Relevance and Frequency enabled Trip Planning Model based on Socio – Economic Status

Sesham Anand
M.V.S.R Engineering College
India
seshamanand282015@gmail.com

Padmanabham. P, Govardhan. A
BIET-JNTUH
India

Rajesh Kulkarni
TSSMS BSCOER
India

ABSTRACT: *Planning a trip not only depends on the travelling cost, time and path, but also the socio-economic status of the traveler. This paper attempts to introduce a new trip planning model that is able to work on real time data with multiple socio-economic constraints. The proposed trip planning model processes the real time data followed by extracting the relevant socio-economic attributes and mine the most frequent and feasible attribute to plan the trip. The relevance of the socio-economic constraints is defined by correlation, whereas the frequent and feasible attributes are mined using sequential pattern mining approach. Real time travel information about 38303 trips is acquired from Hyderabad city of India and the proposed model is subjected to experimentation. The proposed model maintains a substantial tradeoff between multiple performance metrics, though trip mean model performs statistically.*

Keywords: Correlation, Pattern, Socio-Economic, Frequent, Trip, Planning, Mining

Received: 2 June 2017, Revised 30 June 2017, Accepted 5 July 2017

1. Introduction

Ordinary railway as well as the intercity railway, buses, taxis and similar other vehicles that come under the category of urban hubs of passenger transport readily require an intelligent transport management system to handle the abrupt raise in the networks that involve the transport as well as the traffic tools [1] [2]. The Intelligent transportation systems (ITS) are global systems, which draw the attention of automotive industry, transportation professionals and political decision makers from all around the globe [1].

A larger number of countries, which include Europe, USA and Japan [6] in particular, perform a wide variety of research-oriented tasks that are highly focused towards exploring and setting up ITS. Executing such tasks allow the transmission technology of data communication, advanced information technology, electric sensor technology, computer processing technology and electric control technology to integrate and apply over the entire transport management system in an effective way, in addition to setting

up a complete transport management system that is direct, precise and effective [4].

The ITS is found to have emerged in 1930s and only in a gradual sense, it has influenced the current world [3]. This 21st century is greatly dedicated to road transport intelligentization and hence, the setting up of intelligent transport management systems and transport management systems that are increasingly integrative and comprehensive in accordance with the modem electric information technology have gained utmost interest [1]. Yet, several countries of the modern era are attempting to restructure the present transport system and transport management system using novel and sophisticated technologies for yielding a considerably large traffic capacity and quality of service from the road networks [3]. The need for an efficient ITS is trivial in case of public transportation systems as well as the private sectors like factories, harbors and many more [12].

Despite the fact that the methods, which are presented in the literature, determine the travel time reliability [8], the generation of realistic reliability measures in the output of traffic simulation models and planning models continue to be a serious issue.

1.1 Motivation

Day by day, plenty of issues are found to be produced in transport and transport management. Serious interference that exists among the regional distributed traffic and cross-border traffic, heavy traffic load, inadequate static traffic facilities and severe conflicts that result from person/vehicle mixed travel [2] can be stated as few instances. These issues are created due to the urban traffic hubs, which possess a bulkier traffic and entirely overlapped travel space of a variety of transportation means that include bicycles, public transport and minibuses. The main cause for the traffic and vehicle congestion is the progression in the socio-economic factors [11]. So, in case of urban sectors, the intelligent transport management system that effectively controls the transport system has become highly crucial [2].

With ITS, it is possible to bring about the enhancements in driver support, road transport and mobility. In the upcoming years, the potential investments on ITS would rise in a rapid manner. During the appraisal phases of ITS projects, various kinds of assessments that pertain to technology, user acceptance, traffic, environment and socio-economy should be considered. But, in most of the IT assessments, only a few of the above-mentioned factors are addressed with little consideration on the socio-economic factors.

The socio-economic assessments play a lead role in making government policy decisions. In fact, while appropriate evaluation guidelines for ITS projects in United States and Europe were formulated, extensive works have been carried out previously and at present in relation to the socio-economic assessments. Yet, the information regarding the way the impacts can be evaluated is missing and in addition, it was more troublesome to measure or define several of the benefits in a required form. Numerous efforts have been put forth for disclosing a wide variety of potential advantages, while having decreased level of stress over the cost. Further, the comparison between the results of two different projects is not an easy task because the projects would have their own guidelines and cost as well as benefit evaluation schemes [13].

2. Related Works

Na ZENG *et al* [2] have insisted the need to have an intelligent management system in the sophisticated urban transport system. The authors have exploited a number of technologies, wherein, the objective was to design and create an intelligent transport management system that is more vital in traffic hubs. They have also addressed the problems that are associated with design, framework and functional modules, so that the degree of exploitation, protection and ease can be enhanced along with the improvement in the government's decision making ability.

Jing Dong and Hani S. Mahmassani [5] have developed ITS system through the inclusion of vehicular technology, in which they have considered mobility and speed of the vehicles to model the breakdowns that occur in traffic very often. The objective of this scheme was to envisage the change in the travel time, which was produced due to this kind of stochastic events. The breakdown was supposed to be occurring at varying flow limits with some probability and further, it was found to prolong for an arbitrary amount of time. The modeling at the microscopic level was achieved through the consideration on the variations in speed, wherein, a leading vehicle has caused the initiation and the following vehicles that contain correlated-distributed behavioral parameters have enabled the propagation. The numerical results that were obtained using the Monte Carlo simulation has revealed the effectiveness of the proposed stochastic modeling approach in offering the realistic macroscopic traffic flow behavior and generating travel time distributions.

T. Korhonen *et al* [9] have inspected the state of telecommunication technology, service and economy of ITS in Helsinki metropolitan area council (termed as YTV). As a result of this investigation, they have suggested a number of ways to develop a triumphant ITS and they are: (1) The networking services/ technologies have to be acquired from the telecommunication operator; (2) The networks have to be constructed individually and (3) A hybrid solution has to be produced. The authors have recommended the exploitation of the characteristics of timeline diagrams and investment sensitivity estimations for attaining a successful decision making process in the planned ITS.

V. Di Lecce and A. Amato [10] have introduced the flexible ITS, in which they have determined a suitable route for every vehicle through the mitigation of the impact that is produced from transporting hazardous material. They have implemented a negotiation process between the intelligent agents. The system keenly observes the route that every single vehicle follows and checks whether this route and the previously specified route matches or not. An increasingly flexible on-board unit that is placed in the vehicles serve as the most important component of ITS. It owns a modular structure, wherein, several sensors can be linked as per the application requirement. Here, a range of parameters that include the position, speed, load balance and accelerations of the vehicles were examined in detail. With this information, the system can detect few characteristics like, the activities of the driver and the risky operating condition of the vehicle. The authors have put forth a multi-level metaphor-based graphic user interface (GUI) to make a representation of this information. The initial stage of the interface offers a rapidly understood view of the state that is under observation and it relies on the on-site first response as well as the route monitoring operators. The second level gives the information in a large and thorough form to the incident managers and the additional operators. A human panel is employed, while the interface is being assessed. Once a volunteer has made use of the interface for half an hour, he/she would compile a questionnaire. The objective of this work is to jot down few methods that can be employed for route planning and the user interface.

Junping Zhang *et al* [14] have systematically studied the development of data – driven ITS, under which they have described about the various functions that are related to the key elements and additionally, the problems that are encountered at the time of deploying was also depicted. Their review has summarized the key research gap and the directions that are persisting for further development in data-driven systems.

Neal Lathia *et al* [15] have worked on personalized trip time estimation model to notify the traveler. They have acquired the travel history of the user for developing such personalized trip time estimation model. They have proposed a prediction model to predict the personalized trip times for the travelers and it is followed by a ranking method to rank the stations based on the interest of the travelers and hence, the future mobility patterns can be predicted.

2.1 Review

The contributions that have been associated with ITS is vast and diverse. The researchers have attended various motivation scenarios to improve the ITS. However, the primary objective of providing a smart ITS to facilitate the travelers have to be achieved. For instance, in [5], the authors have worked to simulate the vehicle breakdowns in a congested road plot. Though the simulation has based its version on vehicular mobility modeling, the primary intention is to aid the precision of predicting the travel time reliability. However, it is complex and imprecise apart from analytical models. Similarly, transportation of hazardous materials has been seriously considered in [10] to aid safe route planning, monitoring and accident management. This reveals that an urban planning decision system depends on the simulation environment, which should consider multiple parameters. For instance, a traveler's decision depends on numerous factors such as, the travel time reliability [5] [7]. Hence, precise modeling of the entire travelling details is significant for any trip or transportation planning system.

Few personalized prediction models [15], which are reported in the literature, have also found our attention greatly because they have been associated with the travel history and the statistical analysis. However, these models remain static. For instance, the trip familiarity, the trip context and the combined models are experimented for the ability to handle nonlinear and diverse data characteristics such as, the travelers' personal details, transportation mode of interest, etc.

3. Article Overview

3.1 Problem Formulation

Let us assume that a traveler wants to have a trip from source S to destination D. The trip (S,D) includes numerous socio-economic constraints such as the vehicle to be used, travelling stages, availability of parking facilities, travelling cost, parking cost, etc. These socio-economic constraints remain unpredictable. The traveler will be benefited if any of these socio-economic constraints

are precisely estimated and recommended for the trip. For instance, if the traveler comes to know the availability of the parking facilities and its usage in the past trips, then it will be highly helpful on deciding the trip, mode of travel, travel cost and many more other constraints. However, it is not just a process of mining the historical data. Moreover, it is challenging to acquire the relevant socio-economic constraints that plays key role in the trip.

3.2 Our Contributions

Our contributions in the paper is listed as follows

Contribution 1: An in-depth survey is carried out to acquire the real time data about the trips carried out by various travelers.

Contribution 2: A recommendation system is introduced in which the algorithm for mining interesting, significant and relevant socio-economic constraints associated with the trip is proposed.

Contribution 3: Interestingness metric is proposed to evaluate the significance of the mined socio-economic patterns.

Contribution 4: Performance consistency and reliability are investigated using polynomial fitting model.

3.3 Data Acquisition

The data acquisition is the significant task in our work. We have collected the data from the Hyderabad city of India. The entire Hyderabad city has been divided into 147 zones from where the travelers are met and they are asked to fill up four individual forms. Form 1 and 2 collect the family details and personal socio-economic status of each traveler and the way in which they are connected with the travel. The socio-economic status of the family includes the education level, income level, satisfactory level, vehicle details, travelling schedule and travelling cost. Form 3 and Form 4 provide the specific information about the trip. They have the details of every individual trip of the travelers of these zones, their travel details such as origin, destination, and vehicles used, parking facilities, travel comfort, travel cost, parking cost, parking and travelling stages, etc. Totally, the information of 48853 trips has been acquired, where the trips have been taken place within 303 places of Hyderabad city. Since the entire details have been filled up in the user understanding format, it is essential to process as per the machine understandable format. In the rest of the paper, the socio-economic attributes of the trip and the travelers will be simply called as attributes or characteristics.

4. Our Methodology

4.1 Preprocessing

The sequence of steps involved in our methodology is illustrated in Figure 1. The first step in the proposed methodology, termed as acquisition of raw data, has been described in the earlier Section. The data preprocessing steps include extracting the socio-economic attributes of user interest and organizing them along with the origin and destination of the trips so that further processing can be performed easily. Despite numerous details about the trip and the traveler information have been collected, this paper considers 29 most significant attributes along with the origin and destination of the trips. The considered 29 attributes are listed in Table I. The data preprocessing step includes three stage processes, named as filtering, filling and splitting.

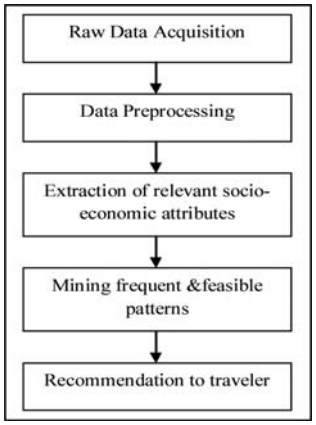

Figure 1. Sequential steps involved in the proposed methodology

Attribute Name	
1. Purpose of travel	
2. Waiting time in Stage I	16. Waiting time in Stage III
3. Stage I distance	17. Stage III distance
4. Mode of travel in Stage I	18. Mode of travel in Stage III
5. Travel time in Stage I	19. Travel time in Stage III
6. Type of parking in Stage I	20. Type of parking in Stage III
7. Cost of parking in Stage I	21. Cost of parking in Stage III
8. Travel cost in Stage I	22. Travel cost in Stage III
9. Waiting time in Stage II	23. Waiting time in Stage IV
10. Stage II distance	24. Stage IV distance
11. Mode of travel in Stage II	25. Mode of travel in Stage IV
12. Travel time in Stage II	26. Travel time in Stage IV
13. Type of parking in Stage II	27. Type of parking in Stage IV
14. Cost of parking in Stage II	28. Cost of parking in Stage IV
15. Travel cost in Stage II	29. Travel cost in Stage IV

Table 1. Selected Attributes Associated With Every Trip

Filtering: Since the raw data are filled up on field, it includes numerous erroneous as well as missing fields. The trips that have only complete information are acquired for further processing. Among 48853 trips, 38303 trips are found to be complete and useful. For these 38303 trips, the 29 attributes are acquired and the database $D^{raw}:\left|D^{raw}\right| = M^{raw} \times N^{raw}$ is constructed. Each element of D^{raw} is represented as: $d_{ij}^{raw}: 0 \le i \le M^{raw}-1, 0 \le j \le N^{raw}-1$, where M^{raw} and N^{raw} refer to the total number of trips and fields in the raw dataset, respectively. Here, $M^{raw} = 38303$ and N^{raw} consist of source field, destination field and 29 attributes. Hence, $N^{raw} = 31$.

Filling: The D^{raw} is further subjected to filling process in which irrelevant stage information is filled up with zero. For instance, trip 1 starts from location *A* to location *D* and trip 2 starts from location *A* to location *B*. Trip 1 has two stage travels, i.e., from location A to location *C* and then from location *C* to *D*, whereas trip 2 has only one stage, i.e. directly from location *A* to location *B*. In these circumstances $d_{1\forall j}^{raw}$ has the filled fields related with stage 1 and stage 2, but $d_{2\forall j}^{raw}$ has incomplete information in the fields related with stage 2. These fields are filled up with zeros.

Splitting: The splitting process segregates D^{raw} into multiple sub-dataset d_{tid}^{proc}, where $0 \le m^{(tid)} \le M_{tid}^{proc}-1$ and $0 \le n^{(tid)} \le N_{tid}^{proc}-1$. Each sub-dataset consists of multiple trips with same origin and destination. The pseudo code of the algorithm used to split the D^{raw} and to construct D^{proc} is given. The splitting algorithm first extracts the set of origins $\{O\}$ and destinations (targets)$\{T\}$ from D^{raw}.

The representation of $\{O\}$ and $\{T\}$ can be given as $\{O\} \in d_{j=1} \forall i$ and $\{T\} \in d_{j=2} \forall i$, respectively. Hence, the $\{O\}$ and $\{T\}$ exhibit the property $\{O\},\{T\} \subset D^{raw}$ and $|O||T| < M^{raw}$. The *tid* refers to trip *ID*, where each refers to a pair of origin and destination. The D^{proc} consists of all the elements of D^{raw}, which is obtained after filling, but structured based on the origin and destinations of the trips.

Remark 1: The number of elements in D^{proc} is equal to the number of elements in filled D^{raw}, i.e., $|D^{proc}| = |D^{raw}|$

Algorithm 1: Splitting filled D^{raw}

Input: D^{raw}

Output: D^{proc}

for every *s* in $\{O\}$

for every *t* in $\{T\}$

set *tid* to unity

for every *t* in

if (*s*, *t*) equals $(d_{i,1}, d_{i,2})$

$d_{tid}^{proc} \leftarrow d_i^{raw}$

Increment *tid* by 1

end if

end for

end for

end for

Figure 2. Pseudo code for splitting algorithm

4.2 Extraction of Relevant Attributes

In order to extract the relevant attributes, we determine the correlation of every attribute with other attributes. Based on the positivity of the correlation, the attributes are defined as relevant and significant for the particular trip. The algorithm for extracting the relevant attributes is illustrated in Figure 3.

Definition 1: Relevance of an attribute is defined here as the attribute that is highly correlating with the other attributes of the trip. The algorithm first extracts only the attributes $att_{tid}(m)$ from $d_{tid}^{proc}(m,n)$ because $d_{tid}^{proc}(m,n)|_{n=0}$ and $d_{tid}^{proc}(m,n)|_{n=1}$ refers to the origin and destination of the trip, respectively. The correlation coefficient of $att_{tid}(C_{tid})$ is determined and it is subjected to diagonal elimination in which the diagonal elements are set as zero. Based on C_{tid}, the correlation index for every attribute, C_{tid}^{idx}, is determined as follows

$$C_{tid}^{idx}(l) = \frac{\sum_{m=1}^{M_{tid}^{proc}} C_{tid}(m,l)}{M_{tid}^{proc}} : 0 \le l \le N_{tid}^{proc} - 1 \qquad (1)$$

where, M_{tid}^{proc} is the number of trips under the trip ID *tid* and N_{tid}^{proc} is the number of attributes that are considered. The lth attribute participates in d_{tid}^{corr}, only if $C_{tid}^{idx}(l)$ is in the positive plane. Hence, the dataset with highly correlating (relevant) attributes, termed as $d_{tid}^{corr}(m,k) : d_{tid}^{corr} \subseteq d_{tid}^{proc}$, is constructed, where $0 \le m^{(tid)} \le M_{tid}^{corr} - 1$ and $0 \le k^{(tid)} \le N_{tid}^{corr} - 1$.

Algorithm2: Extracting relevant attributes

Input: D^{proc}

Output: D^{corr}

for every tid of d_{tid}^{proc}

$att_{tid}(m) \leftarrow d_{tid}^{proc}(m,n): \forall m, 2 \le n \le N_{tid}^{proc} - 1$

Calculate C_{tid} from att_{tid}

$diagonal(C) \leftarrow 0$

Calculate C_{tid}^{idx}

Set $k \leftarrow 1$

for every l in C_{tid}^{idx}

if $C_{tid}^{idx}(l)$ is positive

$d_{tid}^{corr}(k) \leftarrow d_{tid}^{proc}(l+2) \forall m$

Increment k by 1

end if

end for

end for

Figure 3. Pseudo code of algorithm to extract the relevant attributes

Algorithm 3: Mining frequent and feasible patterns

for every tid in d_{tid}^{corr}

Determine att_{tid}^{idx}

Set $r \leftarrow 1$

for every p in att_{tid}^{idx}

Calculate $p_{tid,p}^{freq}(q)$

Calculate $F\left(p^{freq}(q)\right)_{tid,p}$

$p_{tid,p}^{seq}\left(r, att_{tid}^{idx}(p)\right) \leftarrow p_{tid,p}^{freq}(q) \forall q$

$F_{tid,p}^{seq}\left(r, att_{tid}^{idx}(p)\right) \leftarrow F\left(p_{tid,p}^{freq}(n)\right) \forall q$

Increment r by 1

end for

end for

Figure 4. Pseudo code for mining frequent and feasible patterns

Remark 2: d_{tid}^{corr} is equal to d_{tid}^{proc}, when all the N_{tid}^{proc} attributes enable positive correlation with each other

Remark 3: d_{tid}^{corr} remains as multiples of number of attributes that enable positive correlation with other attributes, provided $M_{tid}^{proc} = M_{tid}^{corr}$

4.3 Mining Frequent and Feasible Patterns

As stated, this step first extracts frequent patterns followed by reconstructing the frequent patterns to determine feasible patterns. The step also includes a preliminary process to skip the trips that do not have any relevant attribute. The steps are described in the pseudo code illustrated in Figure 4. This process generates *L* length patterns from the available attributes of d_{tid}^{corr}. In this process, possible combinations of the available attributes att_{tid}^{idx} are determined.

Records	A1	A2	A3	A4
1	5	2	5	4
2	5	2	5	4
3	6	2	5	4
4	6	2	5	6
5	6	2	5	6
6	6	2	1	7
7	5	2	1	7
8	1	2	1	1
9	1	2	3	1
10	2	1	3	1

(a)

A1	A2
A1	A3
A1	A3
A2	A3
A2	A4
A3	A4

(b)

(A1, A2)		(A1, A3)		(A1, A4)		(A2, A3)		(A2, A4)		(A3, A4)	
p^{freq}	F	p^{freq}	F	p^{freq}	F	p^{freq}	F	p^{freq}	F	p^{freq}	F
(6, 2)	4	(6,4)	3	(5,4)	2	(2,5)	5	(2,4)	3	(5,4)	3
				(6,6)	2						
				(1,1)	2						

(c)

A1		A2		A3		A4	
p^{freq}	F	p^{freq}	F	p^{freq}	F	p^{freq}	F
6	4	2	4	0	0	0	0
6	3	0	0	4	3	0	0
5	2	0	0	0	0	4	2
6	2	0	0	0	0	6	2
1	2	0	0	0	0	1	2
0	0	2	5	5	5	0	0
0	0	2	3	0	0	4	3
0	0	0	0	5	3	4	3

(d)

A1		A2		A3		A4	
p^{freq}	F	p^{freq}	F	p^{freq}	F	p^{freq}	F
6	4	2	5	5	5	4	3

(e)

A1	A2	A3	A4
p^{freq}	p^{freq}	p^{freq}	p^{freq}
6	2	5	4

(f)

Figure 5. An example to illustrate the algorithm for mining frequent and feasible patterns

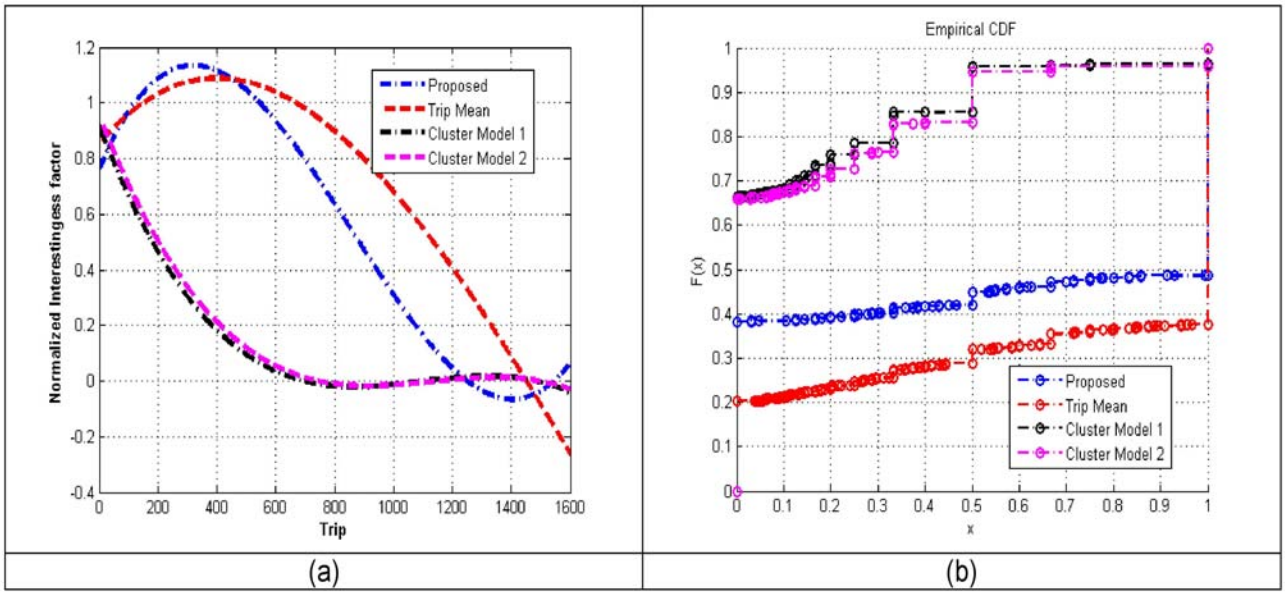

Figure 6. Performance comparison using (a) normalized interestingness factor over every trip plan and (b) its distribution for zero tolerance

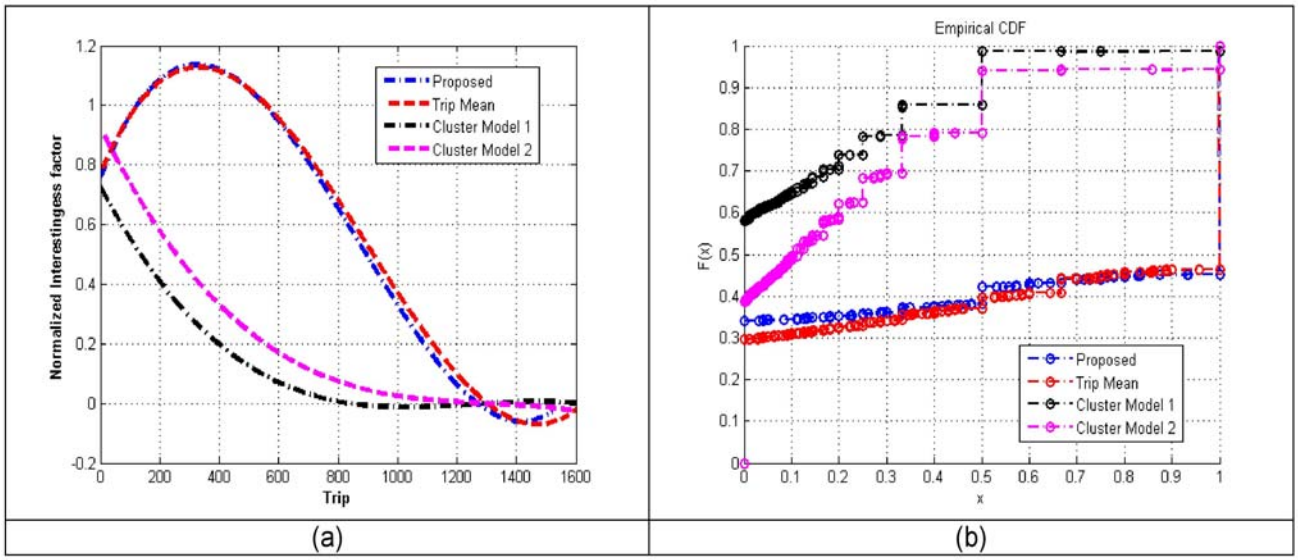

Figure 7. Performance comparison using (a) normalized interestingness factor over every trip plan and (b) its distribution for 10% tolerance

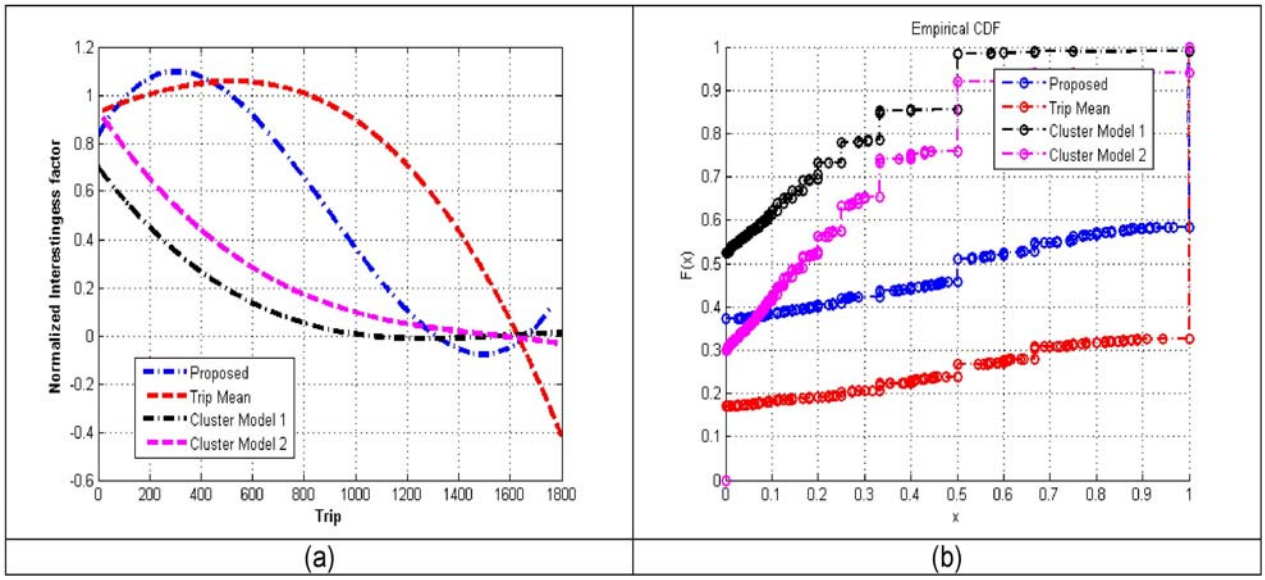

Figure 8. Performance comparison using (a) normalized interestingness factor over every trip plan and (b) its distribution for 20% tolerance

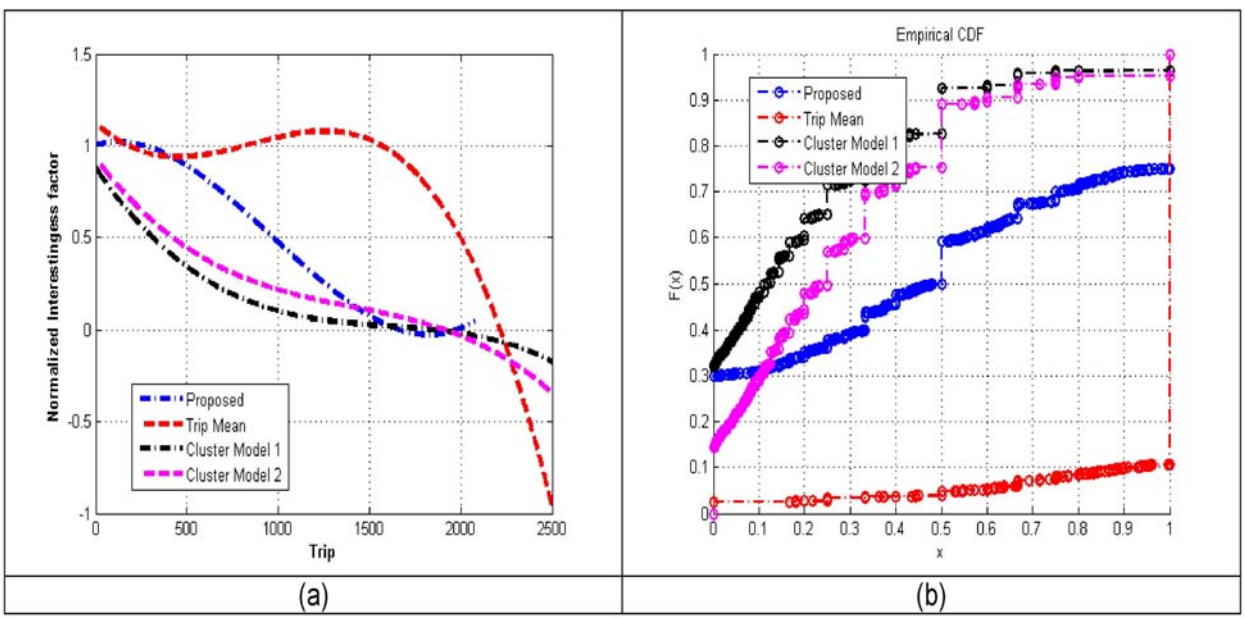

Figure 9. Performance comparison using (a) normalized interestingness factor over every trip plan and (b) its distribution for 50% tolerance

In Figure 4, $p_{tid,p}^{freq}(q)$ refers to frequent pattern elements and $F\left(p^{freq}(q)\right)_{tid,p}$ refers to its frequency. Example 1 interprets the pseudo code with sample data.

Example 1: Consider $d_{tid}^{corr} \mid tid = 1$, where $M_{tid}^{corr} = 10$ and $N_{tid}^{corr} = 4$ for which the data samples are presented in Figure 5 (a). Let $L = 2$ and hence six possible combinations can be generated as in Figure 5 (b). For every combination, the two length patterns that exhibit maximum frequencies are presented in Figure 5 (c). The high frequent patterns are relocated in the respective attribute column along with its frequency in Figure 5 (d). In every attribute, high frequent pattern element is retained and the other pattern elements are eliminated Figure 5 (e) and presented in Figure 5 (f).

Batch trip		1	2	3	4	5	6	7	8	9	10	Cumulative Mean	Cumulative Rank
Best case scenario	Proposed	104	102	82	67	90	73	82	71	76	70	81.7	2
	Trip mean	83	98	100	92	91	90	107	97	111	125	99.4	1
	Cluster model 1	2	4	5	7	6	5	4	8	8	10	5.9	4
	Cluster model 2	6	12	8	2	4	13	7	8	4	3	6.7	3
Worst case scenario	Proposed	47	52	59	62	55	57	68	65	71	74	61	2
	Trip mean	53	38	34	26	33	36	31	29	19	23	32.2	1
	Cluster model 1	128	117	105	83	95	102	126	108	96	98	105.8	4
	Cluster model 2	132	104	95	101	108	93	113	108	90	105	104.9	3
Mean score	Proposed	0.67	0.65	0.57	0.50	0.61	0.56	0.54	0.52	0.51	0.49	0.56	2
	Trip mean	0.56	0.67	0.70	0.73	0.67	0.66	0.72	0.71	0.78	0.79	0.70	1
	Cluster model 1	0.08	0.12	0.12	0.15	0.15	0.12	0.09	0.13	0.17	0.17	0.13	4
	Cluster model 2	0.11	0.18	0.14	0.12	0.12	0.18	0.13	0.14	0.17	0.15	0.14	3
Median score	Proposed	1	1	1	0.3333	1	1	0.8571	0.6667	0.5	0.5	0.78	2
	Trip mean	0.6667	1	1	1	1	1	1	1	1	1	0.96	1
	Cluster model 1	0	0	0	0	0	0	0	0	0	0	0	3
	Cluster model 2	0	0	0	0	0	0	0	0	0	0	0	3
Standard deviation	Proposed	0.44	0.45	0.47	0.48	0.47	0.47	0.47	0.47	0.47	0.47	0.47	4
	Trip mean	0.45	0.42	0.42	0.40	0.42	0.44	0.41	0.41	0.36	0.37	0.41	3
	Cluster model 1	0.18	0.22	0.22	0.26	0.25	0.24	0.20	0.25	0.26	0.27	0.23	1
	Cluster model 2	0.23	0.29	0.25	0.21	0.22	0.31	0.25	0.26	0.24	0.2	0.25	2

Table 2. Statistical Report On Trip Planning Models For Zero Tolerance

5. Results and Discussion

The proposed trip planning model is simulated along with three personalized trip planned models to demonstrate the performance of the proposed model. Despite few personalized models are reported in the literature [15], we consider trip mean model and clustering model, because of its personalization effect to handle our data.

The trip mean model takes the mean of every attribute of each trip and recommends it. Since it takes global average of every attribute, it is believed that it reflects highly frequent attributes. This can also be said as a simple statistical model. In the clustering model, every trip data is clustered into two and their centroids are determined. Here, the clustering is performed using agglomerative hierarchical clustering [16]. First centroid refers to the recommendation from cluster model 1 and the second refers to the recommendation from cluster model 2. The existing models recommend non-integer plans and hence a tolerance value is adjusted. Based on the adjustments of tolerance, the experimentation is categorized into two test cases.

5.1 Test Case 1 (Zero Tolerance)

Batch Trip		1	2	3	4	5	6	7	8	9	10	Cumulative Mean	Cumulative Rank
Best case scenario	Proposed	111	104	83	73	92	69	86	73	73	67	83.1	1
	Trip mean	77	79	84	81	67	74	86	72	92	102	81.4	2
	Cluster model 1	5	1	2	1	2	0	1	2	4	1	1.9	4
	Cluster model 2	12	7	11	6	8	6	7	11	13	5	8.6	3
Worst case scenario	Proposed	45	40	52	53	48	45	57	49	67	62	51.8	2
	Trip mean	68	42	48	42	50	40	41	45	40	35	45.1	1
	Cluster model 1	109	93	86	84	83	82	101	87	76	79	88	4
	Cluster model 2	76	59	54	51	55	60	73	62	46	51	58.7	3
Mean score	Proposed	0.68	0.70	0.59	0.56	0.65	0.59	0.59	0.59	0.51	0.52	0.60	2
	Trip mean	0.51	0.62	0.62	0.62	0.55	0.62	0.64	0.60	0.66	0.71	0.61	1
	Cluster model 1	0.12	0.13	0.13	0.12	0.14	0.13	0.10	0.12	0.18	0.15	0.13	4
	Cluster model 2	0.20	0.20	0.23	0.21	0.21	0.20	0.17	0.21	0.28	0.20	0.21	3
Median score	Proposed	1	1	1	1	1	1	1	1	0.5	0.5	0.9	1
	Trip mean	0.5	0.8	1	1	0.7	1	1	1	1	1	0.9	1
	Cluster model 1	0	0	0	0	0	0	0	0	0.0	0	0.0	4
	Cluster model 2	0.0	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.2	0.1	0.1	3
Standard deviation	Proposed	0.45	0.43	0.46	0.48	0.46	0.46	0.47	0.47	0.48	0.47	0.46	4
	Trip mean	0.47	0.43	0.45	0.46	0.45	0.46	0.44	0.46	0.44	0.43	0.45	3
	Cluster model 1	0.22	0.19	0.20	0.20	0.21	0.20	0.17	0.20	0.23	0.20	0.20	1
	Cluster model 2	0.28	0.25	0.27	0.25	0.27	0.26	0.25	0.29	0.29	0.24	0.27	2

Table 3. Statistical Report on Trip Planning Models For 10% Tolerance

In case 1, there is no tolerance considered for trip plans. Here, the plans from trip mean model and the cluster models are rounded to the nearest integer and the frequency of its occurrence is identified. Similarly, the frequency of occurrence of the plans recommended by proposed model is also observed. Based on the results, a normalized interestingness factor is calculated as follows

$$Interestingnes\,(mdl) = \frac{F(mdl)}{\max(F)\forall mdl} \tag{2}$$

where, *mdl* indicates the planning model and *F* (*mdl*) represents the frequency of occurrence of the trip plan proposed by mdl in the dataset. Further, the distribution of the interestingness factor is determined using cumulative distribution function (*CDF*). Hence obtained interestingness factor and the distribution metrics for all the trips are plotted in Figure 6.

5.2 Test Case 2 (With Tolerance)

In this case, a percentage tolerance is defined between the plan and the data elements. For instance, let the trip data be[12], where 1 refers to travel by bus and 2 refers to medium travel cost. However, assume the trip plan as $[1.1 \quad 2.3]$ for which the deviation is defined as $MEAN\left(\frac{|1.1-1|}{1}, \frac{|2.3-2|}{2}\right) = 12.50\%$. Here, a tolerance of 10%, 20% and 50% are applied so that the trip plan is considered under

Batch trip		1	2	3	4	5	6	7	8	9	10	Cumulative Mean	Cumulative Rank
Best case scenario	Proposed	88	90	72	71	80	70	71	58	63	70	73.3	2
	Trip mean	117	116	123	118	112	103	125	118	133	121	118.6	1
	Cluster model 1	3	1	1	1	0	3	1	2	6	0	1.8	4
	Cluster model 2	7	8	13	12	9	10	5	17	13	10	10.4	3
Worst case scenario	Proposed	52	50	64	69	65	55	71	74	86	71	65.7	2
	Trip mean	39	30	29	26	34	28	22	29	26	37	30	1
	Cluster model 1	100	89	88	90	96	85	104	101	81	90	92.4	4
	Cluster model 2	56	44	43	54	54	56	65	63	43	50	52.8	3
Mean score	Proposed	0.60	0.61	0.51	0.50	0.53	0.53	0.50	0.45	0.44	0.49	0.52	2
	Trip mean	0.70	0.75	0.76	0.76	0.71	0.74	0.78	0.76	0.78	0.73	0.75	1
	Cluster model 1	0.13	0.15	0.14	0.14	0.12	0.15	0.11	0.11	0.19	0.14	0.14	4
	Cluster model 2	0.22	0.25	0.27	0.25	0.23	0.23	0.18	0.25	0.28	0.24	0.24	3
Median score	Proposed	0.8	1	0.5	0.48	0.67	0.5	0.5	0.33	0.26	0.5	0.55	2
	Trip mean	1	1	1	1	1	1	1	1	1	1	1	1
	Cluster model 1	0	0.02	0.00	0	0	0	0	0	0.08	0.01	0.01	4
	Cluster model 2	0.17	0.2	0.17	0.18	0.14	0.12	0.1	0.14	0.25	0.18	0.17	3
Standard deviation	Proposed	0.44	0.44	0.45	0.46	0.46	0.45	0.46	0.45	0.46	0.46	0.45	4
	Trip mean	0.43	0.39	0.39	0.38	0.42	0.40	0.37	0.40	0.38	0.41	0.40	3
	Cluster model 1	0.21	0.20	0.19	0.20	0.17	0.23	0.18	0.18	0.25	0.18	0.20	1
	Cluster model 2	0.25	0.25	0.29	0.28	0.26	0.28	0.23	0.31	0.28	0.26	0.27	2

Table 4. Statistical Report on Trip Planning Models For 20% Tolerance

20% and 50% tolerance (since 12.50% > 20% and 50%), whereas it is not considered under 10% tolerance. Hence the entire trip plans are investigated under these three tolerances and the results are plotted in Figure 7, 8 and 9, respectively.

5.3 Statistical Analysis

The statistical analysis includes five first order statistical metrics such as best case performance, worst case performance, mean performance, median performance and standard deviation. The best case performance is defined as the number of normalized interesting factor being 1 throughout the entire trips, whereas the worst case performance refers to the normalized interesting factor being 0. Hence obtained values are tabulated in Table 2 – 5.

5.4 Discussion

In all the performance illustrations, i.e. from Figure 6 to Figure 9, the trip mean model remains in the first position, whereas the proposed model grabs the second position in terms of normalized interestingness factor. The two cluster models take third positions, respectively. This demonstrates that the socio-economic constraint defined by trip mean model is highly interested based on the user history, while, the plan of the proposed model are second likely interested. However, the CDF illustrations of each case shows the trip mean model and proposed model grab fourth and third position, respectively, whereas the cluster models take first two positions. According to this, outcomes obtained from the trip mean model are loosely distributed. In other words, the dominating performance is not achieved throughout the trips. This can also be seen from the interestingness plot that the trip mean model has reached negative score at the final stage of trips. Despite the distribution is high in the cluster models, it remain in final positions. In other words, their plans are less likely interested throughout the trips, whereas the plans of trip mean model are either highly interested or not at all interested. Under these circumstances, the proposed model maintains a good trade-off between these two metrics by accomplishing second position in the interestingness factor, whereas the distribution almost reaches 0.5.

Batch trip		1	2	3	4	5	6	7	8	9	10	Cumulative Mean	Cumulative Rank
Best case scenario	Proposed	67	65	47	47	60	50	51	33	51	47	51.8	2
	Trip mean	177	188	192	188	175	176	185	196	193	186	185.6	1
	Cluster model 1	7	4	9	2	10	11	7	8	8	9	7.5	4
	Cluster model 2	6	9	9	4	8	17	14	10	8	16	10.1	3
Worst case scenario	Proposed	38	39	61	77	67	59	75	76	71	60	62.3	2
	Trip mean	9	8	7	1	8	7	5	4	2	0	5.1	1
	Cluster model 1	76	62	60	71	63	65	89	69	53	60	66.8	4
	Cluster model 2	39	23	37	22	34	29	31	38	23	24	30	3
Mean score	Proposed	0.58	0.57	0.47	0.43	0.46	0.46	0.43	0.38	0.43	0.47	0.47	2
	Trip mean	0.92	0.94	0.95	0.96	0.93	0.94	0.94	0.96	0.96	0.97	0.95	1
	Cluster model 1	0.18	0.21	0.22	0.17	0.20	0.23	0.16	0.19	0.24	0.21	0.20	4
	Cluster model 2	0.24	0.30	0.27	0.27	0.27	0.30	0.28	0.28	0.30	0.31	0.28	3
Median score	Proposed	0.6	0.6	0.5	0.4	0.4	0.5	0.4	0.3	0.4	0.5	0.5	2
	Trip mean	1	1	1	1	1	1	1	1	1	1	1	1
	Cluster model 1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	4
	Cluster model 2	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.3	0.3	0.3	0.2	3
Standard deviation	Proposed	0.38	0.37	0.39	0.40	0.42	0.39	0.41	0.37	0.39	0.39	0.39	4
	Trip mean	0.24	0.21	0.19	0.14	0.22	0.21	0.20	0.17	0.14	0.14	0.19	1
	Cluster model 1	0.24	0.23	0.25	0.20	0.25	0.28	0.23	0.24	0.26	0.25	0.24	2
	Cluster model 2	0.24	0.24	0.25	0.23	0.25	0.29	0.27	0.26	0.24	0.28	0.25	3

Table 5. Statistical Report on Trip Planning Models For 50% Tolerance

The similar results have also been observed from the statistical report tabulated in Table 2 – 5. For instance (refer Table II – best case scenario), the proposed model produces 81.7 (on average) highly interesting patterns (means, the pattern which has normalized interestingness score as 1) for every trip, whereas the clustering model produces only 5.9 and 6.7 patterns, respectively. Similarly (in the worst case scenario), the proposed model produces 61 non-interested patterns, whereas the cluster models produce around 105 non-interested patterns. Despite the trip mean model is dominating under both the scenarios (99.4 interested patterns and only 32.2 non-interested patterns), the distribution over the 50% interested patterns are worst (as per Figure 6 (b), 7(b), 8 (b) and 9 (b)).

The mean scores and median scores are highly different from the values produces in the best case scenario and the worst case scenario. For instance, the value 104 (Table I – best case scenario by proposed model) refers to the number of plans that achieved the normalized interestingness factor 1. The value 0.67 (Table I – mean by proposed model) refers the average interestingness factor accomplished by all the plans proposed by our model. The performance of the proposed model over the clustering models in the statistical report and the trip mean model in the distribution performance demonstrates that the proposed model maintains a good trade-off between these two metrics, which are essential to ensure reliable and consistent performance.

6. Conclusion and Future Scope

This paper introduced a new trip planning model using data mining approaches. Real-time travel information has been acquired

from Hyderabad city of India and the experimentation has been carried out to demonstrate the performance of the proposed planning model. The proposed planning model was able to produce the socio-economic constraints which are highly relevantto the trip, rather than its frequency. Three level of performance investigation has revealed that the proposed model has maintained adeq uate trade-off between all these performance metrics. The obtained results are encouraging and hence the performance will be substantially improved than the trip mean model, since trip mean model dominates in the statistical analysis. Moreover, the tolerance analysis will be extended along with varying pattern lengths.

References

[1] Figueiredo, L., Jesus, I., Machado, J.A.T., Ferreira, J. R., de Carvalho, J. L. M. (2001). Towards the development of intelligent transportation systems. *In: Proceedings of the IEEE Conference on Intelligent Transportation Systems, p* 1206-1211, 2001.

[2] Na Zeng., Ke-jing Qin., Jun Li. (2010).Intelligent transport management system for urban traffic hubs based on an integration of multiple technologies. *In: Proceedings of the IEEE 17th International Conference on Industrial Engineering and Engineering Management (IE&EM), p* 1178-1183, October 2010.

[3] Masaki., I. (1998) .*Machine-Vision Systems for intelligent transportation systems. Intelligent Systems and their Applications* 13 (6), 24-31.

[4] Xinping Yan., Hui Zhang., Chaozhong Wu. (2012). Research and Development of Intelligent Transportation Systems. *In: Proc. of the* 11th *International Symposium on Distributed Computing and Applications to Business, Engineering & Science (DCABES, 2012), p* 321-327, October 2012.

[5] Jing Dong., Mahmassani., H.S. (2011). Stochastic modeling of traffic flow breakdown phenomenon: Application to predicting travel time reliability. *In: Proceedings. of the* 14th *International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 2112- 2117, Oct. 2011

[6] Yu Yao. (2014). The track research on the latest developments of Intelligent Transportation Systems abroad. *In: Proc. of the* 11th *World Congress on Intelligent Control and Automation (WCICA)*, pages 5132 – 5137, Jun./Jul. 2014.

[7] Robert, B., Noland, Polak, J.W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transport Reviews*, 122 (1) 39- 54.

[8] Al-Deek, H., Emam, E. B. (2006). New methodology for estimating reliability in transportation networks with degraded link capacities. *Journal of Intelligent Transportation Systems*, 10 (3), 117-129.

[9] Korhonen, T., Vaaramaki, T., Riihimaki, V., Salminen, R., Karila, A. (2012). Selecting telecommunications technologies for intelligent transport system services in helsinki municipality. *IET Intelligent Transport Systems*, 6 (1). 18-28.

[10] Di Lecce, V., Amato, A. (2011). Route planning and user interface for an advanced intelligent transport system. *IET Intelligent Transport Systems*, 5 (3), 149-158.

[11] Wang Peng., Wang Jiang-Ping., Xia Jing. (2009). The application of particle swarm optimization on intelligent transport system. *In: Proc. of the International Colloquium on Computing, Communication, Control, and Management*, 4, 389 – 391.

[12] Chiew, K., Shaowen Qin. (2009). Scheduling and Routing of AMOs in an Intelligent Transport System. *IEEE Transactions on Intelligent Transportation Systems*, 10 (3), 547 – 552.

[13] Juan Zhicai, Wu Jianping, Mike McDonald. (2006). Socio-Economic Impact Assessment of Intelligent Transport Systems. *Tsinghua Science and Technology*, 11(3), 339-350.

[14] Junping Zhang, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, Cheng Chen. (2011). Data-Driven Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 12 (4), 1624-1639.

[15] Lathia, N., Froehlich, J., Capra, L. (2010). Mining Public Transport Usage for Personalised Intelligent Transport Systems. *In: Proc. of the IEEE 10th International Conference on Data Mining (ICDM), p* 887-892, December.

[16] Rokach., Lior., Oded Maimon. (2005). Clustering methods. *Data mining and knowledge discovery handbook, Springer US* 321-352.