AN EFFICIENT MINING METHODOLOGY FOR TRAVEL ROUTE INTERPRETATION

Mr. Marri Rathna Kumar[1], Mrs. V. Tejaswini[2*]

1 Final Year MCA Student, QIS College of Engineering and Technology, Ongole

2* Assistant Professor, MCA Dept., QIS College of Engineering and Technology, Ongole

Abstract: *An advancement in online networking (e.g., Facebook and Flicker), clients can without much of a stretch offer their registration records and photographs amid their treks. In perspective on the colossal number of client chronicled portability records in online life, we mean to find head out encounters to encourage trip arranging. When arranging a trek, clients dependably have explicit inclinations with respect to their outings. Rather than confining clients to constrained question choices, for example, areas, exercises or timeframes, we consider discretionary content portrayals as watchwords about customized necessities. Besides, a different and delegate set of prescribed travel courses is required. Earlier works have explained on mining and positioning existing courses from registration information. To address the issue for programmed trip association, we guarantee that more highlights of Places of Interest (POIs) ought to be extricated. In this way, in this paper, we propose a proficient Keyword-mindful Representative Travel Route structure that utilizes learning extraction from clients' chronicled portability records and social connections. Unequivocally, we have planned a catchphrase extraction module to arrange the POI-related labels, for powerful coordinating with inquiry watchwords. We have additionally structured a course remaking calculation to develop course applicants that satisfy the necessities. To give befitting inquiry results, we investigate Representative Skyline ideas, that is, the Skyline courses which best depict the exchange offs among various POI highlights. To assess the viability and productivity of the proposed calculations, we have directed broad investigations on genuine area based interpersonal organization datasets, and the test results demonstrate that our strategies do for sure exhibit great execution contrasted with cutting edge works.*

Keywords: ***Data Markets, Truthfulness and Privacy preserving, RECSdataset.***

I. INTRODUCTION

Location-Based social network (LBSN) services allow users to perform check-in and share their check-in data with their friends. In particular, when a user is traveling, the check-in data are in fact a travel route with some photos and tag information. As a result, a massive number of routes are generated, which play an essential role in many well-established research areas, such as mobility prediction, urban planning and traffic management. In this paper, we focus on trip planning and intend to discover travel experiences from shared data in location-based social networks. To facilitate trip planning, the prior works in [1], [2], [3], [4], [5] provide an interface in which a user could submit the query region and the total travel time. In contrast, we consider a scenario where users specify their preferences with keywords. For example, when planning a trip in Sydney, one would have "Opera House". As such, we extend the input of trip planning by exploring possible keywords issued by users. However, the query results of existing travel route recommendation services usually rank the routes simply by the popularity or the number of uploads of routes. For such ranking, the existing works [6], [7], [8] derive a scoring function, where each route will have one score according to its features (e.g., the number of Places of Interest, the popularity of places). Usually, the query results will have similar routes. Recently, [9][10][11], aimed to retrieve a greater diversity of routes based on the travel factors considered. As high scoring routes are often too similar to each other, this work considers the diversity of results by exploiting Skyline query.

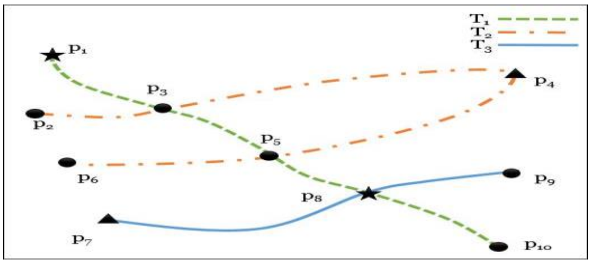


Fig.1. Keyword-aware travel routes query running example.

In this paper, we develop a Keyword-aware Representative Travel Route (KRTR) framework to retrieve several recommended routes where keyword means the personalized [12]-[15], requirements that users have for the trip. The route dataset could be built from the collection of low-sampling check-in records.

Definition 1. (Travel route): Given a set of check-in points recorded as a series of travel routes, each check-in point represents a POI p and the user's checked-in time t. The check-in records were grouped by individual users and ordered by the creation time.

Each user could have a list of travel routes{ }T =

T\{ $_0$, T_1, ... \}, where T_0 = (p_0, t_0), (p_1, t_1), ..., (pi, ti), T_1 = (pi_{+1}, ti_{+1}), (pi_{+2}, ti_{+2}), ... and ti_{+1} ti is greater than a route-split threshold. We set the route-split threshold to one day in this paper.

Table: Example of Route dataset

Tid	Uid	Pid	keyword	time	POI score vector
*T*1	*u*1	*p*1	Opera House	10:00	(0.04, 0.2)
*T*1	*u*1	*p*3	Bar	12:00	(0.25, 0.2)
*T*1	*u*1	*p*5	Bar	15:30	(0.2, 0.8)
*T*1	*u*1	*p*8	Opera House	17:30	(0.04, 0.3)
T 1	*u* 1	*p*10	Bar	19:00	(0.04, 0.2)
*T*2	*u*2	*p*2	Bar	10:30	(0.02, 0.2)
*T*2	*u*2	*p*3	Bar	12:30	(0.25, 0.2)
*T*2	*u*2	*p*4	Sunset	17:00	(0.05, 0.2)
*T*2	*u*2	*p*5	Bar	19:00	(0.2, 0.8)
*T*2	*u*2	*p*6	Bar	19:30	(0.25, 0.8)
*T*3	*u*3	*p*7	Sunset	18:30	(0.4, 0.8)
*T*3	*u*3	*p*8	Opera House	19:30	(0.04, 0.3)
*T*3	*u*3	*p*9	Bar	20:00	(0.1, 0.1)

The above table summarizes the route information. For ease of illustration, each POI is associated with one keyword (though our model can support multiple keywords) and a two-dimensional score vector (each dimension represents the rank of a feature). Assume a tourist plans a date with a set of keywords [**"**Whisky**"** **"**Sydney Cove**"** **"**Sunset**"**]. First, we can find that these keywords vary in their semantic meaning: **"**Sydney Cove**"** is a geographical region; **"**Sunset**"** is related to a specific time period (evening) and locations such as beach; **"**Whisky**"** is the attribute of POI.

II RELATED WORK

Mining people's trips from large scale geo-tagged photos

Photo sharing is one of the most popular Web services. Photo sharing sites provide functions to add tags and geo-tags to photos to make photo organization easy. Considering [16]-[20], that people take photos to record something that attracts them, geo-tagged photos are a rich data source that reflects people's memorable events associated with locations. In this paper, we focus on geo-tagged photos and propose a method to detect people's frequent trip patterns, i.e., typical sequences of visited cities and durations of stay as well as descriptive tags that characterize the trip patterns. Our method first segments photo collections into trips and categorizes them based on their trip themes, such as visiting landmarks or communing with nature. Our method mines frequent trip [21]-[25], patterns for each trip theme category. We crawled 5.7 million geo-tagged photos and performed photo trip pattern mining. The experimental result shows that our method outperforms other baseline methods and can correctly segment photo collections into photo trips with an accuracy of 78%. For trip categorization, our method can categorize about 80% of trips using tags and titles of photos and visited [26], cities as features. Finally, we illustrate interesting examples of trip patterns detected from our dataset and show an application with which users can search frequent [27], trip patterns by querying a destination, visit duration, and trip theme on the trip.

Keyword-aware optimal route search

Identifying a preferable route is an important problem that finds applications [28], in map services. When a user plans a trip within a city, the user may want to find "a most popular route such that it passes by *shopping mall*, *restaurant*, and *pub*, and the travel time to and from his hotel is within 4 hours." However, none of the algorithms in the existing work on route planning can be used to answer such queries. Motivated by this, we define [29], the problem of keyword-aware optimal route query, denoted by KOR, which is to find an optimal route such that it covers a set of user-specified keywords, a specified budget constraint is satisfied, and an objective score of the route is optimal. The problem of answering KOR queries is NP-hard. We devise an approximation algorithm OSScaling with provable approximation bounds. Based on this algorithm, another more efficient approximation [30], algorithm BucketBound is proposed. We also design a greedy approximation algorithm. Results of empirical studies show that all the proposed algorithms are capable of answering KOR queries efficiently, while the BucketBound and Greedy algorithms run faster. The empirical studies also offer insight into the accuracy of the proposed algorithms.

Mining significant semantic locations from GPS data

With the increasing deployment and use of GPS-enabled devices, massive amounts of GPS data are becoming available. We propose [31], a general framework for the mining of semantically meaningful, significant locations, e.g., shopping malls and restaurants, from such data.We present techniques capable of extracting semantic locations from GPS data. We capture the relationships between locations and between locations and users with a graph. Significance is then assigned to locations using random walks over the graph that propagates significance among the locations [32]. In doing so, mutual reinforcement between location significance and user authority is exploited for determining significance, as are aspects such as the number of visits to a location, the durations of the visits, and the distances users travel to reach locations. Studies using up to 100 million GPS records [33], from a confined spatio-temporal region demonstrate that the proposal is effective and is capable of outperforming baseline methods and an extension of an existing proposal.

III PROPOSED SYSTEM

The proposed framework KSTR is presented.

KSTR is comprised of two components: the offline pattern discovery and scoring component and the online travel routes exploration component. Offline Pattern Discovery and Scoring Component. Given an LBSN dataset, we first analyze the tags of each POI to determine the semantic meaning of the keywords, which are classified into

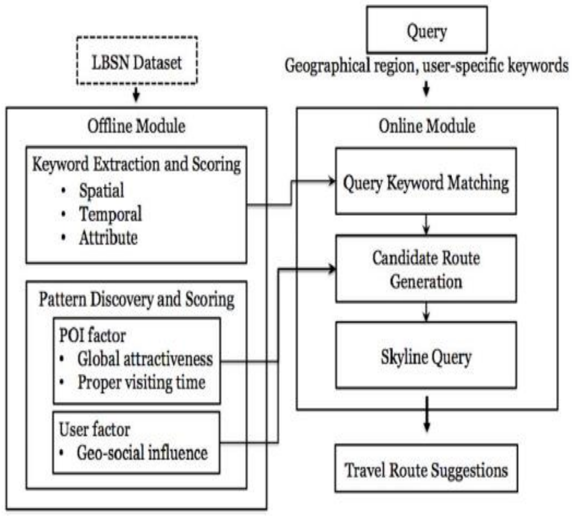
(i) Geo-specific keywords,

(ii) (ii) Temporal keywords,

and (iii) Attribute keywords according to their characteristics. Furthermore, we derive the feature scores of the POIs and generate proper candidate travel routes. Online Travel Routes Exploration Componentz. In this component, we aim to provide an interface for users to specify query ranges and preference-related keywords. Once the system receives a specified range and time, the online component will retrieve those travel routes that overlap the query range and the stay time period. Then, it will compute a matched score of how well the travel route is connected to the keywords.

IV METHODOLOGY

The system architecture is designed with the following components.



Travel Routes Exploration:

In this component, we aim to provide an interface for users to specify query ranges and preference-related keywords. Once the system receives a specified range and time, the online component will retrieve those travel routes that overlap the query range and the stay time period. Then, it will compute a matched score of how well the travel route is connected to the keywords. Consequently, the online component returns the k most representative routes considering the aforementioned feature scores to the users.We first explain the matching function to process the user query. Next, we introduce the background of why we apply a skyline query, which is suitable for the travel route recommendation applications, and present the algorithm of the distance-based representative skyline search for the online recommendation system. Furthermore, an approximate algorithm is required to speed up the realtime skyline query.

With the featured trajectory dataset, our final goal is to recommend a set of travel routes that connect to all or partial user-specific keywords. We first explain the matching function to process the user query. Next, we introduce the background of why we apply a skyline query, which is suitable for the travel route recommendation applications, and present the algorithm of the distance-based representa-tive skyline search for the online recommendation system. Furthermore, an approximate algorithm is required to speed up the real-time skyline query. The *Travel Route Exploration* procedure is presented as Algorithm.

Algorithm: Travel routes exploration

Input: User *u*, query range *Q*, a set of keywords *K*;

Output: Keyword-aware travel routes with diversity in goodness domains *KRT*.

1 Initialize priority queue *CR, KRT*;

2 Scan the database once to find all candidate routes covered by region *Q*;

/* Fetch POI scores and check keyword matching */

3 **For each** *route r found* **do**

4 *r.kmatch* ← 0;

for each *POI p* ∈ *r* **do**

5. *r.kmatch* ← *r.kmatch* + KM(p,k);

6 if *r.kmatch* ≤ *g* **then**

Push *r* into *CR*;

/* Initialize an arbitrary skyline route*/

7. *CR*.r_0 ← route *r* with the largest value of an arbitrary dimension;

/* Greedy algorithm for representative skyline, see Algorithm 3 */

8. *KRT* ← I-greedy(*CR*);

9. **return** *KRT*.

Keyword Extraction:

In this component, keyword extraction component 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 we present how we extract the semantic meaning of the keywords and propose a matched score to describe the degree of connection between keywords and trajectories. The keyword extraction component first computes the spatial, temporal and attributes scores for every keyword w in the corpus. At query time, each query keyword will be matched to the pre-computed score of matching w.CCE: A component, Collective Check-in Extraction, of our proposed method, As candidates for the check-in extraction method m, we present the following two baseline extraction method.the performance of check-in extraction from Flickr photos. Beyond simple matching with an official POI name, harvesting more check-ins requires a trade-off between precision and recall. The performance of check-in extraction depends on whether this trade-off is well controlled. our three proposed extraction methods.

Feature Scoring Methods:

With a set of travel route records, feature scoring should be considered to find proper recommendations. In this paper, we also explore three travel factors: “Where: people tend to visit popular POIs”, “When: each POI has its proper visiting time”, and “Who: people might follow social-connected friends’ footsteps”. To achieve the “Where, When, Who” consideration issue of user demands, the pattern discovery and scoring component defines the ranking mechanism for each POI with global attractiveness, proper visiting time and geo-social influence . From the viewpoint of the POI, we store the attractiveness score and the visiting time information in the POI score vector. On the other hand, from the viewpoint of the user, we also consider a score to quantify an individual’s influence in recommendation.

Route Recommendation:

Route recommendation has to take several factors into consideration to emphasize the unique travel factors of travel routes, the user POI, cost, seasonal preference, time preference of visiting locations such details are combined and the package is mined results is given to the Users and in addition, we refine the results and rank according to **Personalized Recommendation system**

❖ Time-Sensitive Routes (TSR). Only consider the visiting time score of routes. The arrival time of the POIs in the recommendation best fits the extracted proper visiting time. Keyword-Aware Representative Travel Route. Our KRTR outputs optimal representative Skyline routes.

❖ Location Recommendation and Prediction:The task of location recommendation is to recommend new locations that the user has never visited before while the task of location prediction is to predict the next locations that the user is likely to visit Also, most of the research has considered “Where, When, Who” issues to model user mobility. For the location recommendation part, pointed out that people tend to visit near-by locations but may be interested in more distant locations that they are in favor of. Finally, it combined user preference, geographical influence, and historical trajectories to recommend check-in locations. recommended a list of POIs for a user to visit at a given time by exploiting both geographical and temporal influences.

Similarity Route Search: Another relevant area is the similarity route search under specific attributes. Research on this subject has focused on finding routes according to location, activity or keyword-related queries. defined a similarity function for measuring how well a trajectory connects the query locations, considering both spatial distance and order constraint. studied the problem of similarity search on an activity trajectory database.

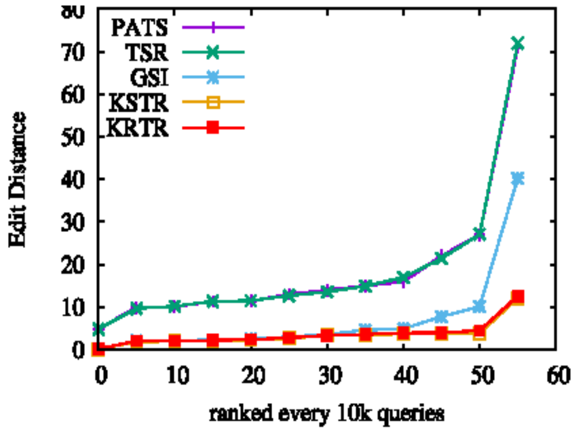
Efficiency:

The online response time of *KRTR* in the three main sub-procedures:

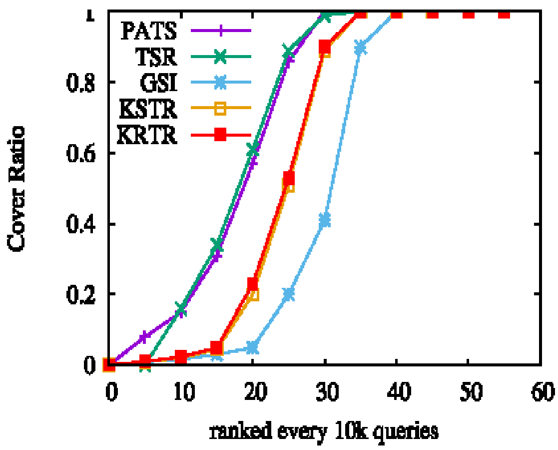
(i) scan the dataset to find the overlap routes and compute the score of candidate routes (O scoring+R scoring),

(ii) Initial skyline point search (I skyline), and

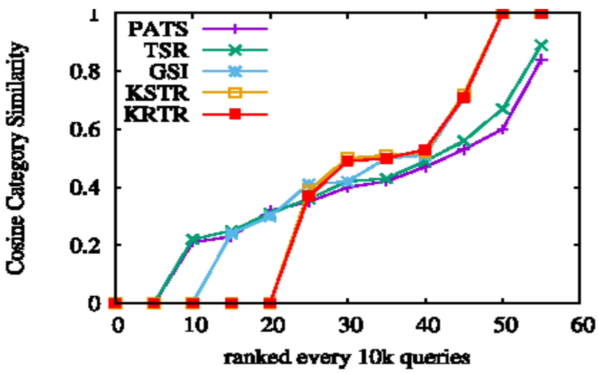
(iii) Representative skyline search (R skyline). We synthesize 34,928 queries from testing users of the *FB* dataset and 39,729 queries from the *CA* dataset. The average response is 1.561708549 seconds. We can find that skyline query (I skyline & R skyline) is the most time-consuming step. we observe the optimal *Nfrac* for approximate candidate route generation. The total running time under different scales is shown.



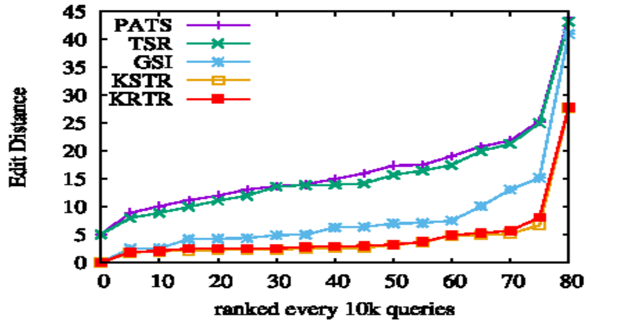
(a) Average edit distance versus to the recommended travel routes of the *FB* dataset



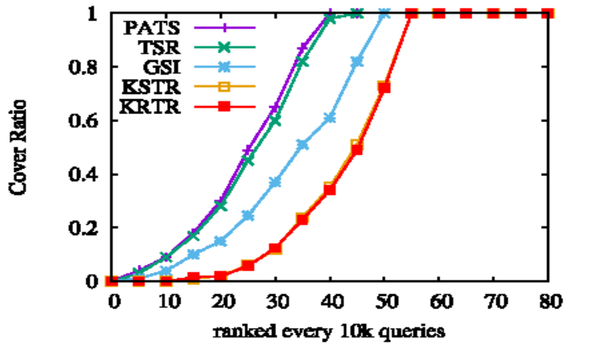
(b) Average region cover ratio versus to the recommended travel routes of the *FB* dataset



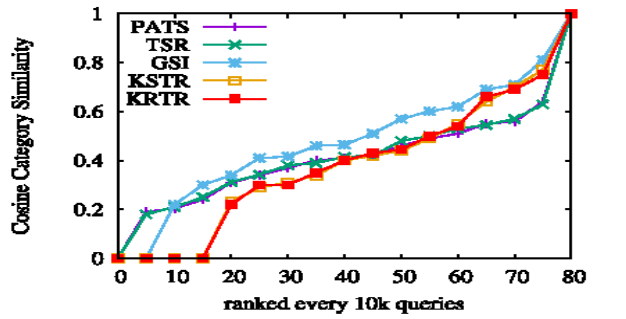
(c) Average category similarity versus to the recommended travel routes of the *FB* dataset



(d) Average edit distance versus to the recommended travel routes of the *CA* dataset



(e) Average region cover ratio versus to the recommended travel routes of the *CA* dataset



(f) Average category similarity versus to the recommended travel routes of the *CA* dataset

V RESULT

Our paper shows the web interval of KRTR in the three main sub-procedures: (i) scan the dataset to find the overlap routes and cypher the score of candidate routes (O scoring+R scoring), (ii) Initial skyline purpose search (I skyline), and (iii) Representative skyline search (R skyline). we tend to synthesize thirty four,928 queries from testing users of the FB dataset and thirty-nine,729 queries from the CA dataset. The average response is 1.561708549 seconds. We are able to realize that skyline question (I skyline & R skyline) is that the most time consuming step. In segment five, we observe the best

Nfrac for approximate candidate route generation. The total running time underneath totally different scales can be identified.

VI CONCLUSION

In this paper, we think about the movement course recommendation problem. We have built up a KRTR structure to suggest travel courses with a particular range and a lot of user preference watchwords. These movement courses are identified with allor fractional client inclination watchwords, and are recommended based on (I) the engaging quality of the POIs it passes, (ii)visiting the POIs at their comparing appropriate landing times, and (iii) the courses produced by persuasive clients. We propose a novel catchphrase extraction component to distinguish the semantic meaning and match the estimation of courses, and have planned a course remaking calculation to aggregate route fragments into movement courses as per query range and timeframe. We influence score capacities for the three previously mentioned highlights and adjust the representative Skyline look rather than the customary best k recommendation system. The examination results show that KRTR is ready to recover travel courses that are intriguing for users, and outflanks the gauge calculations as far as effectiveness and proficiency. Because of the continuous requirements for online frameworks, we intend to diminish the calculation cost by recording rehashed questions and to gain proficiency with the approximate parameters naturally later on.

VII REFERENCES

[1] Y. Arase, X. Xie, T. Hara, and S. Nishio. Mining people's tripsfrom large scale geo-tagged photos. In Proceedings of the 18th ACMinternational conference on Multimedia, pages 133–142. ACM, 2010.

[2] X. Cao, L. Chen, G. Cong, and X. Xiao. Keyword-aware optimalroute search. Proceedings of the VLDB Endowment, 5(11):1136–1147,2012.

[3] X. Cao, G. Cong, and C. S. Jensen. Mining significant semanticlocations from GPS data. Proceedings of the VLDB Endowment, 3(1-2):1009–1020, 2010.

[4] D. Chen, C. S. Ong, and L. Xie. Learning points and routes to recommendtrajectories. In Proceedings of the 25th ACM International onConference on Information and Knowledge Management, pages 2227–2232, 2016.

[5] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng, and X. Xie. Searchingtrajectories by locations: an efficiency study. In Proceedings of the2010 ACM SIGMOD International Conference on Management of data,pages 255–266, 2010.

[6] T. Cheng, H. W. Lauw, and S. Paparizos. Entity synonyms forstructured web search. IEEE transactions on knowledge and dataengineering, 24(10):1862–1875, 2012.

[7] M.-F. Chiang, Y.-H. Lin,W.-C. Peng, and P. S. Yu. Inferring distanttimelocation in low-sampling-rate trajectories. In Proceedings of the19th ACM SIGKDD international conference on Knowledge discoveryand data mining, pages 1454–1457. ACM, 2013.

[8] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties onlocation-based social networks. In ICWSM, 2012.

[9] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani.An energy-efficient mobile recommender system. In Proceedingsof the 16th ACM SIGKDD international conference on Knowledgediscovery and data mining, pages 899–908, 2010.

[10] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi. Trajectory patternmining. In Proceedings of the 13th ACM SIGKDD internationalconference on Knowledge discovery and data mining, pages 330–339,2007.

[11] H.-P. Hsieh and C.-T. Li. Mining and planning time-aware routesfrom check-in data. In Proceedings of the 23rd ACM InternationalConference on Conference on Information and Knowledge Management,pages 481–490, 2014.

[12] H.-P. Hsieh, C.-T. Li, and S.-D. Lin. Exploiting large-scale checkindata to recommend time-sensitive routes. In Proceedings of theACM SIGKDD International Workshop on Urban Computing, pages55–62, 2012.

[13] W. T. Hsu, Y. T. Wen, L. Y. Wei, and W. C. Peng. Skylinetravel routes: Exploring skyline for trip planning. In Mobile DataManagement (MDM), 2014 IEEE 15th International Conference on,volume 2, pages 31–36, 2014.

[14] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura. Travel route recommendationusing geotags in photo sharing sites. In Proceedingsof the 19th ACM international conference on Information and knowledgemanagement, pages 579–588, 2010.

[15] T. Lee, Z. Wang, H. Wang, and S.-w. Hwang. Attribute extractionand scoring: A probabilistic approach. In Data Engineering (ICDE),2013 IEEE 29th International Conference on, pages 194–205, 2013.

[16] X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang. Selecting stars: The kmost representative skyline operator. In Data Engineering. IEEE23rd International Conference on, pages 86–95. IEEE, 2007.

[17] X. Lu, C.Wang, J.-M. Yang, Y. Pang, and L. Zhang. Photo2trip: generatingtravel routes from geo-tagged

photos for trip planning. InProceedings of the 18th ACM international conference on Multimedia,pages 143–152, 2010.

[18] D. Papadias, Y. Tao, G. Fu, and B. Seeger. An optimal andprogressive algorithm for skyline queries. In Proceedings of the2003 ACM SIGMOD international conference on Management of data,pages 467–478, 2003.

[19] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizingpersonalized markov chains for next-basket recommendation. InProceedings of the 19th international conference on World wide web,pages 811–820. ACM, 2010.

[20] A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends andfollowing them to where you are. In Proceedings of the fifth ACMinternational conference on Web search and data mining, pages 723–732, 2012.

[21] Y. Tao, L. Ding, X. Lin, and J. Pei. Distance-based representativeskyline. In Data Engineering. IEEE 25th International Conference on,pages 892–903, 2009.

[22] V. S. Tseng, E. H.-C. Lu, and C.-H. Huang. Mining temporalmobile sequential patterns in location-based service environments.In Parallel and Distributed Systems, 2007 International Conference on,volume 2, pages 1–8. IEEE, 2007.

[23] H. Wang, Z. Li, and W.-C. Lee. PGT: Measuring mobility relationshipusing personal, global and temporal factors. In Data Mining(ICDM), 2014 IEEE International Conference on, pages 570–579, 2014.

[24] W. Wang, H. Yin, L. Chen, Y. Sun, S. Sadiq, and X. Zhou. Geo-SAGE: A geographical sparse additive generative model for spatialitem recommendation. In Proceedings of the 21th ACM SIGKDDInternational Conference on Knowledge Discovery and Data Mining,pages 1255–1264, 2015.

[25] X.-J. Wang, Z. Xu, L. Zhang, C. Liu, and Y. Rui. Towards indexingrepresentative images on the web. In Proceedings of the 20th ACMinternational conference on Multimedia, pages 1229–1238. ACM, 2012.

[26] L.-Y. Wei, W.-C. Peng, B.-C. Chen, and T.-W. Lin. PATS: Aframework of pattern-aware trajectory search. In Mobile Data Management(MDM), 2010 Eleventh International Conference on, pages372–377. IEEE, 2010.

[27] L.-Y. Wei, Y. Zheng, and W.-C. Peng. Constructing popular routesfrom uncertain trajectories. In Proceedings of the 18th ACM SIGKDDinternational conference on Knowledge discovery and data mining,pages 195–203, 2012.

[28] Y.-T. Wen, K.-J. Cho, W.-C. Peng, J. Yeo, and S.-w. Hwang. KSTR:Keyword-aware skyline travel route recommendation. In DataMining (ICDM), 2015 IEEE International Conference on, pages 449–458. IEEE, 2015.

[29] Y.-T. Wen, P.-R. Lei, W.-C. Peng, and X.-F. Zhou. Exploring socialinfluence on location-based social networks. In Data Mining(ICDM), 2014 IEEE International Conference on, pages 1043–1048.IEEE, 2014.

[30] J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activityprediction in location-based social networks. In Proceedings of the2013 SIAM International Conference on Data Mining, pages 171–179.SIAM, 2013.

[31] M. Ye, X. Liu, and W.-C. Lee. Exploring social influence forrecommendation: a generative model approach. In Proceedingsof the 35th international ACM SIGIR conference on Research anddevelopment in information retrieval, pages 671–680. ACM, 2012.

[32] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographicalinfluence for collaborative point-of-interest recommendation.In Proceedings of the 34th international ACM SIGIR conference onResearch and development in Information Retrieval, pages 325–334.ACM, 2011.

[33] H. Yin, B. Cui, Y. Sun, Z. Hu, and L. Chen. LCARS: A spatial itemrecommender system. ACM Transactions on Information Systems(TOIS), 32(3):11,2014.

Authors Profile:

Mr. **Marri Rathna Kumar** pursuing MCA 3rd year in Qis College and Engineering and Technology in Department of Master of Computer Applications, Ongole.

Mrs. **V. Tejaswini** is currently working as an Assistant Professor in Department of Master of Computer Applications in QIS College of Engineering & Technology.