] 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.

[8]P.Jhancy, K.Lakshmi, Dr.S.Prem Kumar "Query Aware Determinization of Uncertain Objects" in International Journal of Computer Engineering In Research Trends, Andhra Pradesh, India.