Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 8, Issue. 4, April 2019, pg.16 – 22

ENSEMBLE HUMAN MOVEMENT SEQUENCE PREDICTION MODEL USING APRIORI AND BAGGED J48 ON MACHINE LEARNING

Dr. M.Gobi[1]; A.Kowshika[2]

Assistant Professor, Research Scholar

[1]Department of Computer Science, Chikkana Govt Arts College, Thirupur-641602

[2]Department of Computer Science, Chikkana Govt Arts College, Thirupur-641602

[1] Mgobimail@yahoo.com; [2] kowshikasurya77@gmail.com

Abstract- The accurate prediction of human movement trajectory has a variety of benefits for many applications such as optimizing day-to-day home movements of old or disabled people to minimize their routine efforts, etc. In order to perform human movement prediction, large amount of historical positioning data from sensors has to be collected and mined by analyzing different human sequential movement prediction approaches and their limitations. In this proposed work, Apriori algorithm which predicts the human movement sequence patterns in indoor environment. The Apriori is integrated into Bagged J48 Machine learning algorithm which results in an ensemble model to predict the future human movement patterns. These patterns are mined based on spatial, temporal and social data which add more accuracy to our prediction. This model also performs clustering mechanism to detect the abnormal patterns.

Keywords- Data mining, Machine learning, Spatial-temporal-social data, Trajectory analysis, Human movement sequence prediction.

I. INTRODUCTION

Data mining is one of the steps for future prediction and knowledge discovery. It is a computational process of discovering patterns in large historical data. Machine learning concerns the construction of systems that can learn from data and act without being explicitly programmed. The human movement pattern prediction has attracted good research interest due to its vast applications. Movement pattern of human's routine life can be mined from their historical collection of movement data. By acquiring the knowledge of people's future sequence of movements, one can improve their convenience and safety. People have their own well-defined criteria for choosing a path when moving from one place to another. For a school student or an office employee who is getting late for the school or the office, the shortest path may hold the utmost importance; a fitness freak or a health conscious person may choose a longer route; an elderly would most likely choose a well-lighted path with least hurdles and so on. But visually impaired people are not privileged to make their own choices when navigating from one place to another. They were try to memorize routes for every environment they visit which besides been a tedious task.

Various applications of human sequence movement prediction are, urban service based applications and some of the fields like healthcare, science, bank, public services, etc., which helps in improving their service by uncovering new patterns .The data which influences human movement pattern are spatial, temporal and social data Hence, by introducing a new ensemble model to mine spatial, temporal and social data of the people for their future movement prediction. The human movement prediction could eventually assure the convenience, precaution and safety of the people. By mining the human movement patterns, the day-to-day routine movement of humans can be extracted for further analysis. The mined pattern gives us the locations which are frequently visited by a person. These frequently visited home locations are termed as Point Of Interest's (POI's). For example, a aged or disabled people in a home walks through different locations and performs various movements to do some work. By predicting the movement sequence of the aged or disabled person, POI's can be arranged optimally according to the predicted sequence of locations.

Aged or disabled care is yet another application of human movement pattern prediction. By reducing the movement of aged people, their effort to do day-to-day routine work is minimized. These applications are also beneficial to the common people, commercial organizations, government agencies, health care units, safety critical systems and context-aware applications such as recommender systems, social networking applications, reminder applications and ambient assistive devices. The backdrop consists of locations, people, Point of Interest's and each individual's trajectory data collected from sensors. The POI's are framed by analysing the most frequently visited locations in the entire record.

II. BACKGROUND AND MOTIVATION

By mining these patterns, one can arrange things accordingly that it matches the sequence or order in which locations are visited. People tend to spend significant amount of time only in certain specific locations. Those specific locations are called stationary locations like place of work, home and other staying places. Although there are people who do not follow the suggested locations, it suits to most of the people in majority of the circumstances. User movement varies from time to time, i.e., early morning, morning, afternoon, evening, night, midnight, working hours, days, and seasons within a temporal period. Therefore, the POI probability calculation with respect to space, time and date is necessary. Observations are carried over in each time interval specified. For instance, the time intervals can be classified as week days and weekends. The POI probabilities are calculated within the time periods and day profiles. The evaluation is performed as 24x7 following the strategy of each day of the week as day profiles and one hour time as time period. The possibility of clustering has to be considered, such as most people do not move at night times and some peoples working hour movements will be same if they ought to do the same job. There are different data clustering methods which can be applied to classify the movement records based on movement pattern similarities. The following fig is the example.

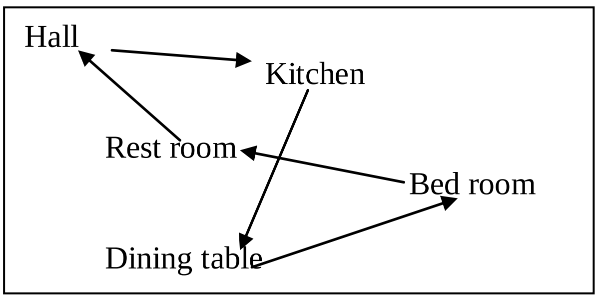

Fig.1 Actual placement of POI's.

The scenarios is given above, The locations are Home, school, classroom, library, canteen, ground and home.

1. Person walk off from hall.
2. He go to the kitchen.
3. Reach his dining table.
4. And he move on to bed room.
5. Then he walks to the rest room.
6. At last he walks through the hall.

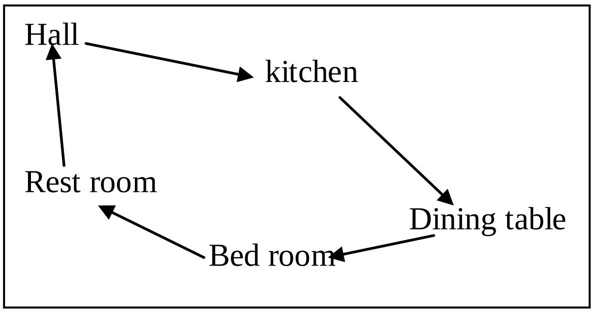

Fig.2 After updating locations using frequently visited

By observing the above sequence with the Fig. 1, the person has to walk through a long distance to complete a task. Fig.2, After the update of locations using frequent visited area motivates us to optimize the movements through which can save time and energy.

III. APRIORI ALGORITHM

Apriori algorithm which is easy to execute and very simple, is used to mine all frequent item sets in database. This algorithm makes various searches in database to discover frequent item sets where k-item sets are used to generate k+1-itemsets. Each k-item set must be greater than or equal to minimum support threshold to be frequency. Otherwise, it is called candidate item sets. In the first, the algorithm scan database to find frequency of 1-itemsets that contains only one item by counting each item in database. The frequency of 1-itemsets is used to find the item sets in 2- item sets which in turn is used to find 3-itemsets and so on until there are not any more k-item sets.

Definition 1: Suppose T={T1, T2, … , Tm},(m≥1) is a set of transactions, Ti= {I1, I2, … , In},(n≥1) is the set of items, and k-item set = {i1, i2, … , ik},(k≥1) is also the set of k items, and k-item set ⊆ I.
Definition 2: Suppose σ (item set), is the support count of item set or the frequency of occurrence of an item set in transactions.
Definition 3: Suppose Ck is the candidate item set of size k, and Lk is the frequent item set of size k.
In this proposed approach, by enhance the Apriori algorithm to reduce the time consuming for candidates item set generation. It firstly scan all transactions to generate L1 which contains the items, their support count and Transaction ID where the items are found. And then use L1 later as a helper to generate L2, L3 ... Lk.

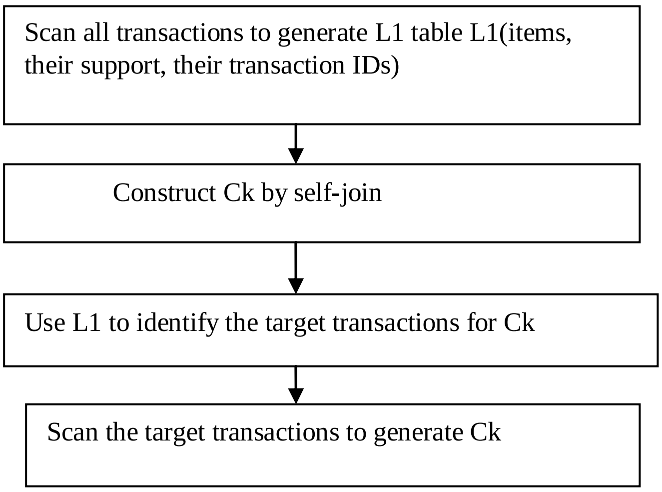

Fig:3 Steps for Ck generation

When want to generate C2, we make a self-join L1 * L1 to construct 2-itemset C (x, y), where x and y are the items of C2. Before one can scanning all transaction records to count the support count of each candidate, use L1 to get the transaction IDs of the minimum support count between x and y, and scan for C2 only in these specific transactions. The same thing for C3, construct 3-itemset C (x, y, z), where x, y and z are the items of C3 and use L1 to get the transaction IDs of minimum support count between x, y and z, then scan for C3 only in these specific transactions and repeat these steps until no new frequent item sets are identified.

The Improved Algorithm Of Apriori:

//Generate items, items support, their transaction ID
(1) L1 = find_frequent_1_itemsets (T);
(2) for(k = 2; Lk-1 ≠ Φ; k++)
//Generate the Ck from the Lk-1
(3) Ck = candidates generated from Lk-1;
//get the item iw with minimum support in Ck using L1, (1≤w≤k).
(4) x = Get _item_min_sup(Ck, L1);
// get target transaction IDs that contain item x.
(5) Tgt = get_Transaction_ID(x);
(6) For each transaction t in Tgt Do
(7) Increment the count of all items in Ck that are found in Tgt;
(8) Lk= items in Ck ≥ min_support;
(9) End;
(10) }

Firstly, scan all transactions to get frequent 1-itemset which contains the items and their support count and the transactions ids that contain these items, and then eliminate the candidates which are less frequent or their support are less than min_sup.

IV. BAGGED J48 MACHINE LEARNING ALGORITHM

Bagging is an ensemble method that creates separate samples of the training dataset and creates a classifier for each sample. Bootstrap aggregation is also called as bagging ensemble meta data algorithm designed to improve the stability and accuracy level of the machine learning algorithm used in statistical classification and regression. The results of these multiple classifiers are then combined (such as averaged or majority voting). That each sample of the training dataset is different, giving each classifier that is trained in different focus and perspective on the problem.

Ensemble techniques are used for improving the performance and result accuracy of the normal classifiers. The accuracy is increased by combining the decisions of different classifiers into a single combined classifier. This bagged J48 algorithm works by applying the two ensemble learning approaches, bagging and boosting on five traditional classifiers. The performance is evaluated in terms of F-measure. This proposed work is automated which increases the accuracy of the final outcome. Finally, data set is formed and is used to verify the prediction. The prediction results can be compared with the training data sets and test data with the help of the association rules. Hence the bagged algorithm has to integrate with the apriori algorithm into a common and uniform representation.

V. APRIORI ALGORITHM INTO BAGGED J48

In this paper, propose Apriori algorithm which predicts the human movement sequence patterns in indoor environment. To extend our approach into machine learning mechanism, by integrate Apriori into Bagged J48 Machine Learning algorithm which results in an ensemble model to predict the human movement patterns. The patterns are mined based on spatial, temporal and social data which adds accuracy to our prediction results. The different phases are data collection, data pre-processing, classification, association rules formation, prediction, similarity computation with the traditional Apriori algorithm and clustering of abnormal patterns. Ensemble techniques are used for improving the performance and result accuracy of the normal classifier. This bagged J48 algorithm works by applying the two ensemble learning approaches, bagging and boosting. J48 algorithm acts as the base classifier, using which integrated into apriori algorithm. Finally, the combined ensemble model is formed and test data is used to verify the prediction. The next location is predicted by finding the matches between a person's current trajectories with the frequent patterns.

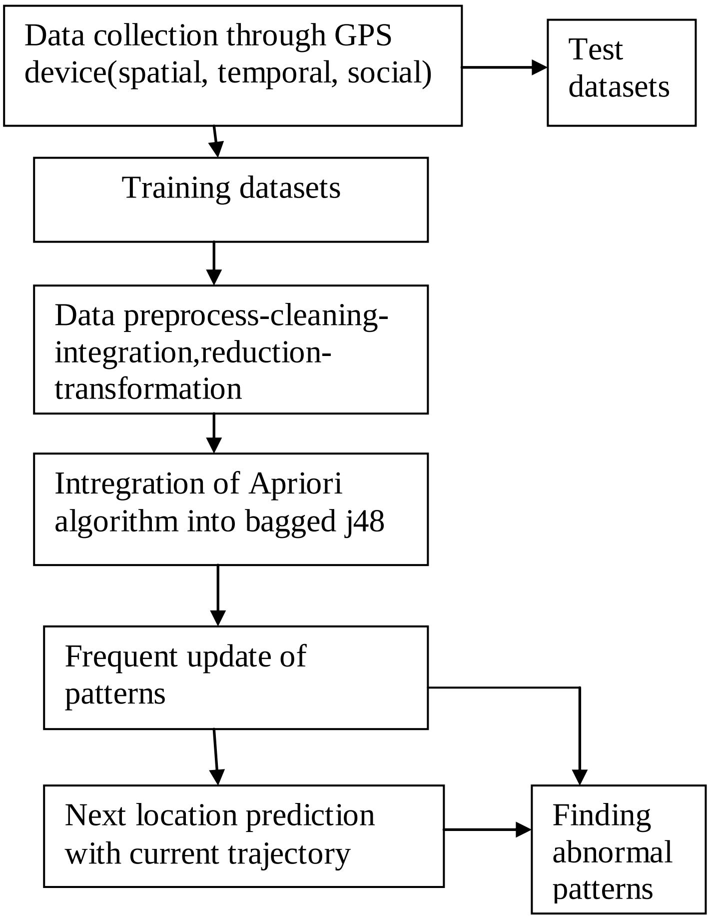

Fig.3 Data collection and frequent update of patterns

SAMPLE COLLECTION OF DATA SETS.

Movement patterns	Order in which the locations are visited (Home Scenario)
1	kitchen, dining table, hall
2	kitchen, dining table, bed room, kitchen
3	dining table, rest room
4	kitchen, dining table, hall

Trajectory data collection from GPS history The training data is collected from the GPS device. From the data collected, the activity transitions (semantic data) are considered for further processing. Since, the activity transitions are more regular than the location transitions. Here, transition denotes the movement from one location to another. In the data pre-processing phase, all the four pre processing techniques such as cleaning, integration, reduction and transformation are performed.

Movement pattern mining In order to mine and extract the frequent patterns, which have the probability value of each location. Once the apriori trained with the frequent patterns, the test data can be used to compute the result.

Association rules formation It is a classic algorithm used in data mining for learning association rules. It is nowhere as complex as it sounds; on the contrary it is very simple. Suppose there are large set of movement pattern records in a home routine which is denoted in Fig. 1,

- Learning association rules basically means finding the locations that are visited together more frequently.

- For example, in the observation that kitchen and dining table are visited together frequently. Association rules are formed in this manner and the things are altered in such a way that it matches the frequently visited pair of locations.
- Based on a simple golden rule: it is said that a location is visited at least 60% of the time (relative to the highest visited location) that can be considered as frequently visited location and further processing is carried over only on those frequently visited locations. The remaining locations which occurred less than 60% are eliminated.

VI. CONCLUSION AND FUTURE DIRECTIONS

This proposal of an ensemble model for human movement sequence prediction based on spatial, temporal and social data. The model is proposed mainly to assist the aged or disabled people by optimizing their day-to-day movement patterns. The datasets from UC Irvine Machine Learning Repository and modified it by adding the temporal and social information. The Apriori algorithm which considers the spatial, temporal and social data of the users as the input. This algorithm is integrated into the bagged J48 machine learning algorithm to obtain better results. By using machine learning algorithms and the human movement, the final results yield a better outcome. The ensemble model used to obtain the frequently visited locations and optimize the movements sequence of disabled people. And it also predict next location based on current trajectory area. Lack of social data usage for human movement prediction is addressed in our model. Our future work is to concentrate more on the social influence of human movements by incorporating artificial intelligence to a prototype.

REFERENCES

[1] Andrienko,N., Andrienko, G.,2008. Basic concepts of movement data. Mobility Data Mining Privacy,15–38.
[2] Banitaan, S., Azzeh, M., Nassif, A. B., 2017. Users Movement Prediction:- the contribution of machine learning techniques. In: Proceedings - 2016 15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, pp. 571–575.
[3] Boldrini, C., Passarella, A., 2010. on data modelling , spatial and temporal properties of human mobility driven by users social relationships. Comput. Commun. 33 (9), 1056–1074.
[4] Cho, E., Myers, S.A., Leskovec, J., 2011. **Users Movement in Locations** Based Social Networks. Proceed of 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1082– 1090.
[5]Aha, D. and D. Kibler (1991) Instance-based learning algorithms. In Machine Learning, 6: 37{66.
[6] Breiman, L. (1996) Bagging Predictors. Machine Learning 24(2): 123{140.
[7] Dietterich, T. G. (1997) Machine-Learning Research: Four Current Directions. AI Magazine 18(4): 97{136.
[8] J. Han, M. Kamber, Data Mining: Concepts and Techniques, second ed., Morgan Kaufman, San Francisco, 2006.
[9] Y.L.Chen, C.C.Ling, Algorithms for mining association rules in bag databases, Information Sciences 166(2004) 31-47.
[10] Y.C Hu, G.H. Tzeng, C.M. Chen, Deriving two-stage learning sequences from knowledge in fuzzy sequential pattern mining, Information Sciences 159(2004) 69-86.
[11] K. Lagus, S. Kaski, T. Kohonen, Mining massive document collections by the websom method, Information Sciences 163(2004) 135-156.
[12] Y. Li, S. Zhu, X.S. Wang, S. Jajodia, Looking into the seeds of time: discovering temporal patterns in large transaction sets, Information Sciences 176(2006) 1003-1031.
[13] J. Berlin ska, M.Drozdowski , Scheduling divisible Map Reduce computations, Journal of Parallel and Distributed Computing 71(2011) 450- 459.
[14] L. Breiman, "Bagging Predictors," Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.
[15] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5- 32, 2001
[16] J.R.Quinlan, C4.5, Bagging, and Boosting., Proceedings of the National Conference on Artificial Intelligence, 1996, pp. 725-730.
[17] L. Breiman, "Bagging Predicators, Machine Learning, 1996, Vol. 24, No. 2, pp. 123-140.
[18] K. Bache and M. Lichman. UCI machine learning repository, 2013
[19] Thomas G. Dietterich. The Experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning, 40(2):139–157, 2000.
[20] Quinlan, J.R.: C4.5: programs for machine learning. Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
[21] Cohen, W.W.: Fast effective rule induction, A., Russell, Proceed of the 12th International Conference on Machine Learning, Tahoe City, CA, pp. 115–123. Morgan Kaufmann, San Francisco (1995)
[22] D., Albert, Aha, D.W, Kibler, M.K.: Instance-based learning algorithms. Mach. Learn. 6(1), 37–66 (1991)

[23] M. Kamber, J. Han, "Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Book, 2000.
[24] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, Data mining to knowledge discovery in databases, AI magazine, vol. 17, no. 3, p. 37, 1996.
[25] Y. H. Jbara, F. H. AL-Zawaidah, and A.L.Marwan, An Improved Algorithm for Mining Association Rules in Large Databases," Vol. 1, No. 7, 311-316, 2011
[26] T. C. Corporation, "The Introduction to Data Mining and Knowledge Discovery, Two Crows Corporation, Book, 1999.