Speculating the Travel Purposes of Passenger Groups in Railway Transportation System for Better understanding of Passengers

Shirin A. Maniyar
M.S.Bidwe Engineering College
SRTM University, Nanded.

Pooja K. Shinde
M.S.Bidwe Engineering College
SRTM University, Nanded.

ABSTRACT

There are different areas for speculations from the researcher's point of view. Speculations about persons who are traveling together is a very interesting area for research. This presumption can benefit to many bearing services so that they can improve personalized carrier services and get more benefit from that presumptions. This will results a drastic change in bearing services. To get this benefit we have implemented a system, in which we have to perform iterative classification algorithm on data, which results presumption of passengers. We have to build Co-travel network first then we have to generate features to form a grouping of passengers then use iterative classification algorithm. Iterative classification algorithm collectively finds the overlapping of passengers from different groups and results speculation of passengers who are traveling together from same source to destination. Experimental results on data set of passenger travel record in the field of railway transportation system demonstrate that our proposed iterative classification approach can efficiently speculate the travel purposes of passenger group.

General Terms

Travel purpose, transportation system, co-travel network.

Keywords

Speculate, co-travel network, iterative classification, passenger group, travel purpose, collective inference.

1. INTRODUCTION

As we know life is one time fruitful opportunity so generally peoples are living their life with full of happiness. Every one want to live life happy, enjoy every moment of life. Some people get satisfied by traveling to different places, visiting to relatives, some of them find happiness by spending moments with friend,chatting with friends, watching movies. Every person in this world have their own way of satisfaction. Generally most of the peoples travel for enjoying moments in life with families, friend, coleagues, dear ones sometimes alone also. Every one has their comfortable satisfaction zone. Human travel usually together with other peoples from same source to same destination and they utlistize every seconds of life which results to their fruitful happiness factor and this could be highest rated area for enjoyment refered by today's generation peoples.

Spectulations about traveling of peoples with other people together from same soruce to same destination can be very interesting area for resercher's today, because this can be very important factor to make meaningful changes in personalised travel transportations and decision making of traveling agencies,organisations even governments too.Presumptions

about behaviours of persons who are traveling in large groups [1], [2] are different as compaire to indivizuals traveling behaviour [3]. In general survey represents peoples enjoy their journey in groups based traveling and remember that moments too for long time.

Every persons traveling agenda is different, it might be for business,tourisum,education or any other [4], [5]. For example japenese peoples travel repetedly to U.S. [6]. Japenese travelers have been one of the the largest groups of traveler to U.S. So they normally called japanese plesure travellers to U.S. Whenever they travel to U.S. They arrange gettogether for relative's. Their relatives meet them and gives them their requirements and gifts as a wish at the meeting times and japanese return new gifts to their relatives while they are return from their journey's. So basically here the japanese traveller's agenda is to travel U.S. is to spend pleasure time there and purchase return gifts to their relatives. In this way every person have their own purposes [7], [8].

Spectulation about traveling purposes of passenger groups can be beneficial for any transportation system might be railway transportation,civil aviation or bus tranportation system. With this research any tranportation system improve their service providing patterns. If the persons are traveling for business purposes repetedly then service provider can arrange meeting places like meeting halls for smooth conduction of meeting's. They can also provide basic needs for business groups as an package so that business groups can't weaste their important busniess time. For tourist groups,they can recommend different routes so that tourist can spend their pleasure time with different routes and know beauti of nature. Hence this idea is somewhat beneficial for government also. With this idea government can improve econimical strategy.

Railway transportation system can adjust their trains according to passenger's purposes. They can also leave extra train's on seasonal times. They can adjust trains according to situation. Air transportion can adjust their flights, bus transportation system can also change their schedule according to season. Any transportation system can manage their policies according to passengers or current situation. It will be better for transportation system to take decision instantly without any compulsion, but restricted to take action as per situation so the authorities will always take a decision which is beneficial to that respective system because the authorities are only responcible for taking that decisions so they will take decision very carefully as expected in the advantage of that system and serves benefit to nation also.

For implementing this application we should have the travel purposes of passengers and that should be single and clear in general. For sucessful conduction of this application we should know the passengers historical travel record as well as basic information. The basic information of passengers will be

taken at the time of reservation. Every transportation system keeps the basic information of passengers in their passenger name record i.e. PNR. To form this application we have to find the social relationships between passengers that will be recognize from the passenger name record to reperesent the social relationship between passengers the Co-travel network [9] is used.
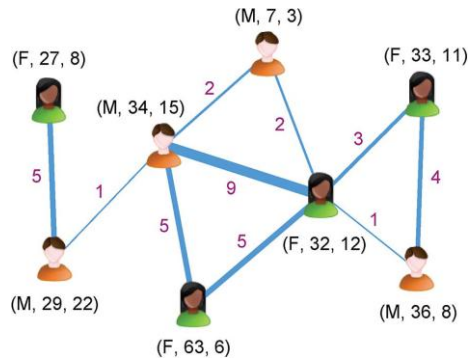

Figure 1 : Example of a civil aviation passenger group [9].

Figure 1 shows the co-travel network of civil aviation passenger group. The letter and numbers in a bracket in Figure 1 indicate the gender, age, and historical travel times of a passenger, whereas the number beside an edge indicates the historical co-travel times between two passengers (it is noted that the current co-travel is not counted in). Then, the problem is how to infer the travel purpose of this passenger group.

Spectulating travel purposes of passengers is a classification problem. To solve this problem we have to apply traditional clssifier. The traditional clssifiers work on the concept of IID (independent and identical distribution) [10]. IID states any of two passengers are independent of each other. Considering this fact results to loss of dependancy factor which directly ignores the dependancy between passengers. For example any of two passengers can be overlap in between other groups this fact is ignored in traditional classifiers. For this problem the solution is to spectulate collectively all passenger groups so that no passenger will lose from any other groups. Which results to recognize overlapping of passengers in many groups so no one will lose to spectulate from all groups [9]. If repetedly spectulating the passengers from all groups will results to clear travel purposes of passengers. Ierative classification algorithms [9] results the overlapping of passengers sucessfully.

The traditional classifiers neglect the fact of collective inference. So the traditinal clssifiers leads to inaccuracy. Collective inference provides better results as compaire to traditional classifiers. Collective inference uses auto correlation dependance relation [11], [12] between the variables of related entities which improves prediction.

In this application also the collective inference concept is used as base concept. The PNR records of the railway transportation are extracted from the systems, based on that records the co-travel network is constructed. The co-travel network represents the social relation between the passengers. The features are extracted for forming the groups of passengers & assigning labels to all passengers. The iterative classification algorithm is used to successfully find out the overlapping of passengers [13], [14], [15]. Finally with the help of collective classification inference all labels are detected successfully.

The rest of the paper contains literature survey which explains the necessary methods, algorithms respective to work.

2. LITERATURE SURVEY

Many researchers have done the work with this application and have different views regarding to this. Jang et al. [16] recognizes the travel expenditure and travel behavior of pleasure travelers of U.S. They have concluded this research with some important point like Japanese pleasure traveler repeatedly visits to U.S. They are mostly educated & spend heavily on lodging & gifts to close relatives and friends and spend less on foods. So that they are normally called as "white collars". Older travelers spend more as compare to new travelers.

Jenson et al. [17] describes collective inference reduces classification error as compare to traditional classifiers. Collective inference observed characteristics of relational data. They also have given reasons why collective inference has to use. They also have described the cleared concept about constraint on Independent and identical distribution and explained with different models like intrinsic model, relational R_1, R2 models. Finally they have describes how collective inference can be used on relational data by relational collective inference.

Xiang et al. [18] propose a novel latent relational model based on "copulas" which allows use to make predictions in a discrete label space at the time of ensuring identical marginal. They describes copulas are used for descriptive modeling. In this research they have compare existing collective inference approaches with copula network.

Neville et al. [19] describes a research in which they propose an iterative classification algorithm. They have used simple Bayesian classifier in iterative fashion and dynamically updating the attributes of some object and guesses are made about its related objects. They have proposed generalized iterative classification algorithm. They also have mentioned static relational attributes and dynamic relational attributes with the help of one manufacturing company.

N. Eagle et al. [20] presumed friendship network by capturing mobile phone data. They have detected social relationship between people by collecting data by media devices. They have some views depend on that finally infer the friendship network.

You fang Lin et al. [21] have done very important research that states infers high quality co-travel network. They have given very different ways through which we can build high quality co –travel network For implement this application they have covered all minors things like co-flight times, co-flight costs, large passenger group, small passenger group, co-journey times with this tremendous work they have given strong reasons how to build high quality network.

Wan et al. [22] have research on the real data of civil aviation IT provider in china. They have discovered family groups from social networks. They also have done this research with the help of collective inference and iterative classification algorithm. They have also used new concept to discover family group i.e. Conditional Random Field. They have made their decision for detection of family group members from social network on use of weighted community detection algorithm. Weighted community detection algorithm is superior community detection algorithm. There are various general community detection algorithms. BGLL [23] and Info map [24] are very popular general community algorithms. These researchers have used these general community

algorithms to discover family groups from social networks.

3. PROBLEM DEFINITION

Here we explain some necessary definitions and formulas.

Definition 1—Passenger Group: A passenger group g = (Pg , orig , dstg , disg , d_dateg , r_dateg) is a collection of passengers who book tickets together for a certain travel, where Pg = {pi} is a set of passengers, orig and dstg are the common origin and destination of the travel, disg is the mileage of the travel, and d_dateg and r_dateg are the departure date and return date of the travel, respectively. Obviously, we have |Pg | ≥ 1, where |Pg | = 1 means a passenger travels alone [9].

In real world, passenger service providers keep the data of all their passenger groups in their information systems as record. Let's consider a large set of passenger groups S = {g1, g2. . . gn } for certain period of time and our goal is to speculate the travel purpose of each passenger group in the set, from all kinds of information mentioned in Definition 1. formally, we can define the problem as follows.

Problem 1—Speculate the Travel Purposes of Passenger Groups in Railway transportation system: Given a travel record set S = {g1, g2. . . gn}, the objective is to learn a function such as [9].

$$f: S \rightarrow L$$

Where L is the label space of the problem, i.e., all the possible types (travel purposes) of passenger groups. Our goal in this paper is to classify each passenger group into one of the three categories: tourist groups and business groups and education groups. In this case, the problem can be treated as a binary classification problem [9].

4. PROPOSED SYSTEM

In this section we present the solution for the problem of speculate travel purposes in railway transportation system in detail. At first we have to construct co travel network from historical travel records of passenger groups. Then several features of passenger groups, including all types of features (basic and network based) for classification. Finally we have to propose a novel iterative classification algorithm.

Table 1: List of notations [9]

Notation	Description
G	Passenger group
P_g	Member set of passenger group g
ori_g	Origin of passenger group g
dst_g	Destination of passenger group g
dis_g	Distance between ori_g and dst_g
d_date_g	Departure date of passenger group g
r_date_g	Return date of passenger group g
$\#male_g$	Number of male of passenger group g
P	Passenger
i,j,k	Index of passenger
age_i	Age of passenger p_i
$gender_i$	Gender of passenger p_i

$seat_i$	Seat number of passenger p_i
$mile_i$	Total historical travel mileage of passenger p_i
$\#travel_i$	Total historical travel times of passenger p_i
G=(V,E)	Co-travel network
$G=(V_g,E_g)$	Corresponding sub graph of passenger group g in G
d_i	Degree of passenger p_i
e_{ij}	Edge between passenger p_i and p_j in G
w_{ij}	Weight of edge e_{ij}
T_i	Set of neighbors of passenger p_i
$\#comp_g$	Number of components of sub graph G_g

Definition 2.Co-travel Networks: A Co-travel network is a graph G = (V, E), where V is a node set, and each node pi ∈ V represents a passenger; E is an edge set, and each edge eij ∈ E indicates that passengers pi and pj have traveled together at least once. Following sections explains the important steps need to complete this application. Table 1 represents the necessary notations.

4.1 Building a co-travel network

As mentioned previous, service providers maintain record of all the passengers historical behavior in their information systems, and actually each record is a passenger group. For example in the railway transportation system, a PNR contains the travel information of a group passengers who book the railway tickets might be together or alone. From such records, we can construct a co-travel network, which represents the social relationships between passengers. The co-travel network algorithm is as follows.

Algorithm 1: Constructing co-travel network [9].

Input: S={g₁,g₂,g₃....gₙ};

Output: G = {V,E};

  for each passenger group g ϵ S

    for each passenger pair (pᵢ,pⱼ) (pᵢ ,pⱼ ϵ Pg,pᵢ ≠ pⱼ)

      if pᵢ ∉ V **then**

        V← V ∪ {pᵢ};

      end if;

      if pⱼ ∉ V **then**

        V← V ∪ {pⱼ};

      end if;

      if eᵢⱼ ∉ E **then**

        Wᵢⱼ = 1;

        E← E ∪ {eij };

      else

        Wᵢⱼ = Wᵢⱼ + 1;

      end if;

    end for;

  end for;

4.2 Generating features

After construction of co-travel network, we have to generate a series of features for the classification of passenger groups. We divide the features into two different categories. First one is basic features and the second feature is network based features. The detailed information is as follows.

4.2.1 Basic features

Basic features are generated within each passenger group. These are very simple characteristics of passenger groups but without co-travel network. These features are again divided into three different types holistic features, demographic characteristics and historical travel statistics of group members.

4.2.1.1 Holistic group attributes

Passenger groups are bounded to a certain travel that's why these features are related to current travel. The holistic group attributes are very important basic feature to speculate the travel purpose of passengers. For example different types of passengers may have different scales of group sizes, travel duration, travel distances, whether the travel is on holiday, whether travel is on return trip and seat proximity is also important which reflect the closeness between the group members and also help us to identify the type of traveling passenger [9]. Table 2 shows list of holistic group attributes and formulas.

Table 2: Holistic group attributes [9].

Feature	Formulization
group size	$size_g = |P_g|$
travel distance	dis_g (i.e., mileage of the current travel)
travel duration	$dur_g = \begin{cases} 0, & \text{if } r\_date_g \text{ is null} \\ r\_date_g - d\_date_g, & \text{otherwise} \end{cases}$
whether holiday	$is\_holiday_g = \begin{cases} 0, & \text{if } d\_date_g \text{ is a workday} \\ 1, & \text{otherwise} \end{cases}$
whether return	$is\_returned_g = \begin{cases} 0, & \text{if } r\_date_g \text{ is null} \\ 1, & \text{otherwise} \end{cases}$
seat proximity	$seat\_proximity_g$ (ref. Algorithm 2.)

4.2.1.2 Demographic features

The second most important basic features of group members. These features are related to age distribution of passenger groups because different types of passengers have different types of age ratio. Table 3 shows all the demographic features of passenger groups [9].

Table 3: List of demographic features [9]

Feature	Formulization
average age	$avg\_age_g = \frac{\Sigma_{p_i \in P_g} age_i}{|P_g|}$
age variance	$age\_var_g = \sqrt{\frac{\Sigma_{p_i \in P_g}(age_i - avg\_age_g)^2}{|P_g|}}$
ratio of minors	$minor\_ratio_g = \frac{|\{p_i | age_i < 18, p_i \in P_g\}|}{|P_g|}$
ratio of seniors	$senior\_ratio_g = \frac{|\{p_i | age_i > 60, p_i \in P_g\}|}{|P_g|}$
ratio of adults	$adult\_ratio_g = \frac{|\{p_i | 18 \le age_i \le 60, p_i \in P_g\}|}{|P_g|}$
ratio of male	$male\_ratio_g = \frac{\#male_g}{|P_g|}$

4.2.1.3 Historical travel attributes of group

The third most important basic feature is historical feature. These types of features represent all historical record of passengers in history. These types of features gives very useful information like average travel times and average travel mileage. These features are listed in table 4.

Table 4: List of historical travel attributes of group [9].

Feature	Formulization
total mileage	$sum\_mile_g = \Sigma_{p_i \in P_g} mile_i$
average mileage	$avg\_mile_g = \frac{sum\_mile_g}{|P_g|}$
total travel times	$\#sum\_travel_g = \Sigma_{p_i \in P_g} \#travel_i$
average travel times	$\#avg\_travel_g = \frac{\#sum\_travel_g}{|P_g|}$

4.2.2 Network based features

The Network based features are drawn in the context of overall Co-travel network. Network based features represents network related components to Co-travel network. Network based feature are summarized in Table 5 with formulization.

Table 5: List of network based features [9].

Feature	Formulization
edge density	$density_g = \frac{|E_g|}{|V_g|(|V_g|-1)/2}$
average weight of edge	$avg\_weight_g = \frac{\Sigma_{e_{ij} \in E_g} w_{ij}}{|E_g|}$
ratio of components	$comp\_ratio_g = \frac{|V_g| - \#comp_g}{|V_g|-1}$
average clustering coefficient	$\#avg\_clus_g = \frac{\Sigma_{P_i \in V_g} \frac{|\{e_{jk}|p_j,p_k \in \Gamma_i, e_{jk} \in E_g\}|}{d_i(d_i-1)/2}}{|V_g|}$
average embeddedness	$avg\_emb_g = \frac{\Sigma_{P_i \in V_g}(\frac{1}{d_i}\Sigma_{P_j \in \Gamma_i} \frac{|\Gamma_i \cap \Gamma_j|}{|\Gamma_i \cup \Gamma_j|})}{|V_g|}$

4.3 Iterative classification algorithm

For classification of passengers in specified group, the features are extracted and then can apply traditional classifier but still one problem is still exist. The one important fact is still ignored. Overlapping of passengers still exist between passenger groups. It is the thing that make passenger dependent of each other, because the travel purposes is relevant in the social relationship. For example a passenger can be travel for many purposes and they can have different group relation with other passenger like a passenger can travel with his/her colleagues for business purpose while with her/his family or friends he can travel for visiting relatives or sightseeing or for educational purpose. So there may be huge chances of passenger to be overlap in many groups and that is ignored in traditional classifier or existing system.

To find Solution we have propose a novel Iterative Classification Algorithm (ICA) which more accurately speculate travel purposes of passenger group. Iterative classification [9], [13], [14], [25] describes that there are some statistical dependence relations rather than being completely independent between the purposes of target object, and that's the reason break ICA breaks the constraint of the IID assumption. ICA collectively speculates all of the purposes in the whole data set, which results to improve the accuracy of classification.
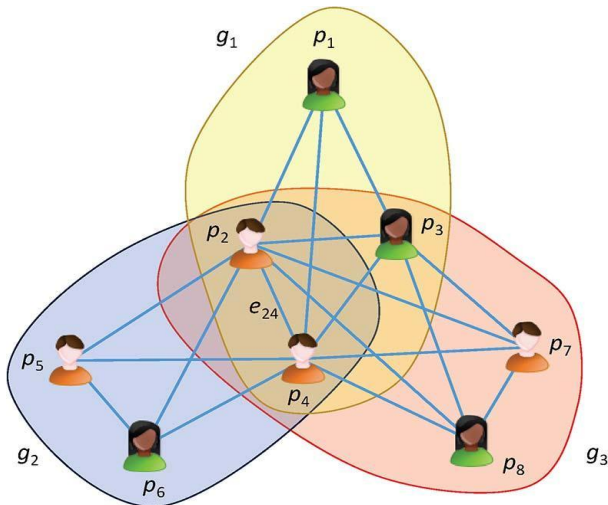

Figure 2: Simple example of overlapping passenger groups [9].

According to the analysis before, when the overlap between two groups contains only one passenger, it is hard to say that there are some correlations between them. Therefore, we take an overlapping edge as a basic unit to propagate the probabilistic dependence relations between passenger groups. For doing this we have calculate Label Preference of Edge [9] which is to accurately gives idea about overlapping between passengers. The iterative classification algorithm is as follows.

Algorithm 2: Iterative classification algorithm [9].

Input: S= {$g_1, g_2, g_3 \ldots g_n$}; G = {V,E};

     F:;

Output: P(Y) = {P (y_1), P(y_2),….. P(y_n) };

    for each edge e_{ij} ϵ G

        for each possible label value l ϵ L

            $\mathcal{P}^{(0)}$ (y_{ij}= l) = 1/|L|;

        end for;

    end for;

    repeat

      for each passenger group g ϵ S

        for each possible label value l ϵ L

$$\mathcal{P}^{(0)}(y_g = l) = \frac{\sum e_{ij} \epsilon E_g \, \mathcal{P}^{(t-1)}(y_{ij}=l)}{|E_{ij}|};$$

$$P^{(t)}(y_g = l) = f(F, \mathcal{P}^{(t)}(y_g = l));$$

        end for;

      end for;

      for each edge e_{ij} ϵ G

        for each possible label value l ϵ L

$$\mathcal{P}^{(t)}(y_{ij}=l) = \frac{\sum G_g : E_g \ni e_{ij} \mathcal{P}^{(t-1)}(y_g=l)}{w_{ij}};$$

        end for;

      end for;

    until $| P^{(t)}(y_g = l) - P^{(t-1)}(y_g = l)| < \xi$;

5. EXPERIMENTS

In this section we present our set of experiments to evaluate proposed system. We performed travel purpose speculation on own created railway transportation data set and mentioned the results that are tried to match with existing system.

5.1 Dataset

We created a passenger group data set for railway transportation system. The dataset contains the information of passengers (e.g. Passenger name, passenger ID, gender, age, mobile number, address, email id) and tickets (passenger ID , railway code,origin,destination,seat number, departure time, return date).It should be stress that personal and travel information we used in our experiments is private and must be seriously protected.

For training and testing our approach, the passenger groups in the data set should be labeled in advance. Fortunately, some

railway reservation tickets labeled with the tags of business visit, tourist, educational visit, visiting relatives. Finally we have obtained 70 passenger data set. The statistics of the data set we collected are listed in table 6.

Table 6: Statistics of the data set.

S.R.NO	Title	Quantity
1)	Number of PNRs(i.e. Passenger groups)	3
2)	Number of passengers	70
3)	Number of overlapping relations	4
4)	Number of labeled business passenger groups	20
5)	Number of labeled tourist passenger groups	25
6)	Number of labeled educational passenger groups	20

5.2 Experimental Results

In our experiments, we perform iterative classification algorithm on dataset. To evaluate contribution of different types of features and the performance of our proposed ICA in railway reservation system, we perform 1) ICA with demographic features for every group (i.e. Business, tourist, educational) 2) ICA with network based features for every group.3) ICA with both the network based features and demographic features. 4) ICA with historical features for every group. 5) ICA analysis for overlapping passengers 6) GroupWise passenger analysis. All the experiments are performed on 70 labeled passenger groups, 3 passenger groups, each group consist of at least 20 passengers. All the experiments performed on fully labeled groups as supervising information for training and test.

From the experiments we have observed that group based traveling usually do for tourisms purpose then business purpose after that for education purpose. Our proposed ICA approach significantly improves the accuracy of classification. Relations between passengers are very useful information for speculating their travel purposes, and the iterative classification approach we proposed is a very effective method to solve this problem. We plotted the graph (figure) for each features on own created railway reservation dataset and our ICA model successfully speculate the travel purposes of different passenger groups in railway transportation system.
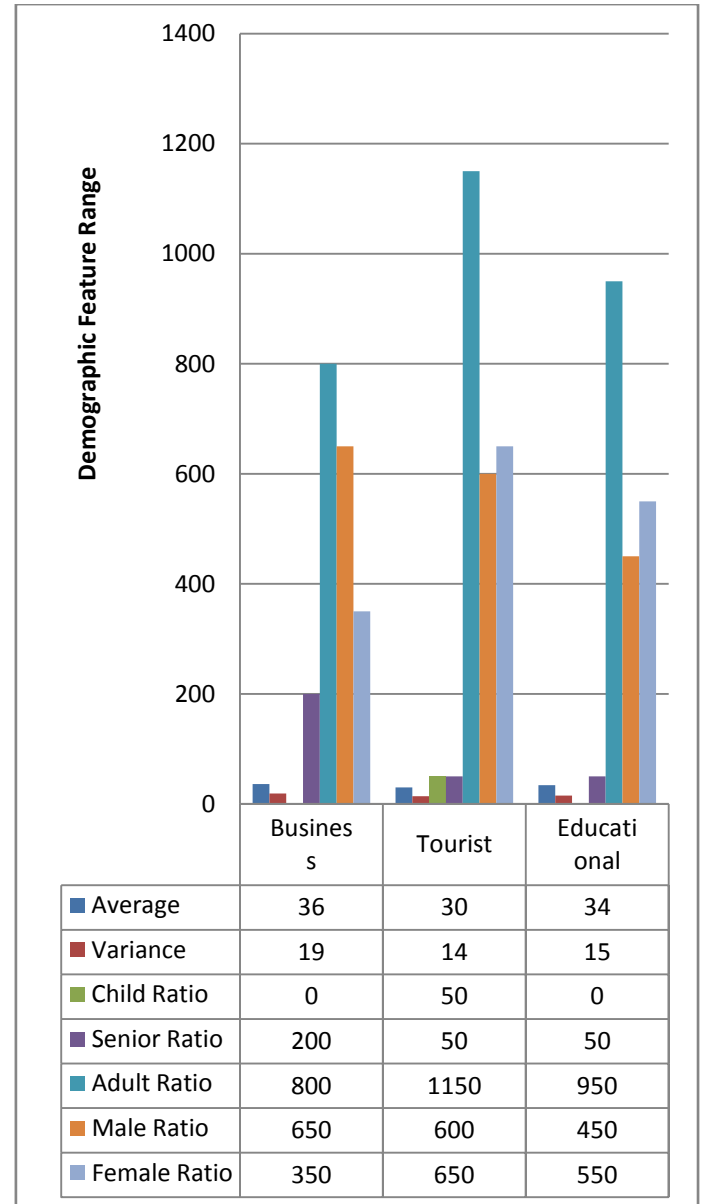

	Busines s	Tourist	Educati onal
■ Average	36	30	34
■ Variance	19	14	15
■ Child Ratio	0	50	0
■ Senior Ratio	200	50	50
■ Adult Ratio	800	1150	950
■ Male Ratio	650	600	450
■ Female Ratio	350	650	550

Figure 3: Demographic feature analysis graph for all groups.

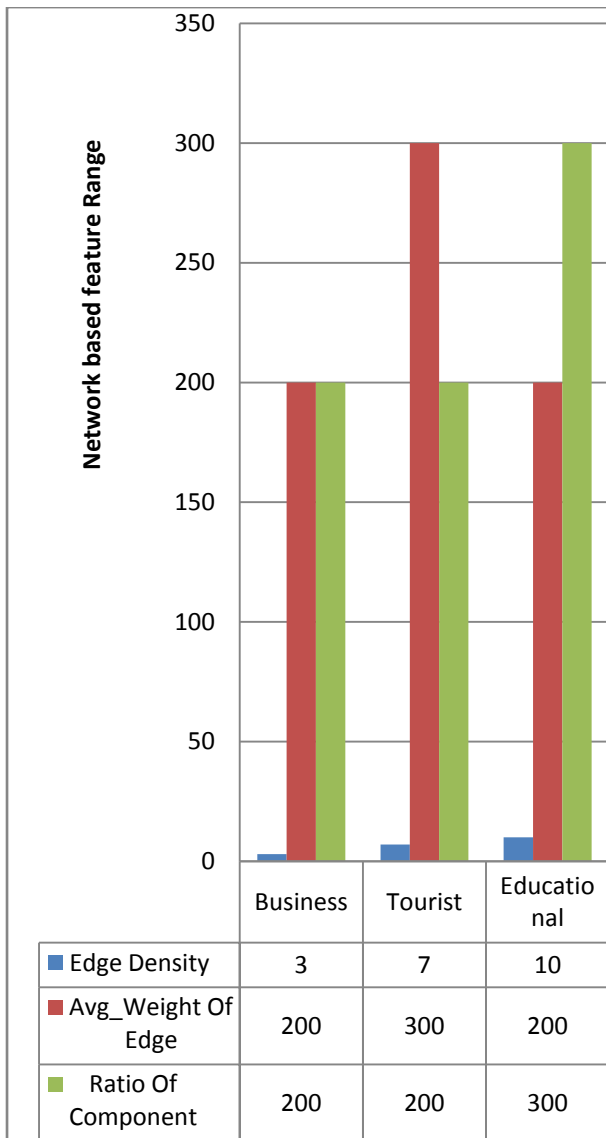|	Business	Tourist	Educational
Edge Density	3	7	10
Avg_Weight Of Edge	200	300	200
Ratio Of Component	200	200	300

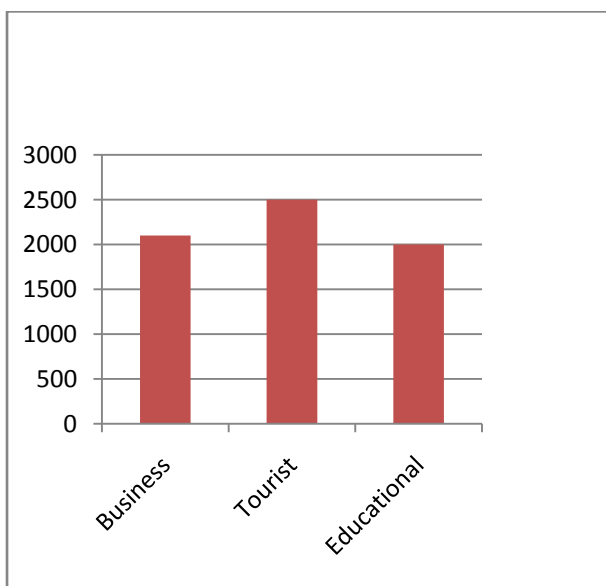**Figure 4: Network based feature graph.**

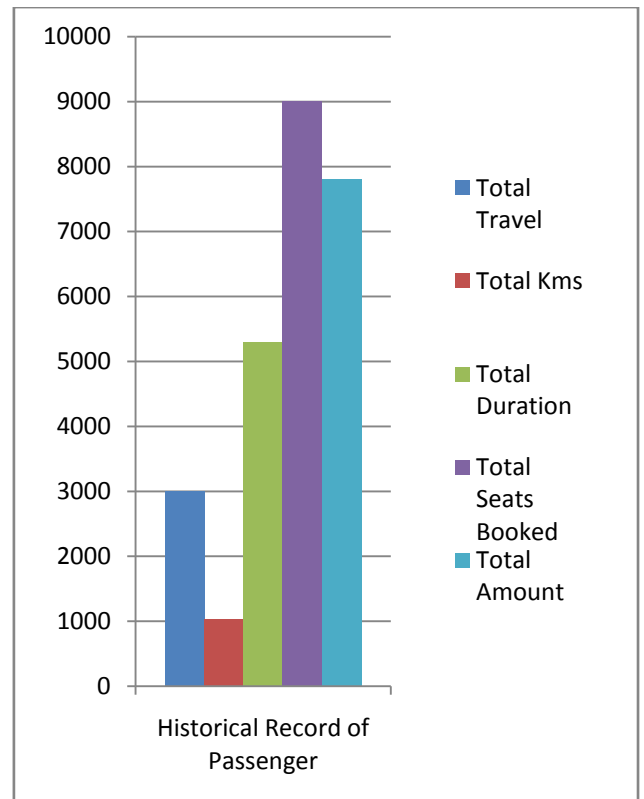**Figure 5: Group wise passenger analysis.**

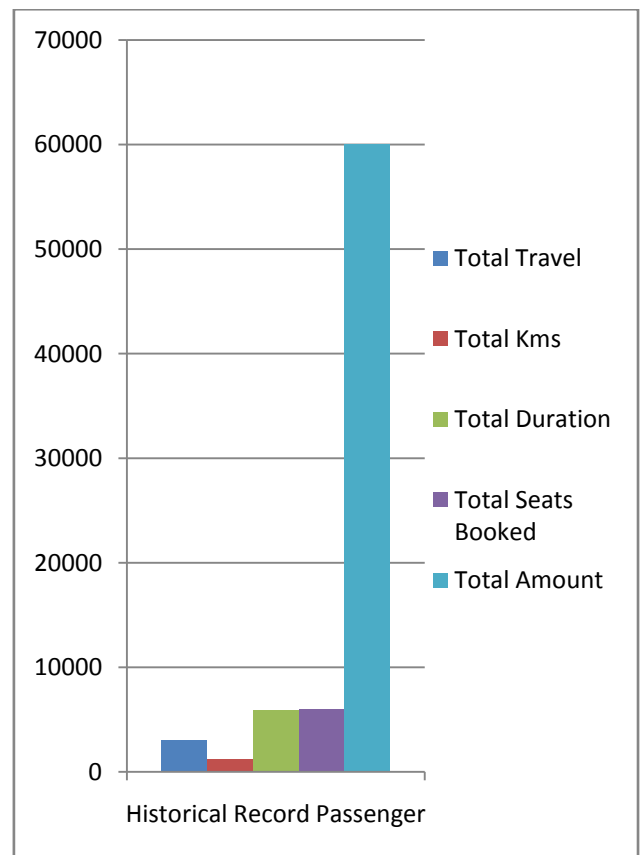**Figure 6: Historical record of overlapping passenger first.**

Figure 7: Historical record of overlapping passenger second.

6. CONCLUSION

In this paper, we have studied the problem of speculating the travel purposes of passenger groups and proposed a iterative classification approach for railway transportation system based on collective inference. We first created original dataset for railway transportation system. Then build co-travel networks by extracting social relations between passengers from their historical travel record. Then, we generate series of basic features as well as network based features and used overlapping relations between passenger groups to represent the dependence relations between their labels. Finally, we collectively speculate all the labels by using an ICA.

We performed our proposed approach on a real world data set in the field of railway transportation system. The experimental results showed that our proposed iterative classification performed well. One can implement this application as web based system in future.

7. REFERENCES

[1] R. Nicole, J. Carlson, and P. J. Rosenberger, III, "Factors affecting group-oriented travel intention to major events," *J. Travel Tourism Marketing*, vol. 29, no. 2, pp. 185–204, 2012.

[2] L. Wu, J. Zhang, and A. Fujiwara, "Representing tourists heterogeneous choices of destination and travel party with an integrated latent class and nested logit model," *Tourism Manage.*, vol. 32, no. 6, pp. 1407–1413, 2011.

[3] R. March and A. G. Woodside, "Testing theory of planned versus realized tourism behavior," *Ann. Tourism Res.*, vol. 32, no. 4, pp. 905–924, 2005.

[4] S. Boo and D. L. Jones, "Using a validation process to develop market segmentation based on travel motivation for major metropolitan areas," *Travel Tourism Marketing*, vol. 26, no. 1, pp. 60–79, 2009.

[5] K. Hung and J. F. Petrick, "Why do you cruise? Exploring the motivations for taking cruise holidays, the construction of a cruising motivation scale," *Tourism. Manage,* vol. 32, no. 2, pp. 386–393, 2011.

[6] S. C. S. Jang, B. Bai, G. S. Hong, and J. T. O'Leary, "Understanding travel expenditure patterns: A study of Japanese pleasure travelers to the United States by income level," *Tourism Manage.*, vol. 25, no. 3, pp. 331–341, 2004.

[7] S. I. So and X. Y. Lehto, "The situational influence of travel group com-position: Contrasting Japanese family travelers with other travel parties," *Travel Tourism Marketing*, vol. 20, no. 3/4, pp. 79–91, 2007.

[8] Sirakaya and A. G. Woodside, "Building and testing theories of decision making by travelers," *Tourism Manage.*, vol. 26, no. 6, pp. 815–832, 2005.

[9] Youfang Lin, Huaiyu Wan, Rui Jiang, Zhihao Wu, and Xuguang Jia," Inferring the travel purposes of passenger groups for better understanding of passengers," 2014.

[10] D. Jensen, J. Neville, and B. Gallagher, "Why collective inference improves relational classification," in *Proc. 10th SIGKDD*, 2004.

[11] D. Jensen and J. Neville, "Linkage and autocorrelation cause feature selection bias in relational learning," in *Proc.19th ICML*, vol.19, pp.259–266, 2002.

[12] J. Neville and D. Jensen, "Leveraging relational autocorrelation with latent group models," in *Proc. 5th ICDM*, vol. 5, pp. 322–329, 2005.

[13] J. Neville and D. Jensen, "Iterative classification in relational data," in *Proc.AAAI Workshop SRL*,vol.17, pp.13–20, 2000.

[14] P. Sen *et al.*, "Collective classification in network data," *AI Magazine*,vol. 29, no.3, pp.93–106, 2008.

[15] L. K. McDowell, K. M. Gupta, and D. W. Aha, "Cautious collective classification," *J. Mach. Learn. Res.*, vol. 10, pp. 2777–2836, 2009.

[16] S.C. S. Jang, B. Bai, G. S. Hong, and J. T. O'Leary, "Understanding travel expenditure patterns: A study of Japanese pleasure travelers to the United States by income level," *Tourism Manage.*, vol. 25, no. 3, pp. 331–341, 2004.

[17] D. Jensen, J. Neville, and B. Gallagher, "Why collective inference improves relational classification," in *Proc. 10th SIGKDD*, 593–598, 2004.

[18] R. Xiang and J. Neville, "Collective inference for network data with copula latent Markov networks," in *Proc. 6th WSDM*, pp. 647–656, 2013.

[19] J. Neville and D. Jensen, "Iterative classification in relational data," in *Proc. AAAI Workshop SRL*, pp. 13–20, 2000.

[20] N. Eagle, A. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

[21] Youfang Lin, Huaiyu Wan, Rui Jiang, Zhihao Wu, and Xuguang Jia,"inferring high quality co-travel networks," 2013.

[22] Youfang Lin, Huaiyu Wan, Rui Jiang, Zhihao Wu, and Xuguang Jia, "Discovering family groups in passenger social networks,"journal of computer science & technology,30(5):1141-1153,2015.

[23] Blondel V D, Guillaume J L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): Article No. 10008, 2008.

[24] Rosvall M, Bergstrom C T. Map of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4): 1118-1123, 2008.

[25] L. K. McDowell, K. M. Gupta, and D. W. Aha, "Cautious collective classification," *J. Mach. Learn. Res.*, vol. 10, pp. 2777–2836, 2009.