International Journal of Engineering and Techniques - Volume 4 Issue 2, Mar-Apr 2018

RESEARCH ARTICLE OPEN ACCESS

Travel Route Recommendation Using GA Optimization

Kanchana J.S, Aarthy M.G, Deekshitha Rani T.R.K, Jeeva Lakshmi A.D
[1,2,3,4] (Information Technology, K.L.N.College of Engineering, Sivagangai)

Abstract:

With the favour of several websites, the stakeholders can share their photos and check-in records during their trips. In the perception of several users historical mobility records in several websites by finding travel experiences to facilitate trip planning. While planning a trip, the stakeholders always have explicit preferences regarding the trips. Instead of restricting users to limited query options such as locations, activities, or time periods, this paper considers arbitrary text descriptions as keywords about personalized requirements. Even though, a different and representative set of recommended travel routes is needed. To meet the need for automatic trip organization, the system claim that more features of Places of Interest (POIs) should be extracted. Therefore, this paper proposes about keyword based representative travel route recommendation using Depth First Search Algorithm and Genetic Algorithm. Based on the precision value top ten keywords are extracted and possible routes for the locations are identified using Depth First Search Algorithm. Optimal solution for the routes will be generated using Genetic Algorithm. To provide benefiting query results, this paper proposes genetic algorithm with improved performance.

Keywords -– **check-in records, Travel routes, Place of Interests (POI), Optimal solution**.

I. INTRODUCTION

Web mining is the process of using data mining techniques and algorithms to extract information directly from the Web by extracting it from Web documents and services, Web content, hyperlinks and server logs. The goal of Web mining is to look for patterns in Web data by collecting and analyzing information in order to gain insight into trends, the industry and users in general. In this paper, it proposes route optimization technique using Evolutionary algorithm. It is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs ofWeb-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

Web usage mining itself can be classified further depending on the kind of usage data considered. Web Server Data in which the user logs are collected by the Web server. Typical data includes IP address, page reference and access time. Application Server Data provides the commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs. Application Level Data provides the new kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above.

Depth-first search (**DFS**) is an algorithm for traversing or searching tree or graph data structures. One starts at the root (selecting some arbitrary node as the root in the case of a graph) and explores as far as possible along each branch before backtracking. Using this algorithm, the possible set of candidate routes are generated for the top ranking keywords. For applications of DFS in relation to specific domains, such as searching for solutions in artificial intelligence or web-crawling, the graph to be traversed is often either too large to visit in its entirety or infinite (DFS may suffer from non-termination). In such cases, search is only performed to a limited depth; due to limited resources, such as memory or disk space, one typically does not use data structures to keep track of the set of all previously visited vertices. When search is performed to a limited depth, the time is still linear in terms of the number of expanded vertices and edges (although this number is not the same as the size of the entire graph because some vertices may be searched more than once and others not at all) but the space complexity of this variant of DFS is only proportional to the depth limit, and as a result, is much smaller than the space needed for searching to the same depth using breadth-first search. For such applications, DFS also lends itself much better to heuristic methods for choosing a likely-looking branch.

Genetic Algorithm (**GA**) is meta-heuristic inspired by the process of natural selection that belongs to the larger class of Evolutionary Algorithms (EA). Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, cross over and selection. Using these processes the optimal solution will be generated from the set of candidate routes. The evolution usually starts from a population of randomly generated individuals, and is an iterative process, with the population in each iteration called a generation. In each generation, the fitness of every individual in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's genome is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

A typical genetic algorithm requires a genetic representation of the solution domain and a fitness function to evaluate the solution domain. A standard representation of each candidate solution is as an array of bits. Arrays of other types and structures can be used in essentially the same way. The main property that makes these genetic representations convenient is that their parts are easily aligned due to their fixed size, which facilitates simple crossover operations. Variable length representations may also be used, but crossover implementation is more complex in this case. Tree-like representations are explored in genetic programming and graph-form representations are explored in evolutionary programming; a mix of both linear chromosomesand trees is explored in gene expression programming.

II. RELATED WORK

In his paper, Yu-Ting Wen represented Efficient Keyword-aware Representative Travel Route framework that uses knowledge extraction from users' historical mobility records and social interactions. Explicitly, this system has designed a keyword extraction module to classify the POI-related tags, for effective matching with query keywords. This System designed route reconstruction algorithm to construct route candidates that fulfill the requirements [1].

Hsieh told in his paper regrading Location-based services allow users to perform check-in actions, which not only record their geo-spatial activities, but also provide a plentiful source for data scientists to analyse and plan more accurate and useful geographical recommender system. This paper presents a novel Time-aware Route Planning (TRP) problem using location check-in data. The central idea is that the pleasure of staying at the locations along a route is significantly affected by their visiting time [3].

In this paper, Trajectory search has long been an attractive and challenging topic, in which Zheng blooms various interesting applications in spatial-temporal databases. This Paper studies a new problem of searching trajectories by locations, in which context the query is only a small set of locations with or without an order specified, while the target is to find the *k* Best-Connected Trajectories (*k*-BCT) from a database such that the *k*-BCT best connect the designated locations geographically [6].

With the advent of location-based social networking services (LBSNs), W.C. Peng represents travel planning and location-aware information recommendation based on LBSNs have attracted much research attention. This paper shows the study of the impact of social relations hidden inLBSNs, the social influence of friends. This system proposes a new social influence-based userrecommender framework (SIR) to discover the potential value from reliable users (i.e., Closefriends and travel experts) [4].

In this paper with the availability of user check-in data in large volume from the rapid growing location-based social networks (LBSNs) enables a number of important location-aware services.A. Sun focus on the problem of time-aware POI recommendation, which aims at recommending a list of POIs for a user to visit at a given time. To exploit both geographical and temporal influences in time aware POI recommendation, the system proposed the Geographical-Temporal influences Aware Graph (GTAG) to model check-in records, geographical influence and temporal influence [8].

III. PROPOSED WORK

The proposed approach is used to generate the optimized route for the top ranking keywords. In this work, GATR is presented. Given an LBSN dataset, the system first analyse the tags of each POI to determine the semantic meaning of the keywords, which are classified into (i) Geo-specific keywords, (ii) Temporal keywords, and (iii) Attribute keywords according to their characteristics. Furthermore, the system derive the feature scores of the POIs and generate proper candidate travel routes. Geo specific keywords are extracted from some tags which are specific to a location, which represents its spatial nature. To quantify the geo-specificity of a tag, an external database identifies geo-terms in the overall tag set and then the tag distribution on the map rates the identified geo-terms. Temporal keywords are extracted from the tags which are specific to a time interval, which represents its temporal nature. To quantify the temporal-specificity of a tag, time distribution on a tag rates the identified temporal-terms. Using the time distribution of tags, we can find tags associated with a specific time interval like 'sunset'. Attribute keywords are extracted by considering tags frequently associated with a POI. From the extracted keywords DFS algorithm is used to generate the possible set of candidate routes. From the generated set of recommended routes the optimal solution will be identified by applying GA.The approach consists of Data Pre-processing, Keyword Extraction, DFS-Route Generation and GA-Route Optimization.

DATA PREPROCESSING

Data Pre-processingis used for removing the noisy, incomplete and inconsistent data. The Data is normalized, aggregated and generalized.To eliminate irrelevant and redundant data in the data setis done by collecting real-world LBSN datasets from online sources in text file format.The text file format files are converted into Database table to determine the semantic meaning of the keywords. The keywords are classified into (i) Geo-specific keywords, (ii) Temporal keywords and (iii) Attribute keywords.

KEYWORD EXTRACTION

Extraction of keywords first computes the spatial, temporal and attributes scores for every keyword in the corpus. The classified keywords used to provide an interface for users to specify query ranges and preference-related keywords.The proper candidate travel routes are generated by the feature scores of the keywords.

DFS ROUTE GENERATION

The DFS algorithm is a recursive algorithm that uses the idea of backtracking. It involves exhaustive searches of all the nodes by going ahead, if possible, else by backtracking. It is based on the procedure proposed to generate possible routes. The new candidate routes are constructed by combining the subsequence of trajectories. The featured dataset is used to recommend a set of travel routes that connect user specific keywords.

```
DFS(G, u)
        u.visited = true
for each v ϵ G.Adj[u]
        ifv.visited == false
DFS(G,v)
Init( ) {
For each u ϵ G
        u.visited = false
For each u ϵ G
DFS(G,u)   }
```

GA ROUTE OPTIMIZATION

Genetic Algorithm has a great potential for search problems.Further, unlike many search algorithms which perform a local or greedy search and GA performs a global search. The proposed travel route recommendation system with GA is used for optimal travel route search process. The results show that the proposed GATR is effective and beats other baselines and state-of- the-art methods in terms of route prediction accuracy. Tournament Selection is chosen for identifying the route by selecting the individual population.

TOS_pseudo-code

```
For i=1 to ndo{
for j=1 to n do{
  p1 = t[j]
for m=1 to n do{
  p2 = t[j+m]
if(p1 > fp2) then select p1 else select p2;
}end for m
}end for j
}end for i
```

IV. EXPERIMENTAL RESULTS

The performance is measured for Category and Similarity by computing the Ranked Queries. Table 1 shows the Comparison of Category and Similarity for Keyword Route Travel Recommendation (KRTR) and Genetic Algorithm Travel Recommendation (GATR) using the formula

$$\text{Category-Similarity} = \frac{\text{No. of overlapped category}}{\sqrt{\text{No. of Category1} * \text{No. of Category 2}}}$$

TABLE 1 COMPARISION OF CATEGORY SIMILARITY

S.NO	RANKED QUERIES IN Ks	KRTR	GATR
1	0	0.00	0.00
2	5	0.00	0.00
3	10	0.00	0.00
4	15	0.00	0.00
5	20	0.00	0.00
6	25	0.40	0.45
7	30	0.50	0.50
8	35	0.50	0.55
9	40	0.55	0.60
10	45	0.65	0.75
11	50	0.70	0.95
12	55	0.85	1.00

Fig 1 shows the Performance analysis by comparing the Category-Similarity which is tabulated in Table 1. The optimized travel routes are recommended by GA Tournament Selection process.The reconstructed travel routes from original routes are generated for GA optimization. The category similarity of recommended routes are calculated for the performance accuracy. As per the ranked queries the Category-Similarity value is calculated for the computed locations using Genetic Algorithm. The results show that the proposed GATR is effective and beats other baselines and state-of- the-art methods in terms of route prediction accuracy.

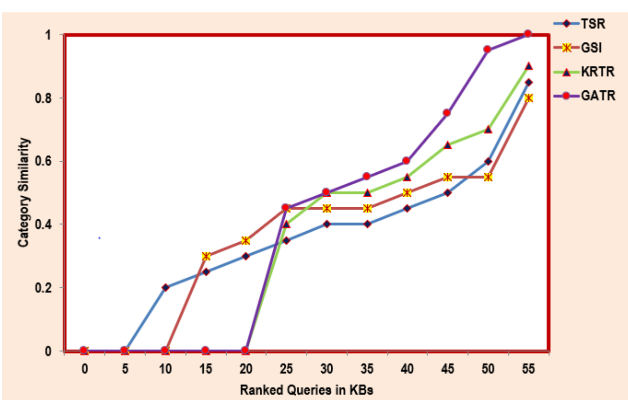

FIG 1 Performance Analysis

V. CONCLUSION

In this paper the travel route recommendation problem is studied and developed a GATR framework to suggest travel routes with a specific range and a set of user preference keywords. These travel routes are related to all or partial user preference keywords, and are recommended based on (i) the attractiveness of the POIs it passes, (ii) visiting the POIs at their corresponding proper arrival times, and (iii) the routes generated by influential users. This paper proposes a novel keyword extraction module to identify the semantic meaning and match the measurement of routes, and have designed a route reconstruction algorithm to aggregate route segments into travel routes in accordance with query range and time period. The experiment results demonstrate that GATR is able to retrieve travel routes that are interesting for users, and outperforms the baseline algorithms in terms of effectiveness and efficiency. Due to the real-time requirements for online systems, we aim to reduce the computation cost by recording repeated queries and to learn the approximate parameters automatically in the future.

VI. REFERENCES

[1] Yu-Ting Wen, Jinyoung Yeo, Wen-Chih Peng, Member, IEEE, and Seung-Won Hwang, "Efficient Keyword-Aware Representative Travel Route Recommendation" in IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 8, AUGUST 2017.

[2] H-P. Hsieh, C-T. Li, and S.-D. Lin, "Exploiting large-scale checkin data to recommend time-sensitive routes", in Proc. ACM SIGKDD Int. Workshop Urban Comput., 2012, pp. 55–62.

[3] H-P. Hsieh and C-T. Li,"Mining and planning time-aware routes from check-in data", in Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage., 2014, pp. 481–490.

[4] Y-T. Wen, P.-R. Lei, W.-C. Peng, and X.-F. Zhou, "Exploring social influence on location-based social networks," in Proc. IEEE Int. Conf. Data Mining, 2014, pp. 1043–1048.

[5] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp.899–908.

[6] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng, and X. Xie, "Searching trajectories by locations: An efficiency study," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2010, pp. 255–266.

[7] X. Cao, G. Cong, and C. S. Jensen, "Mining significant semantic locations from GPS data", Proc. VLDB Endowment, 2010, vol. 3, no. 1/2, pp. 1009–1020.

[8] Q. Yuan, G. Cong, and A. Sun, "Graph-based point-of-interest recommendation with geographical and temporal influences", in Proc. 23rd ACM Int. Conf. Conf. Inf. Knowl. Manage., 2014, pp. 659– 668.

[9] M. Ye, X. Liu, and W-C. Lee, "Exploring social influence for recommendation: A generative model approach", in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp.671– 680.

[10] J. Ye, Z. Zhu, and H. Cheng, "What's your next move: User activity prediction in location-based social networks" in Proc. SIAM Int. Conf. Data Mining, 2013, pp. 171–179.

[11] A. Sadilek, H. Kautz, and J. P. Bigham, "Finding your friends and following them to where you are," in Proc. 5th ACM Int. Conf. Web Search Data Mining, 2012, pp. 723–732.