Indexing Strategy to speed up Computation: Efficient storage of huge volume of trajectories is a big challenge. We have introduced UFHM *(User Footprint Hash Map Structure)* to index all geo-tagged information along with user details. $UFHM(H) = \{(B_1, L_1), ..., (B_n, L_n)\}$. H is a chain hash map structure and each bucket (B_i) of UFHM is associated with a list (L_i). $B_i =< Lat, Long, place >$: each tuple stores latitude, longitude along with land use information. Each list contains user details, geo-tagged places and timestamp of visiting the place. $L_i =< u_i, Geo_{tag}, t_i >$. A pairing function $H(l_x, l_y)$ is used as the hash function to map normalized latitude (l_x) and longitude (l_y) in *UFHM.*

$$H(l_x, l_y) = (l_x + l_y)(l_x + l_y + 1)/2 + l_y \quad (1)$$

UFHM structure provides efficient access to a particular geographic region and interesting land use information.

Spatio-Temporal Clustering and Extracting Common pattern: To analyse movement pattern, TNTG *(Temporal Node based T-Graph Structure)* is generated for each user's trajectory path. $TNTG = \{(V, E)|1 < v_i < |V|, 1 < e_i < |E|\}$, where each node $v_i \in |V|$ denotes stay point of the trajectory segment. We have represented v_i as **Temporal Node** (TN). $TN =< node_{id}, user, time >$. Each TN stores footprints of users who visited that particular node, a unique node id and stay duration. Directed edges from one (TN) to another defines the transition from one place to another place based on time series of GPS traces. To cluster trajectory segments, location (geographical and land use) information, time duration in a stay-point, speed of movement, transportation mode and direction of the trajectories are considered. Each TN and edge of TNTG are weighted based on the above features. To measure similarity among trajectory segments of different users, minimum stay duration in a common node is computed along with other mentioned features, namely geo-tagged information of the stay points (e.g., university, restaurant), transportation mode (e.g., cycle, car), timestamp of the visits etc. Based on the similarity measurement the users' trajectories are clustered.

For finding common patterns among a set of trajectories an extension of LCS (*Longest Common Sub-sequence*) problem, namely *Temporal Common Sub-sequence* (TempCS) is introduced. $TempCS(X_i, Y_j)$ finds common sub-sequence among trajectories X and Y with i and j stay points respectively as depicted in (2).

$$TempCS(X_i, Y_j) = \begin{cases} 0 & if(i == 0) \\ & or(j == 0) \\ TempCS(x_{i-1}, y_{j-1}) & if((x_i == y_j) \\ +Min(XT_{Score_i}, YT_{Score_j}) & and(x_{i+1}) \neq (y_{j+1})) \\ TempCS(X_{i-1}, Y_{j-1}) + C\times & if((x_i == y_j) \\ Min(XT_{Score_i}, YT_{Score_j}) & and(x_{i+1}) == (y_{j+1}) \\ MAX(TempCS(X_{i-1}, Y_j), & \\ TempCS(X_i, Y_{j-1})) & if(x_i \neq y_j) \end{cases} \quad (2)$$

TempCS for a cluster of users trajectory returns a set of TN, average stay duration in each TN and directed edges which represents common trajectory path followed by all users in the particular cluster.

Categorization and Correlation: Each common trajectory path of clusters is analysed to categorize users based on the movement patterns. In our experiment (using GPS dataset of Microsoft GeoLife [2]), users' footprints on frequently visited places, namely student dormitory, professor office, laboratory etc. are analysed. Users are categorized into four broad categories, namely $Student$, $Professor$, $Staff$ and *Guest* according to the frequency of visiting places, timestamp of the visit and few assumptions, like students visit library, laboratory, student-cafe more frequently than professor/staff etc.

Given a dataset of *user-user* or *user-place* or *place-place* pair, association/correlation can be automatically determined using THUMP framework.

3. EVALUATION OF THUMP

In this section, we evaluate the effectiveness of the THUMP framework using the real data set of GeoLife Trajectory [2] with GPS traces of 182 users in a period of over five years around Beijing, China and over 17,621 trajectories. Few representative results are presented here.

1. Computational speed-up: Using *UFHM*, search time reduction is captured in a dataset of 258 places/ locations.

Type of Query	Search Time (UFHM)	Linear Scan (Naive)
Point	6.3s	18.5s
Range	8.6s	41.2s

Table 1: Computational efficiency of THUMP

2. Association/Correlation: Association/Correlations between GPS traces of users are determined from *TNTG* and *TempCS*. One of the generated association rules is shown below:

$$Users(Visits\ Health - care\ center\ frequently) \rightarrow$$
$$Users(Less\ Hospital\ Visits)$$
$$[Support : 70\%, Confidence : 80\%] \quad (3)$$

4. CONCLUSION

This paper presents a novel framework for analysing the mobile trajectory traces in spatio-temporal context, model human movement patterns using semantic aspects and categorize people based on these patterns. It also helps in extracting interesting association rules from the GPS traces.

5. REFERENCES

[1] S. Shang, R. Ding, K. Zheng, C. S. Jensen, P. Kalnis, and X. Zhou. Personalized trajectory matching in spatial networks. *In The VLDB Journal, The International Journal on Very Large Data Bases*, 23(3):449–468, 2014.

[2] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. *In Proceedings of the 18th international conference on World wide web*, pages 791–800, 2009.