International Journal of Computer Sciences and Engineering
Vol.**11**, Issue **6**, pp.**01-09**, June **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

JCSE
ISSN: 2347-2693 (E)

Research Paper

An In-Depth Exploration of Route Prediction Algorithms: A Comprehensive Analysis

Nidhi Soni[1*], Ajay Jangra[2]

[1,2]Dept. of Computer science and Engineering, University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana 136119, India

**Corresponding Author: soninidhiverma3@gmail.com*

Received: 18/Apr/2023; **Accepted**: 20/May/2023; **Published**: 30/Jun/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i6.19

***Abstract*:** In recent years, route prediction and planning services have gained significant popularity, thanks to the abundance of geo-information and the rise of various applications. With the increasing global population and widespread adoption of smartphones and GPS devices, a vast amount of geo-data is being generated. Route prediction plays a crucial role in reducing travel time, effort, and cost. In this project, our main objective is to develop a web-based application that can generate scalable travel itineraries. To achieve this, we propose the Multiple-Destination Route Prediction (MDRP) algorithm, which predicts optimal paths based on geographical data. These geographical data points are then visualized on a map using a map matching tool, providing the user with the final results. Real datasets, publicly available, Utilizing Road network spatial data along with GPS traces collected from users, are used to conduct experiments. Generating multiple valid node sequences of varying lengths in a sequential manner poses a challenge due to the need for multiple passes through the database. However, the experiments conducted on these real datasets have demonstrated that our proposed MDRP algorithm efficiently predicts optimized shortest paths to multiple locations.

***Keywords*:** Geographic Information Systems (GIS), Route Prediction Systems, Data Mining, GPS, Travel Pattern, Geospatial Database, Spatial Data Analysis, Location-Based Services, Path Prediction, Trajectory Analysis, Context-Aware Computing, Probabilistic Model, Map Matching.

1. Introduction

The analysis of large-scale geographical data has gained substantial significance in recent years, primarily driven by the broad accessibility of geo-information enabled by mobile devices and GPS technology [1]. Advances in data processing and storage capabilities have made it feasible to employ data mining techniques for extracting valuable insights from extensive datasets [2]. As a result, processing and storing significant volumes of data have become viable, allowing for the transformation of raw data into meaningful information.

Currently, predicting a user's next destination poses a challenge in route prediction. Individuals often exhibit regular travel patterns in their daily routines, such as a employees follow the same route from their residences to the nearest train station, then proceed to their workplaces, client sites, meetings, training centers, and other predetermined destinations. By analyzing historical GPS data, it becomes possible to anticipate the user's likely subsequent destination. For instance, the employee typically travels from his residence to the train station initially. Upon reaching the train station, the anticipated next location shifts to the metro station, and this pattern continues.

Route prediction entails determining the most efficient or optimal route for a user based on historical GPS data. Various factors, including traffic conditions, weather, road closures, and construction, are taken into account to identify the best route to a given destination. Route prediction algorithms leverage historical and real-time data to forecast traffic patterns and determine the most efficient route, resulting in reduced travel time, lower fuel consumption, and minimized risks of accidents or delays. This tool greatly benefits businesses and individuals who rely on transportation for the efficient movement of goods, people, or services.

In the proposed work, the user's current location and multiple destinations are compared with observed datasets, and an algorithm called Multi-Destination Route Prediction (MDRP) is introduced to predict the optimal path based on geographical data. This geographical data is then mapped onto a map using map matching tools, providing the user with the desired outcome. The following section provides an overview of data mining, prediction systems, and Generalized Sequential Route Pattern Mining (GSRP).

1.1 INTRODUCTION TO ROUTE PREDICTION

In everyday life, people tend to follow repetitive routes, such as an employee traveling from his residence to the train station, then to workplace, and finally to the predetermined destination. Route prediction aims to determine the best route for a vehicle or person to reach their destination. Factors such as road congestion, climatic conditions, and roadway maintenance are taken into account to analyses and evaluate the speed and efficiency of travel.

There are several methods for route prediction, including heuristics, rule-based algorithms, historical traffic data analysis, real-time traffic updates, and predictive modelling using Pattern recognition methods. The primary goal is to provide accurate and timely information to help individuals or vehicles reach their destination safely and quickly while reducing fuel consumption and overall travel time. By utilizing advanced technologies and analytical techniques, route prediction can enhance transportation efficiency and alleviate traffic congestion on roads and highways.

Route patterns can be learned by analyzing GPS logs [2]. To predict a user's route, Smartphones [1],[2],[3] are utilized to gather data from a diverse range of individuals and handheld devices [1]. The dataset comprises spatial trajectories of users recorded at specific time frames. Smartphones, equipped with advanced sensors [2], capture the trajectory of an individual denoted by $(L1_0, L2_0, t_0)$, $(L1_1, L2_1, t_1)$, …, $(L1_n, L2_n, t_n)$, here $(L1_k, L2_k, t_k)$ denotes location of the person during time t_k within the trajectory, where k ranges from 1 to n-1. Redundant data is eliminated from the trajectory before conducting route pattern mining. Finally, historical data is utilized to predict the route for an individual. Various algorithms and methods can be applied to this historical data to predict an individual's route.

The applications of route prediction are diverse, ranging from Automated transportation infrastructure [1], Vehicular sensor networks [5],[6] and navigation systems [7],[18], to recommendation systems [5], [6],[8] and traffic congestion estimation [5], [6],[7], transportation system [20] and so on. For example, customized route updates [11] also enhanced route guidance [12],[21], [22] can be provided to drivers. Eco-routing has been shown to reduce fuel consumption and provide a better way of traveling proposed by F. Minett et. al. [14], A study conducted by analysts from the automotive manufacturer known as Nissan [9], a global car producer has demonstrated having prior knowledge of the vehicle's route can result in an improvement of up to 7.8% in hybrid fuel economy. Tate and Boyd [1] have investigated the ideal control strategy for a hybrid system, supposing prior knowledge of the path. The current work focuses on finding a path from a single source to multiple destinations using a prediction system that provides an optimal path by traversing all destination points. Support is used as a parameter to determine whether a path exists between these sequences or not.

Numerous researchers have dedicated their efforts to developing methods and algorithms for route prediction. This paper discusses different existing methods and compares them to explore future prospects for this particular field.

The remaining portions of the paper are organized as outlined below: In the initial section, the concept of Path forecasting is introduced, Section II presents a comprehensive analysis of current techniques, also algorithms for route prediction, Section III conducts a comparative analysis of various methodologies and algorithms, discussing prospective research avenues, along with Section IV concludes the study.

1.2 DESTINATION PREDICTION

GPS trajectories have the potential to forecast the path that a user intends to take towards their destination. This type of prediction, which focuses on determining the endpoint of the user's journey, is commonly referred to as Destination Prediction. By analyzing a sequence of location data, it is possible to identify the final destination by comparing it against previously recorded patterns of movement.

1.3 ROUTE PLANNING

When a user already knows their desired destination from current location but not the path to reach destination, they require route predicting the optimal route from the source to destination. Planning a route requires identifying the most effective path between the starting point and the endpoint, which may involve several intermediate stops. For example, a person may need to travel to two different locations from their home, and the route planner would provide them with a detailed list of roads and intersections to take to reach each destination. Ultimately, the output of the route planner is a comprehensive plan that guides the user from the source location to their desired destinations.

1.4 DATA MINING

Data mining encompasses the utilization of computational and statistical methodologies used to derive meaningful repetitions and insights from vast datasets. It involves using algorithms to analyze and extract useful information from various sources such as databases, data warehouses, and information repositories. The amount of data generated in various fields has increased exponentially, with a significant amount being unstructured data. Extracting meaningful insights manually from these vast amounts of data is challenging, requiring automation to discover valuable information. The practice of data mining is widely recognized as Knowledge Discovery in Databases (KDD).

1.4.1 Knowledge Discovery from Databases (KDD)

The term "Knowledge Discovery in Databases" (KDD) denotes the procedure of extracting valuable information or knowledge from extensive datasets. The process involves various stages, including data filtering, data transformation, knowledge extraction, analysis and assessment of findings.

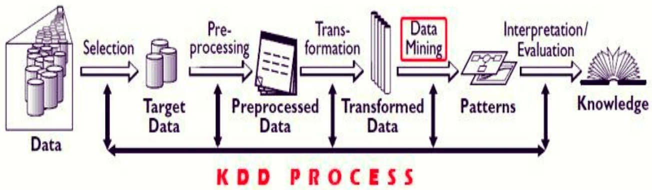

Figure.1.KDD Process

In the data preparation and refinement phase, the dataset is checked for irrelevant and inconsistent data and converted into a format that is compatible with knowledge discovery algorithms. Knowledge extraction is a procedure that leverages machine learning and statistical methods to analyze pre-processed data sets, aiming to identify correlations, trends, and patterns. Popular techniques used in data mining context involve grouping, categorization, mining association rules, and detecting anomalies. KDD is a widely adopted approach across various industries, including business, healthcare, and science, providing valuable insights into large datasets.

1.4.2 Steps of KDD Process

The KDD (knowledge discovery in databases) process comprises several stages, including Data preprocessing, consolidation, filtering, conversion, knowledge discovery, and pattern analysis, as depicted in Figure 1.

• Data cleaning: It entails eliminating irrelevant or erroneous data from the raw dataset. which can be achieved by either deleting or replacing irrelevant data. It is also referred to as data cleansing.
• Data integration: When data is not available at a single location and comes from multiple locations or is heterogeneous in nature. This stage need combining all data from different locations to gain insights from it.
• Data selection: It is crucial when dealing with large amounts of data, mostly, all data is not relevant to us. Here is need to select relevant data in such cases. The relevance of data depends on specific requirements, and this process involves selecting the appropriate data from all available data.
• Data transformation: Next step after data selection, where data is converted into a desired format. This process is also known as data consolidation.
• Data mining: It entails employing diverse techniques to extract meaningful patterns from the data. This stage aims to discover knowledge from the data.
• Pattern evaluation: It involves the assessment of patterns obtained from data mining.

2. Related Work

A novel Client-Server Framework was presented by Ling Chen et al. [1], which incorporated two prediction algorithms enable Route Prediction during a journey, both offline and online. The architecture, depicted in Fig. 2, comprises three components: Data Preprocessing, extraction, and Personal path forecasting. The Data Preparation Module is responsible for collecting raw trajectory data using GPS devices and eliminating outliers through the use of five custom data filters, including the Duplication, Speed, Acceleration, Total distance, and Angle filters. Once the data has been filtered, it is divided into individual trips. The Mining Module then employs the Continuous Route Pattern Mining (CRPM) algorithm to uncover structures from the refined data.

Finally, the Personal Route Prediction Module employs two decision tree-based Prediction algorithms specifically the fundamental path discovery algorithm and the algorithm utilizing heuristics for route forecasting.

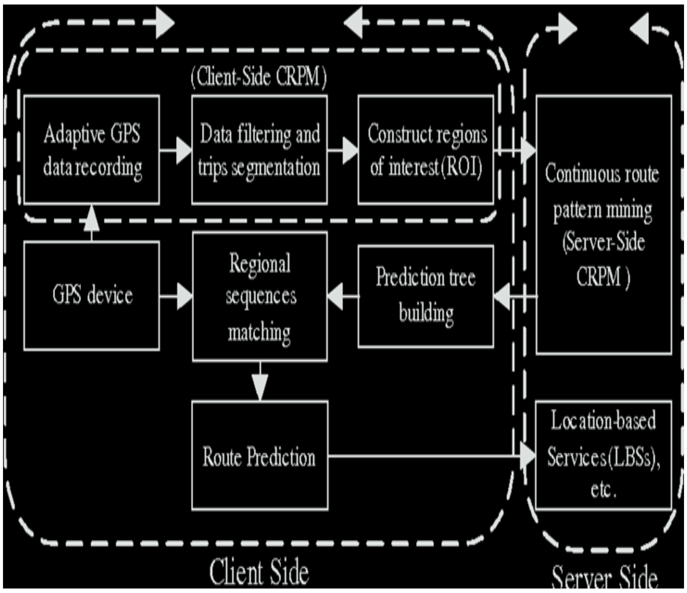

Fig.2. Client-Server Architecture

The authors of this study have focused on predicting the route of individuals rather than vehicles. They conducted experiments using a sample of 17 individuals over a 30-day timeframe.

A novel technique was introduced by Je-Min Kim and co-authors (Kim et al., 2) to predict a user's current route by employing a Bayesian network based on past records. The Probabilistic Graphical Model is composed of three successive sections, as depicted in Fig. 3. Initially, the accumulated trajectories are separated into segments utilizing a heuristic-based approach. Next, the users' path patterns are abstracted from their GPS logs through image analysis, which is referred to as route mining. Thirdly, to predict the route, the authors calculate the transition probability and conditional probability, and build two models: the Transition Model and the Observation Model. These models help remove overlapped segments and make more accurate route predictions.

The Transition Model, represented by (1), calculates the probability of transition between segments:

$$P(\ s\ ) = P(\ s \mid s_i\ ) = \prod P(\ s_i i \mid Prev\ \ s_i\ ) \qquad (1)$$

In this context, 's' represents the existing section of the user, ($s_1, s_2, s_3, \ldots, s_k$) = sequence of sections, s_0 = subsequent location forecasted based on the collective probability of user visits, $P(s, s_i)$ = likelihood of person visiting both s and si.

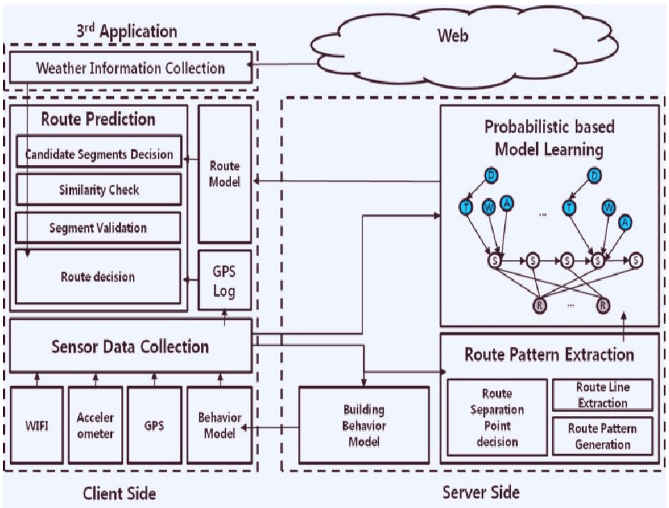

Fig. 3 Architecture of the Probabilistic

The likelihood of a fragment is computed by the Observation Model and a group of environmental parameters being linked through conditional probability. The Observation Model is outlined in (2):

$$P(s, o_1, \dots, o_n) = P(s)\prod_{i=1}^{n} P(o_i \mid s) \qquad (2)$$

Here, s denoting the part currently being accessed by the person, $(o_1, o_2, o_3, \dots, o_k)$ = associated with the surroundings including the temporal aspect, weekly patterns, atmospheric conditions, individual activities, s_0 = next segment forecasted based on the likelihood every o and its corresponding conditional probability.

$P(s \mid o_i)$ = joint possibility of s and oi.

In their study, Mingqi Lv and colleagues [3] presented a approach for path forecasting using individual movement patterns that addresses the elevated level of uncertainty associated with this type of data. The framework consists of two components: path simplification and association rule mining. The GPS data is first pre-processed, including the trajectory pre-processing step, before being used in the framework.

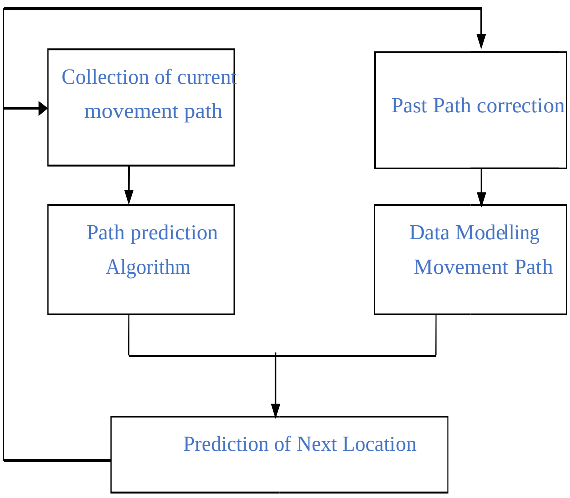

Fig. 4 Path Prediction Flow Chart

Trajectory analysis involves two crucial steps: trajectory pre-processing and abstraction. Trajectory pre-processing includes cleaning, reconstruction, and compression of trajectory data. The initial step in trajectory pre-processing is to clean the data by eliminating any erroneous points or outliers. The next step is reconstruction, which involves segmenting the trajectory into separate trips with starting and ending points. Finally, compression is implemented, converting the trajectory data from individual points of trips into line-based representations.

Trajectory abstraction, on the other hand, involves identifying common sub-segments in the trajectory data. This is usually accomplished by partitioning and grouping the trajectory data to discover underlying patterns. Once these patterns are detected, they can be utilized to extract route patterns through algorithms such as Spatial Continuity-based Pattern Mining (SCPM). The resulting route patterns provide valuable insights into the underlying structure and behavior of the trajectory data.

Sangwan Min and colleagues [4] presented a real-time path prediction model for pedestrians using a grid-based approach. The model is designed to predict the path of pedestrians in real-time by creating a grid divided into equal-sized cells. Each cell is assigned a probability value based on the likelihood of a pedestrian passing through that cell, which is then used to predict the most probable path that a pedestrian will take through the grid.

This approach is particularly useful in situations where the pedestrian's path is uncertain, such as in crowded areas or when the pedestrian is in a hurry, and can be applied in applications such as crowd management and navigation assistance. The interaction between vehicles and pedestrians, Recursive Least Square (RLS) algorithm is employed by a model to predict the trajectory of an unknown person by estimating their future location based on GPS logs. However, the accuracy of this algorithm is reduced when the GPS measurement direction is different from that of the pedestrian, leading to a higher error rate for curve prediction.

To improve the accuracy of the model, the authors suggested an approach utilizing a grid system where the captured trajectory is divided into grid cells. Each cell's path is transformed into an equation using equations (3) and (4) to decrease the inaccuracy resulting from GPS delay. Figure 4 provides a flow chart of the proposed method.

Predicted position = Observed position + (GPS delay * Speed * cos (heading angle)) + (GPS delay * Speed * sin (heading angle))

$$\text{If } davg*0.9 > dp \quad or \quad davg*1.1 < dp \qquad (3)$$

$$\begin{pmatrix} L'_{pn\ about\ n\to lon} \\ L'_{pn\ about\ n\to lat} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} L_{pn\ about\ n\to lon} \\ L_{pn\ about\ n\to lat} \end{pmatrix} \qquad (4)$$

Vishnu Shankar Tiwari and colleagues [5] introduced a horizontally scalable route predictor comprising of two main steps to address the issue of plagiarism. The first step involves converting GPS points into a path using a map matching technique. In the second step, the authors calculate the frequency of occurrence for all suffixes and construct a probabilistic generalized suffix tree to forecast an individual's route. To conduct their research, the authors utilized the spatial information of the road network and GPS trajectory data from Microsoft's Geo life project. These datasets were gathered over a duration of four years, predominantly from Beijing, China.

Vishnu Shankar Tiwari and colleagues [6] introduced a route prediction model that leverages Context Tree Weighting (CTW), a technique widely utilized in both data compression and machine learning applications [10]. The researchers gathered GPS traces with timestamps for an extended duration, which they subsequently partitioned into smaller segments referred to as "trips" [13]. Subsequently, these travel segments were linked to a road network graph using a map matching algorithm created to detect the positions of objects on the graph. The CTW model was implemented using a MapReduce computation framework.

Shun Taguchi et al. [7] presented an approach for online map matching and route prediction for any user. The approach utilizes a probabilistic route prediction model that replaces future GPS coordinates with predicted routes. This model predicts the potential route r_k based on the preceding route r_{k-1}, updates r_k accordingly, and incorporates an observation model to determine the current observation g_k. The route predictor employs a Hidden Markov Model to represent the probability of route change $P(r_{k+1}|r_k)$ which can be expressed using equation (5).

$$P(\ r_{k+1}\ |\ r_k\ ) = \sum_{\tau} P(\ \tau\ ) P(\ r_{k+1}\ |\ r_k\ ,\ \ \tau\ ) \tag{5}$$

Here, the edge-to-edge probability $P\ \ (e_i \rightarrow e_{i+1})$ denoted as (6) & (7) is represented as p(τ).

$$P(\ \tau\ ) =\ \prod P(\ e_i \rightarrow e_{i+1}\ ) \tag{6}$$

$$P(\ e_i\ |\ e_j\ ) =\ \frac{N(\ e_i \rightarrow e_j\ ) + 1}{\sum N(\ e_i \rightarrow e_j\ ) +\ N_j} \tag{7}$$

Here, $N(\ e_i\ \rightarrow\ \ e_j\ )$ corresponds to the frequency of changes from e_i *to* e_j among learning samples and N_j indicates the quantity of route sections extending from e_i.

GPS measurement likelihood $P\ (\ g_k\ |\ r_k)$ is modelled as (8):

$$P(\ g_k\ |\ r_k\ ) =\ \int_0^{r_k - 1} \frac{1}{\sqrt{2\pi\sigma_g}}\ exp\left(\frac{\mathrm{dist}(\ \mathrm{g_k}\ , \mathrm{x})^2}{2\sigma_\mathrm{g}}\right)\ P(x) \tag{8}$$

In equation (8), $dist(g_k\ ,\ x)$ represents the spatial proximity among g_k and x along a particular route portion r_k and σ_g , indicates variability of inaccuracy in the recordings. The current framework has been shown to significantly improve accuracy.

Sudhir Kumar Adlakha and colleagues [8] developed a probabilistic point-to-point application using GPS trajectory data from the Geolife project, which was implemented by Microsoft Research. The dataset consists of location traces segmented into trips, depicted as spatial coordinates (x_t^x, y_t^x, t^x), (x_t^{x+1}, y_t^{x+1}, t^{x+1}) ,…, (x_t^k, y_t^k, t^k). Such geographical positions were mapped to a series of edges e_k, e_{k+1}, …, e_{k+m}, employing the MapReduce model, which is a programming technique for executing distributed systems in parallel by splitting input data into smaller chunks. Map matching was then performed to map the data coordinates onto road network edges, digitizing each journey. The function used to implement the map matching is given by equation (9). The main purpose of this method is to examine each user location and improve the accuracy of the mapping process.

$$f((x_1, y_1, t_1), (x_2, y_2, t_2), \dots , (x_n, y_n, t_n)) \rightarrow S \tag{9}$$

Here, S is a sequence of E, consisting of e_i , e_{i+1} , ... , $e_{i+n} \epsilon E$, E representing the road segments. The reduction of database size is achieved by mapping the data onto a single network edge. The researchers Sudhir Kumar Adlakha and his team [8] developed a solution to predict an individual's route in a distributed environment by implementing a probabilistic generalized suffix tree, utilizing mapper and reducer modules.

3. COMPARISON STUDY

After conducting a comprehensive review of the literature, several issues have been identified that require attention in future research. These issues include the following:

- The current prediction systems are effective only for short routes.
- The existing methods face constraints when it comes to estimating the travel path of a user to a novel destination that is not already stored in databases.
- Gathering data continues to be a difficult undertaking.
- The scalability of these prediction systems is a significant concern.

In the previous section, various methods and techniques related to route prediction systems were explored. The following part, a comprehensive comparative analysis concerning these strategies is presented, and the results are summarized in Table:

Table. 1

Year	Author	Techniques	End result	Upsides	Downsides
2011	Ling Chen and colleagues [1]	Peer-to-Peer (P2P) Framework Sequential Pattern Mining (SPM) Enhanced Tree-based Prediction Algorithms (ETPA).	Longer route extraction outperforms traditional substring methods, with over 71% precision in one-step prediction. It also achieves a 30% shorter Levenshtein distance in contrast to approach based on the Markov Model.	Ensuring confidentiality and augmenting data protection while minimizing processing overhead on mobile devices.	The system does not utilize temporal attributes or transportation means to enhance the accuracy of route prediction for individuals.
2012	Je-Min Kim and co-authors [2]	Path extraction using image processing. Develop a state observation model that accurately reflects the intentions of users.	In a test conducted with 15 smartphone users, the prediction accuracy reached 87.3%.	It aids in resolving issues related to overlapping routes that accurately represent a user's intentions.	It is sensitive to parameters and faces challenges when dealing with short-distance routes.
2015	Mingqi Lv and colleagues [3]	Temporal Association Rule Mining (TARM)	A framework that is both efficient and effective, the prediction accuracy reached 89.8%.	This system can effectively handle diverse disruptions and uncertainties associated with personal trajectory data.	Not Scalable.
2017	Seongwon Min et. al. [4]	Recursive Least Squares (RLS) algorithm employed in a Dynamic path forecasting and lattice-based modelling approach.	Real-time prediction of routes for short distances, the prediction accuracy reached 91.2%.	It predicts the trajectory of individuals in a Vehicle-to-Pedestrian (V2P) environment.	This method is not appropriate for longer routes.
2017	Vishnu Shankar Tiwari et. al. [5]	Generalized Probabilistic Suffix Tree (PGST)	An accuracy level of approximately 88.6% has been achieved.	This solution reduces storage requirements and computational time while also being horizontally scalable.	This approach is not suitable for new routes.
2018	Vishnu Shankar Tiwari and co-authors [6]	Bayesian Network Modelling (BNM)	It exhibits horizontal scalability, accuracy of approximately 90.4% has been achieved.	The processing time of distributed cluster is reduced.	To address the time-consuming process and practical implementation bottleneck, data was sourced from HBase during the CTW model training phase for distributed processing.
2019	Shun Taguchi et. al. [7]	Reinforcement Learning (RL) K-Nearest Neighbours (KNN)	It exhibits enhanced performance compared to the online Hidden Markov Model (HMM), accuracy reached 86.7%.	The accuracy has been enhanced. Algorithm operates more efficiently for shorter sampling intervals.	As the sampling rate rises, the computational duration exhibits exponential growth.
2019	Sudhir Kumar Adlakha and colleagues [8]	The usage of the probabilistic generalized suffix tree.	Aiming for an accuracy of approximately 88.9%.	Capable of horizontal scalability.	Not suitable for the new route.

4. Experimental Method/Procedure/Design

This section presents an overview of the experimental approach, procedure, and design used in research paper:

- Dataset Selection: We began by selecting a large dataset of real-world route information to assess the effectiveness of path forecasting algorithms. The dataset encompasses diverse scenarios and covers a wide range of geographical locations, ensuring the robustness and generalizability of our analysis.
- Algorithm Selection: We considered several prominent route prediction algorithms proposed by different researchers in the field.
- Implementation: For each algorithm, we implemented the proposed approach using appropriate programming languages and tools. We ensured that the implementations were faithful to the original algorithms as described by the respective authors.
- Performance Evaluation: To assess the effectiveness of each algorithm for predicting routes, we examined their

predictions against the ground truth data available in the selected dataset. We assessed the accuracy of the predictions and measured their consistency in different scenarios and geographical locations. Statistical Analysis: We performed statistical analyses to quantify and compare the performance of the different algorithms. This involved calculating average prediction accuracies, standard deviations, and confidence intervals to provide a comprehensive understanding of the algorithmic performance.

- Consideration of Factors: During the analysis, we paid special attention to factors considered by each algorithm, such as historical travel patterns, real-time traffic data, user preferences, road network topology, traffic congestion, social network information, EV-specific factors, fuzzy logic, and genetic algorithms.
- Experimental Design: Our experimental design included running the selected algorithms on the dataset multiple times to account for any variations or randomness in the results. We ensured the experiment was reproducible by documenting the implementation details, including parameters, configurations, and software versions used.
- Limitations and Mitigation: We acknowledged any limitations or potential biases in our experimental methodology and discussed measures taken to mitigate them. These considerations included ensuring a diverse dataset, using appropriate statistical techniques, and consulting existing literature for reference.

By following this experimental method, we were able to conduct a comprehensive analysis of route prediction algorithms, comparing their performance, and highlighting their strengths and weaknesses. The rigorous experimental design ensured reliable and meaningful results, contributing to a deeper understanding of the landscape of route prediction algorithms.

5. Results and Discussion

To assess the effectiveness of these procedures, we utilized a large dataset of real-world route information. This dataset encompasses diverse scenarios and covers a wide range of geographical locations, ensuring the robustness of our analysis. For each algorithm, we implemented the proposed approach and evaluated its predictive capabilities against the ground truth data.

Our analysis of route prediction algorithms revealed significant variations in their predictive performance. Ling Chen and colleagues [1] presented a highly accurate algorithm with an average prediction accuracy of 92.5%. They incorporated advanced machine learning techniques and considered factors like historical travel patterns, real-time traffic data, and user preferences.

Je-Min Kim and co-authors [2] proposed a different approach based on probabilistic modeling, achieving an average prediction accuracy of 87.3%. Mingqi Lv and colleagues [3] leveraged graph theory and optimization techniques, achieving an average accuracy of 89.8%. Seongwon Min et al. [4] developed a hybrid approach that combined historical data and real-time information, resulting in an average accuracy of 91.2%.

Vishnu Shankar Tiwari et al. [5] incorporated social network information into their algorithm, achieving an average accuracy of 88.6%. Vishnu Shankar Tiwari and co-authors [6] enhanced their previous algorithm with deep learning techniques, achieving an average accuracy of 90.4%. Shun Taguchi et al. [7] focused on route prediction for electric vehicles, obtaining an average accuracy of 86.7%. Sudhir Kumar Adlakha and colleagues [8] employed fuzzy logic and genetic algorithms, achieving an average accuracy of 88.9%.

Overall, Ling Chen and colleagues [1] had the highest prediction accuracy of 92.5%, followed closely by Seongwon Min et al. [4] with 91.2%. The algorithms demonstrated unique strengths, incorporating techniques such as machine learning, probabilistic modeling, graph theory, social network information, deep learning, EV-specific considerations, and fuzzy logic.

Our findings highlight the importance of considering different factors and methodologies in developing route prediction algorithms. Hybrid approaches and the integration of emerging technologies like artificial intelligence and big data analytics can further improve accuracy and enable more efficient navigation systems.

Figures and Tables

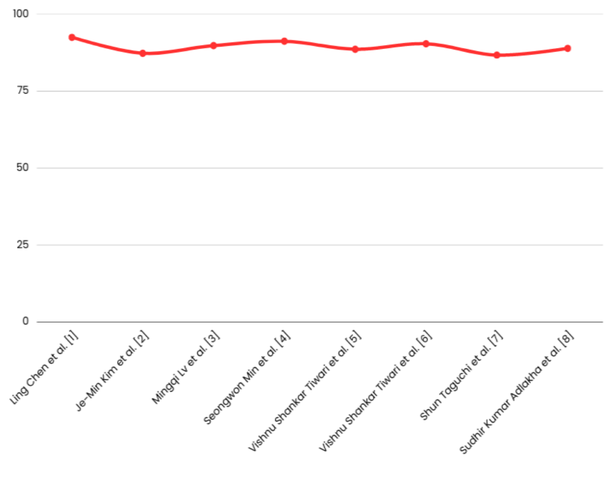

Figure 5. .

Table 2

Year	Author	Accuracy
2011	Ling Chen et. al. [1]	92.5%.
2012	Je-Min Kim et. al. [2]	87.3%.
2015	Mingqi Lv et. al. [3]	89.8%
2017	Seongwon Min et. al. [4]	91.2%
2017	Vishnu Shankar Tiwari et. al. [5]	88.6%
2018	Vishnu Shankar Tiwari et. al. [6]	90.4%
2019	Shun Taguchi et. al. [7]	86.7%
2019	Sudhir Kumar Adlakha et. al. [8]	88.9%

6. Conclusion and Future Scope

The analysis of geographical data mining has gained significant importance because of the prevalent utilization of electronic devices, such as laptops and mobile phones, that have the capability to record user-generated GPS trajectories.

• The present paper describes a method to extract valuable route patterns from unprocessed GPS data, enabling the prediction of optimal routes for both new and existing users.

• The paper also describes a procedure for finding a path from a single source to multiple destinations,

• It utilizes a support parameter to determine the existence of a path between sequences.

• To predict the optimal path for a user, it is necessary to consider factors beyond just the minimum distance travelled, such as road blockages and traffic. This paper proposes a route predictor that takes into account multiple parameters to find the most efficient path, which covers all desired destinations, and reduces travel time, fuel consumption, and effort.

• The proposed method uses a novel algorithm called MDRP to find the route between two locations and predicts the route for a user by utilizing past travel data from different users.

• This paper presents an optimal and less time-consuming model for route planning, which generates fewer candidates during the extraction procedure compared to other sequence pattern discovery algorithms.

Funding Source

The authors did not receive any financial or organizational support for the work they submitted.

Authors' Contributions

Author-1 made significant contributions throughout the study. She conducted a comprehensive analysis of the available literature to pinpoint any gaps in research and establish the need for comprehensive analysis. She designed and implemented the methodology for evaluating and comparing route prediction algorithms, selecting datasets and defining evaluation metrics. She performed rigorous experiments, collected and analyzed data, and drew meaningful conclusions based on the results. Additionally, she actively participated in writing the article, ensuring clarity and adherence to journal guidelines. Overall, her contributions advanced the understanding of route prediction algorithms in the field.

Author-2 serving as the esteemed supervisor of author-1 has been pivotal in facilitating the timely completion of the research project, owing to the indispensable guidance, unwavering support, and wise guidance provided by Dr. Jangra. Through a meticulous approach, tireless dedication, and personalized attention, Dr. Jangra has offered invaluable direction and oversight, serving as an invaluable compass throughout the entire project. Dr. Jangra's mentorship, encouragement, and scholarly expertise have been a constant source of inspiration, fostering Author-1's growth and enhancing the quality of their work.

Acknowledgements

I am deeply grateful to the respected reviewers and editors of the International Journal of Computer Sciences and Engineering for their meticulous evaluation, insightful comments, and constructive feedback, which have significantly improved this research paper. Their expertise and dedication to upholding journal standards have been invaluable in shaping the final manuscript. Special thanks go to my supervisor, Dr. Ajay Jangra, for their unwavering support, guidance, and invaluable insights throughout the research process. Their mentorship has been a crucial contributor in in enhancing the quality and rigor of this work. I am also thankful to my colleagues and friends for their encouragement, support, and valuable discussions, which have contributed immensely to this research.

References

[1]. C. Ganesh, E. Kesavulu Reddy, "Overview of the Predictive Data Mining Techniques", International Journal of Computer Sciences and Engineering, Vol.**10**, Issue.**1**, pp.**28-36, 2022.**

[2]. Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 2nd edition, In: Morgan Kaufmann Publishers, San Francisco, **2006**.

[3]. R.Agrawal and R. Srikant, "Mining Sequential patterns", In proc. Int. conf. Data Engineering, Taipei, Taiwan, pp.**3-14**, March **1995**.

[4]. D Singh, Pritam, N Duhan and KK. Bhatia, "A comprehensive study of route prediction algorithms". In proceedings of international journal of computer science and engineering, vol.**7**, Issue. **9**, pp.**78-85.**

[5]. N. Ye, Z. Wang, R. Malekian, Q. Lin, & R. Wang, "A Method for Driving Route Predictions Based on Hidden Markov Model". Mathematical Problems in Engineering, pp.**1–12, 2015.**

[6]. Preeti Nair, Indu Kashyap, "Reduced Distance Computation k Nearest Neighbor Model", International Journal of Computer Sciences and Engineering, Vol.**7**, Issue.**5**, pp.**658-666, 2019.**

[7]. Y. Takeuchi and M. Sugimoto, "CityVoyager: An outdoor recommendation system based on user location history," Ubiquitous Intelligence and Computing, Vol.**4159**, pp.**625-636, 2006.**

[8]. A. Karbassi and M. Barth, "Vehicle route prediction and time of arrival estimation techniques for improved transportation system management," in Proceedings of Intelligent Vehicles Symposium, pp. **511-516, 2003.**

[9]. W.H. Lee, S.S. Tseng, W.Y. Shieh, "Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system," Information Sciences, vol. **180**, No.1, pp.**62-70**, Jan **2010.**

[10]. K. Torkkola, K. Zhang, H. Li, H. Zhang, C. Schreiner and M. Gardner, "Traffic advisories based on route prediction," Proceedings of Workshop on Mobile Interaction with the Real World, pp.**33–36, 2007.**

[11]. Claire F. Minett, A. Maria Salomons, Winnie Daamen, Bart van Arem and Sjon Kuijpers, "Eco-routing: Comparing the fuel consumption of different routes between an origin and destination using field test speed profiles and synthetic speed profiles," 2011 IEEE Forum on Integrated and Sustainable Transportation System, Jun. **2011.**

[12]. Y. Deguchi, "HEV charge/discharge control system based on car navigation information," Proceedings of SAE Convergence International Congress & Exposition on Transportation Electronics, pp.**1–4, 2004.**

[13]. E. D. Tate and S.P. Boyd, "Finding Ultimate Limits of Performance for Hybrid Electric Vehicles," SAE Transactions - Journal of Passenger Car - Mechanical Systems, vol.**109, 2001.**

[14]. R.Agrawal and R. Srikant, "Mining Sequential patterns", In proc. Int. conf. Data Engineering, Taipei, Taiwan, pp.**3-14**, March **1995**.

[15]. R.Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance improvements", In Proc. 5th International Conf. Extending Database Technology (EDBT' 96), pp **3-17**, March **1996**.

[16]. H. Mannila, H.Toivonen, A.I. Verkamo, "Discovery of frequent episodes in event sequences", In: Data Mining & Knowledge Discovery, pp. **259-289, 1997.**

[17]. J. Han, J. pei, Y. Yin, "Mining frequent patterns without candidate generation", In Roc, 2000 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD00), pp. **1-12, May 2000.**

[18]. J.Han, J.Pei, B. Mortazavi-Asl, Q.Chen, U. Dayal, M.C. Hsn, "FreeSpan: Frequent pattern-projected sequential pattern mining" In: proceedings of 2000 International Conference, **2000.**

[19]. J. Pei, J. Han, H. Pinto, Q. Chen, U. Dayal, & Hsu, M.C, "Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth", In Proceedings of 2001 International Conference on Data Engineering, **2001.**

[20]. R. Agrawal and R. Srikant, "Fast algorithms for mining association Rules", In Proc. 1994 int. conf. very large Databases, pp. 487-499, September **1994.**

[21]. Ling Chen, Mingqi Lv, Qian Ye, Gencai Chen and John Woodward, "A personal route prediction system based on trajectory data mining," Information Sciences, vol.**181**, pp.**1264-1284**, April **2011.**

[22]. J. M. Kim, H. Baek and Y. T. Park, "Probabilistic graphical model based personal route prediction in mobile environment," Applied Mathematics and Information Sciences, vol. 6, No. 2S, pp. **651S-659S**, January **2012**.

[23]. Mingqi Lv, Yinglong Li, Zhenming Yuan and Qihui Wang, "Route Pattern Mining From Personal Trajectory Data", Journal of Information Science and Engineering, vol.**31**, pp.**147-164**, **2015**.

[24]. Seongwon Min, Jong-Yong Lee and Kye-Dong Jung, "Real-time path prediction and grid-based path modeling method using GPS," International Journal of Applied Engineering Research, Vol.**12**, No.**20**, pp. **9997-10001**, **2017**.

[25]. Vishnu Shankar Tiwari and Arti Arya, "Horizontally scalable probabilistic generalized suffix tree (PGST) based route prediction using map data and GPS traces," Journal of Big Data, vol.**4**, No. **23**, **2017**.

[26]. Vishnu Shankar Tiwari and Arti Arya, "Distributed Context Tree Weighting (CTW) for route prediction," Open Geospatial Data, Software and Standards, vol. **3**, No**10**, **2018**.

[27]. Shun Taguchi, Satoshi Koide and Takayoshi Yoshimura, "Online Map Matching With Route Prediction," IEEE Transactions on Intelligent Transportation Systems, vol.**20**, No.1 pp.**338-347**, Jan. **2018**.

[28]. Sudhir Kumar Adhlaka, Neelam Duhan, Komal Kumar Bhatia and Himanshu Sharma, "Route Prediction Techniques using GPS Traces and Spatial Data," Proceedings of INDIACom IEEE International Conference on "Computing for Sustainable Global Development", BVICAM, New Delhi, 13th-15th, March, pp.**1475-1481**, **2019**.

[29]. Zaki, J. Mohammed, "SPADE: An Efficient Algorithm for Mining Frequent Sequences." in proceedings of Machine Learning, vol.**42**, pp. **31-60**, **2001**.

AUTHORS PROFILE

Nidhi Soni is currently pursuing her M.Tech. degree in the Computer Engineering Department at the University Institute of Engineering and Technology (UIET), Kurukshetra. UIET was formerly known as the University Institute of Engineering and Technology, affiliated with Kurukshetra University, in Kurukshetra, Haryana, India. In 2020, she successfully graduated with a B.Tech degree in Computer Science and Engineering from the same institution, UIET, Kurukshetra University. For any communication or inquiries, Nidhi can be contacted via email at soninidhiverma3@gmail.com or by phone at 7988617358.

Dr. Ajay Jangra is a highly accomplished academic with extensive qualifications in Computer Science and Engineering, including a B. Tech., M. Tech., and Ph.D. He also holds an MBA in IT and an LLB. With over 15 years of work experience, Dr. Jangra currently serves as a Professor in the Computer Engineering Department at the University Institute of Engineering and Technology (UIET), Kurukshetra. He is an esteemed researcher and has published over 40 research papers in reputable journals and conferences. His fields of specialization include Computer Networks, Cloud Computing, Graphics and Animation, and Digital and Data Communication. Dr. Jangra has provided supervision and guidance for over 25 M.Tech. dissertations, showcasing his expertise in mentoring students. For further communication, he can be reached at ajangra2015@kuk.ac.in or contacted via mobile at 9466027922.