Webology (ISSN: 1735-188X)
Volume 18, Number 5, 2021

Pattern Discovery Of Gps Trajectory With Data Mining Techniques

Dr. S. CHITRA

Assistant Professor, Department of Computer Science, Srimad Andavan Arts and Science College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tamilnadu, India.

ABSTRACT

When the query time is far distant from the current time, existing prediction methods in moving objects databases are unable to reliably estimate positions. Most strategies presume that the trajectory of an object's movements may be represented by some mathematical formulas of motion functions based on its recent movements, even for near future prediction. The movements of an object, on the other hand, are more complicated than mathematical formulas can depict. Prediction based on an object's trajectory patterns is a powerful method that has been studied in a number of studies. Their biggest concern, though, is figuring out how to find the patterns. This research work describes the Hybrid Prediction Model, an unique prediction approach that guesses an object's future positions based on its pattern information as well as existing motion functions using the object's recent motions in this study.

KEYWORDS: GPS Trajectory, Pattern Discovery, Feature selection, Chi-squared analysis, Apriori Association rule mining

1. INTRODUCTION

The subject of trajectory mining has been investigated in the context of smart cities and the Internet of Things (IoT) [1] [2] [3]. Smart cities and the Internet of Things are the way of the future, since trillions of IoT devices, ranging from coffee machines to mobile objects, generate massive amounts of data that must be successfully modelled and processed to improve daily living [4]. To optimise commuting time to work, for example, many sources of information such as the intended route, calendar, city traffic, weather, and so on must come together to determine the most convenient route. As a result, smart data collection, preparation, and fast algorithms that can work with the incoming data and propose solutions in real time are required.

The infusion of intelligence into cities through mobile intelligence [5] is one of the fundamental concerns in future smart cities. The mobility data collected through the Internet of Things is the major source of mobile intelligence. The information is gathered from a range of sources, such as moving humans or equipment that provide location data and a time stamp to a central repository on a regular basis. Once this data is evaluated, it can provide fascinating information [6], such as which regions of the city are seeing an increase in activity [7], [8], the position of any traffic anomalies [9], which individual or group of people is travelling [10], the most popular stay places [11], and so on.

A trajectory is a time-ordered record of a moving object captured at distinct time intervals. During these times, however, the 'precise' location of a moving object may be unknown. Many studies have been conducted on trajectory uncertainty in order to improve the utility of trajectories. Modeling uncertainty using probable world semantics is achievable with probabilistic databases [12]. The uncertainties in the trajectories could be at the event level, such as the uncertainty associated with the object's location, or at the trajectory level, such as the uncertainty associated with the path recorded versus the path taken, or elsewhere [13]. In this case, recording individual mobile object readings and then creating complicated events using probabilistic event extraction [14] is an intriguing method.

2. PRE-PROCESSING USING FILTERED FEATURE SELECTION METHOD

In recent years, the amount of high-dimensional data that survives and is publicly available on the internet has increased dramatically. As a result, machine learning approaches struggle to deal with the large amount of input features, making modelling an appealing topic for academics. The data must be pre-processed in order to use machine learning algorithms effectively. Feature selection [15] is one of the most common and well-known data preparation procedures, and it has become an essential part of the machine learning process. In machine learning and statistics, this is also known as variable selection, attribute selection, or variable subset selection. It is a technique for deleting irrelevant data and recognising relevant features, as well as noisy or redundant data. This method improves comprehensibility and forecast accuracy while speeding up data mining techniques. Irrelevant features provide no more information than the currently selected features, and unrelated features supply no more information than the currently selected features [16][17][18][19][20][21]. Feature selection presents a set of potential features using one of the three ways in supervised inductive learning [22][23][24][25].

- The exact number of features in a subset that optimises an evaluation metric.
- The smaller the subset that fulfils a specific evaluation measure restriction.
- In general, the subset with the best size and evaluation measure commitment.

Noise or redundant characteristics in the data might obstruct feature selection in many cases because they aren't important or related to the class notion, such as microarray data analysis. Machine learning becomes particularly challenging when the number of samples is significantly fewer than the number of features, because the search space will be sparsely populated. As a result, the model will be unable to distinguish between noise and useful data. When it comes to feature selection, there are two main ways. Individual and subset evaluations are the first and second, respectively. Individual evaluation is the process of ranking the features. The weight of each individual aspect is assigned in Individual Evaluation based on its degree of importance. Candidate feature subsets are produced using a search approach in Subset Evaluation.

3. PROPOSED FRAMEWORK FOR THE TRAJECTORY PATTERNS

For finding the frequent itemset of patterns, in the stage of pre-processing, Chi-Square based feature selection technique has used and in the stage of pattern discovery, Apriori Association rule mining algorithm has utilized.

3.1 Chi-Square Feature Selection Method

By evaluating the GPS trajectory dataset, a Chi-Square Attribute Evaluation methodology is applied to choose the attributes for optimising the business strategy.

Another widely used method is feature selection via the chi-square test. The value of a feature is calculated using the chi-squared statistic for the class in chi-squared attribute evaluation. The initial hypothesis H_o is that the two characteristics are unrelated, which is tested using the chi-squared formula:

$$x^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{o_{ij} - E_{ij}}{E_{ij}}\right)^2$$

The null hypothesis asserts that is the predicted (theoretical) E_{ij} frequency, and o_{ij} is the observed frequency. The higher the value of, the more evidence there is against the hypothesis H_o.

3.2 Association Rule Mining Via Apriori Algorithm

The Apriori method is used to generate association rules using the Association Rule Mining technique. To mine frequent item collections, many sorts of algorithms are utilised. Finding frequent item sets from a transaction dataset and generating association rules is one of the most used data mining techniques. Because of the combinatorial explosion, identifying frequent item sets (item sets with a frequency greater than or equal to a user-specified minimum support) is not easy. It is simple to build association rules with confidence greater than or equal to a user-specified minimum confidence once frequent item-sets have been established.

Apriori is a key technique for employing candidate generation to find common item-sets. It is described as a level-wise comprehensive search method that uses anti-monotonicity of item-sets, which states that "if an item-set is not frequent, none of its supersets are ever frequent." Apriori expects that elements in a transaction or item set are sorted in lexicographic order by convention. Let Fk be the set of frequent item-sets of size k, and Ck be the set of candidates. Apriori first scans the database for frequent item-sets of size 1 by counting each item's count and collecting those that meet the minimal support criterion. It then repeats the previous three processes, extracting all of the frequently occurring item sets. The precursor part of the rule is generated using the apriori approach, and the result is generated using a classification strategy in which the complete Telecommunication dataset is divided into two classes, ordinary and attack, based on the marks provided in the dataset. The following Apriori algorithm is used to find the regular itemsets from the dataset:

Algorithm: Apriori Algorithm for finding continual itemset.
Input: Data was normalised, and the minimum support (minsupp) was set to 0.2.
Step 1: Set k (number of itemsets) to one.
Step 2: From the Ck of all application itemsets, find the normal itemset Lk.

Step 3: Scanning D and including each itemset Ck,
Step 4: If count above minimal support, it is continuous.
Step 5: From Lk, make Ck+1; k = k + 1.
Step 6: Combine the Lk-1 and itself item sets to obtain the new contestant item sets.
Step 7: If a non-continuous subset is discovered, expel it.
Step 8: Add the persistent itemset to the rule pool.
Step 9: Repeat steps 2–9 until Ck is depleted.
Output: Frequent item sets.

4. RESULT AND DISCUSSION

4.1 Description of the dataset

Table 1 gives the description of the dataset used in this research work.

Table 1: Description of the GPS trajectory dataset

Sl.No	Feature Name
1	id
2	id_android
3	speed
4	time
5	distance
6	rating
7	rating_bus
8	rating_weather
9	car or bus
q	latitude
11	longitude
12	track_id
13	time

From the filtered feature selection method, the following features are selected by removing the feature with rank 0. Table 2 depicts the result obtained by chi-square feature selection method in the stage of pre-processing. Only nine features are selected among 13 features.

Table 3: Result obtained by using Chi-Square analysis feature selection method in pre-processing step

Sl.no	Original Dataset	Chi-squared feature selection
1	id	id_android
2	id_android	speed
3	speed	time
4	time	distance
5	distance	rating
6	rating	rating_bus

7	rating_bus	rating_weather
8	rating_weather	car or bus
9	car or bus	track_id
10	latitude	
11	longitude	
12	track_id	
13	time	

Following table 3a to table 3f gives the details of the Apriori Association Rule Mining algorithm for discovering trajectory patterns. Table 4 gives the trajectory pattern generated for above reduced dataset by using Chi-Square Feature Selection method.

Table 3a: Result Details of Apriori Association Rule Mining

Minimum support: 0.2 (33 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16
Generated sets of large item sets:
Size of set of large item sets L(1): 11
Size of set of large item sets L(2): 24
Size of set of large item sets L(3): 18
Size of set of large item sets L(4): 7
Size of set of large item sets L(5): 1

Table 3b: Size of set of large item sets L(1): 11

id_android=1 56
rating=2 45
rating=3 101
rating_bus=0 116
rating_bus=1 34
rating_weather=0 116
rating_weather=2 37
car_or_bus=1 87
car_or_bus=2 76
track_id=1 90
track_id=2 73

Table 3c: Size of set of large item sets L(2): 22

id_android=1 rating_bus=0 39
id_android=1 rating_weather=0 39
id_android=1 car_or_bus=2 36

id_android=1 track_id=1 38
rating=2 rating_bus=0 45
rating=2 rating_weather=0 45
rating=3 rating_bus=0 65
rating=3 rating_weather=0 65
rating=3 car_or_bus=1 57
rating=3 car_or_bus=2 44
rating=3 track_id=1 51
rating=3 track_id=2 50
rating_bus=0 rating_weather=0 116
rating_bus=0 car_or_bus=1 87
rating_bus=0 track_id=1 86
rating_bus=1 car_or_bus=2 34
rating_weather=0 car_or_bus=1 87
rating_weather=0 track_id=1 86
rating_weather=2 car_or_bus=2 37
rating_weather=2 track_id=2 36
car_or_bus=1 track_id=1 57
car_or_bus=2 track_id=2 43

Table 3d: Size of set of large item sets L(3): 15

id_android=1 rating_bus=0 rating_weather=0 39
id_android=1 rating_bus=0 track_id=1 34
id_android=1 rating_weather=0 track_id=1 34
rating=2 rating_bus=0 rating_weather=0 45
rating=3 rating_bus=0 rating_weather=0 65
rating=3 rating_bus=0 car_or_bus=1 57
rating=3 rating_bus=0 track_id=1 48
rating=3 rating_weather=0 car_or_bus=1 57
rating=3 rating_weather=0 track_id=1 48
rating=3 car_or_bus=1 track_id=1 40
rating_bus=0 rating_weather=0 car_or_bus=1 87
rating_bus=0 rating_weather=0 track_id=1 86
rating_bus=0 car_or_bus=1 track_id=1 57
rating_weather=0 car_or_bus=1 track_id=1 57
rating_weather=2 car_or_bus=2 track_id=2 36

Table 3e: Size of set of large item sets L(4): 6

id_android=1 rating_bus=0 rating_weather=0 track_id=1 34
rating=3 rating_bus=0 rating_weather=0 car_or_bus=1 57
rating=3 rating_bus=0 rating_weather=0 track_id=1 48

rating=3 rating_bus=0 car_or_bus=1 track_id=1 40
rating=3 rating_weather=0 car_or_bus=1 track_id=1 40
rating_bus=0 rating_weather=0 car_or_bus=1 track_id=1 57

Table 3f: Size of set of large item sets L(5): 1

rating=3 rating_bus=0 rating_weather=0 car_or_bus=1 track_id=1 40

Table 4: Pattern Discovery of GPS trajectory using Apriori Association Rule Mining method

Pattern of GPS Trajectory	Support and Confidence Level of Pattern
rating_weather=0 116 ==> rating_bus=0 116	<conf:(1)> lift:(1.41) lev:(0.21) [33] conv:(33.45)
rating_bus=0 116 ==> rating_weather=0 116	<conf:(1)> lift:(1.41) lev:(0.21) [33] conv:(33.45)
car_or_bus=1 87 ==> rating_bus=0 87	<conf:(1)> lift:(1.41) lev:(0.15) [25] conv:(25.09)
car_or_bus=1 87 ==> rating_weather=0 87	<conf:(1)> lift:(1.41) lev:(0.15) [25] conv:(25.09)
rating_weather=0 car_or_bus=1 87 ==> rating_bus=0 87	<conf:(1)> lift:(1.41) lev:(0.15) [25] conv:(25.09)
rating_bus=0 car_or_bus=1 87 ==> rating_weather=0 87	<conf:(1)> lift:(1.41) lev:(0.15) [25] conv:(25.09)
car_or_bus=1 87 ==> rating_bus=0 rating_weather=0 87	<conf:(1)> lift:(1.41) lev:(0.15) [25] conv:(25.09)
rating_weather=0 track_id=1 86 ==> rating_bus=0 86q	<conf:(1)> lift:(1.41) lev:(0.15) [24] conv:(24.8)
rating_bus=0 track_id=1 86 ==> rating_weather=0 86	<conf:(1)> lift:(1.41) lev:(0.15) [24] conv:(24.8)
rating=3 rating_weather=0 65 ==> rating_bus=0 65	<conf:(1)> lift:(1.41) lev:(0.11) [18] conv:(18.74)
rating=3 rating_bus=0 65 ==> rating_weather=0 65	<conf:(1)> lift:(1.41) lev:(0.11) [18] conv:(18.74)
rating=3 car_or_bus=1 57 ==> rating_bus=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating=3 car_or_bus=1 57 ==> rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
car_or_bus=1 track_id=1 57 ==> rating_bus=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
car_or_bus=1 track_id=1 57 ==> rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)

rating=3 rating_weather=0 car_or_bus=1 57 ==> rating_bus=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating=3 rating_bus=0 car_or_bus=1 57 ==> rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating=3 car_or_bus=1 57 ==> rating_bus=0 rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating_weather=0 car_or_bus=1 track_id=1 57 ==> rating_bus=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating_bus=0 car_or_bus=1 track_id=1 57 ==> rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
car_or_bus=1 track_id=1 57 ==> rating_bus=0 rating_weather=0 57	<conf:(1)> lift:(1.41) lev:(0.1) [16] conv:(16.44)
rating=3 rating_weather=0 track_id=1 48 ==> rating_bus=0 48	<conf:(1)> lift:(1.41) lev:(0.08) [13] conv:(13.84)
rating=3 rating_bus=0 track_id=1 48 ==> rating_weather=0 48	<conf:(1)> lift:(1.41) lev:(0.08) [13] conv:(13.84)
rating=2 45 ==> rating_bus=0 45	<conf:(1)> lift:(1.41) lev:(0.08) [12] conv:(12.98)
rating=2 45 ==> rating_weather=0 45	<conf:(1)> lift:(1.41) lev:(0.08) [12] conv:(12.98)
rating=2 rating_weather=0 45 ==> rating_bus=0 45	<conf:(1)> lift:(1.41) lev:(0.08) [12] conv:(12.98)
rating=2 rating_bus=0 45 ==> rating_weather=0 45	<conf:(1)> lift:(1.41) lev:(0.08) [12] conv:(12.98)
rating=2 45 ==> rating_bus=0 rating_weather=0 45	<conf:(1)> lift:(1.41) lev:(0.08) [12] conv:(12.98)
rating=3 car_or_bus=1 track_id=1 40 ==> rating_bus=0 40	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.53)
rating=3 car_or_bus=1 track_id=1 40 ==> rating_weather=0 40	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.53)
rating=3 rating_weather=0 car_or_bus=1 track_id=1 40 ==> rating_bus=0 40	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.53)
rating=3 rating_bus=0 car_or_bus=1 track_id=1 40 ==> rating_weather=0 40	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.53)
rating=3 car_or_bus=1 track_id=1 40 ==> rating_bus=0 rating_weather=0 40	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.53)
id_android=1 rating_weather=0 39 ==> rating_bus=0 39	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.25)
id_android=1 rating_bus=0 39 ==> rating_weather=0 39	<conf:(1)> lift:(1.41) lev:(0.07) [11] conv:(11.25)

rating_weather=2 37 ==> car_or_bus=2 37	<conf:(1)> lift:(2.14) lev:(0.12) [19] conv:(19.75)
rating_weather=2 track_id=2 36 ==> car_or_bus=2 36	<conf:(1)> lift:(2.14) lev:(0.12) [19] conv:(19.21)
rating_bus=1 34 ==> car_or_bus=2 34	<conf:(1)> lift:(2.14) lev:(0.11) [18] conv:(18.15)
id_android=1 rating_weather=0 track_id=1 34 ==> rating_bus=0 34	<conf:(1)> lift:(1.41) lev:(0.06) [9] conv:(9.8)
id_android=1 rating_bus=0 track_id=1 34 ==> rating_weather=0 34	<conf:(1)> lift:(1.41) lev:(0.06) [9] conv:(9.8)

5. CONCLUSION

This research paper introduced a novel framework in this paper that projected an item's future positions based on both motion function and object movement patterns. Specifically, object trajectory patterns were discovered and specified. Chi-Squared analysis was utilised to minimise the size of the feature space in this suggested framework. The moving pattern of the moving items was discovered using the Apriori Association Rule Mining technique. The prediction of moving objects can be analysed using this pattern.

REFERENCES

[1] Vera-Baquero, Alejandro, and Ricardo Colomo-Palacios. "Big-Data Analysis of Process Performance: A Case Study of Smart Cities." Big Data in Engineering Applications. Springer, Singapore, 2018. 41-63.

[2] Montori, Federico, Luca Bedogni, and Luciano Bononi. "A collaborative internet of things architecture for smart cities and environmental monitoring." IEEE Internet of Things Journal5.2 (2018): 592-605.

[3] Zhang, Ningyu, et al. "Semantic framework of internet of things for smart cities: case studies." Sensors 16.9 (2016): 1501.

[4] Bifulco, Francesco, et al. "ICT and sustainability in smart cities management." International Journal of Public Sector Management 29.2 (2016): 132-147.

[5] Arasteh, H., et al. "Iot-based smart cities: a survey." Environment and Electrical Engineering (EEEIC), 2016 IEEE 16th International Conference on. IEEE, 2016.

[6] Duan, Zongtao, et al. "Personalized service recommendations for travel using trajectory pattern discovery." International Journal of Distributed Sensor Networks 14.3 (2018): 1550147718767845.

[7] Cai, Haoshu, et al. "Mining frequent trajectory patterns of WIP in Internet of Things-based spatial-temporal database." International Journal of Computer Integrated Manufacturing30.12 (2017): 1253-1271.

[8] Radhakrishna, Vangipuram, P. V. Kumar, and V. Janaki. "DESIGN AND ANALYSIS OF SIMILARITY MEASURE FOR DISCOVERING SIMILARITY PROFILED TEMPORAL ASSOCIATION PATTERNS." IADIS International Journal on Computer Science & Information Systems 12.1 (2017).

[9] Cai, Haoshu, Yu Guo, and Kun Lu. "A location prediction method for work-in-process based on frequent trajectory patterns." Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture (2017): 0954405417708222.

[10] Zhang, Rui, et al. "Understanding Trajectory Data Based on Heterogeneous Information Network Using Visual Analytics." International Conference on Mobile Ad-Hoc and Sensor Networks. Springer, Singapore, 2017.

[11] Khoroshevsky, Faina, and Boaz Lerner. "Human Mobility-Pattern Discovery and Next-Place Prediction from GPS Data." IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction. Springer, Cham, 2016.

[12] Gao, Yifeng, et al. "TrajViz: A Tool for Visualizing Patterns and Anomalies in Trajectory." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2017.

[13] Malinovskiy, Yegor, and Yinhai Wang. "Travel pattern discovery using mobile device sensors." U.S. Patent No. 9,374,661. 21 Jun. 2016.

[14] Moosavi, Sobhan, Rajiv Ramnath, and Arnab Nandi. "Discovery of driving patterns by trajectory segmentation." Proceedings of the 3rd ACM SIGSPATIAL PhD Symposium. ACM, 2016.

[15] Semwal, Vijay Bhaskar, et al. "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification." Multimedia tools and applications76.22 (2017): 24457-24475.

[16] Subhashini, M., & Gopinath, R., Mapreduce Methodology for Elliptical Curve Discrete Logarithmic Problems – Securing Telecom Networks, International Journal of Electrical Engineering and Technology, 11(9), 261-273 (2020).

[17] Upendran, V., & Gopinath, R., Feature Selection based on Multicriteria Decision Making for Intrusion Detection System, International Journal of Electrical Engineering and Technology, 11(5), 217-226 (2020).

[18] Upendran, V., & Gopinath, R., Optimization based Classification Technique for Intrusion Detection System, International Journal of Advanced Research in Engineering and Technology, 11(9), 1255-1262 (2020).

[19] Subhashini, M., & Gopinath, R., Employee Attrition Prediction in Industry using Machine Learning Techniques, International Journal of Advanced Research in Engineering and Technology, 11(12), 3329-3341 (2020).

[20] Rethinavalli, S., & Gopinath, R., Classification Approach based Sybil Node Detection in Mobile Ad Hoc Networks, International Journal of Advanced Research in Engineering and Technology, 11(12), 3348-3356 (2020).

[21] Rethinavalli, S., & Gopinath, R., Botnet Attack Detection in Internet of Things using Optimization Techniques, International Journal of Electrical Engineering and Technology, 11(10), 412-420 (2020).

[22] Poornappriya. T.S., & Gopinath, R., Application of Machine Learning Techniques for Improving Learning Disabilities, International Journal of Electrical Engineering and Technology (IJEET), 11(10), 403-411, 2020.

[23] Priyadharshini, D., Gopinath, R., & Poornappriya. T.S., A fuzzy MCDM approach for measuring the business impact of employee selection, International Journal of Management (IJM), 11(7), 1769-1775, 2020.

[24] Poornappriya. T.S., & Gopinath, R., Rice plant Disease Identification using Artificial Intelligence Approaches, International Journal of Electrical Engineering and Technology (IJEET), 11(10), 392-402, 2020.