Inferring the Travel Purposes of Passenger Groups for Better Understanding of Passengers :A Survey

Shirin A. Maniyar
Department of Computer Science and Engineering
M.S.Bidwe Engineering college,Latur, Maharashtra,India

Pooja Shinde
Department of Computer Science and Engineering,
M.S.Bidwe Engineering college,Latur, Maharashtra,India

Abstract- In this paper the purpose of passenger who are traveling together in group having same destinations & same purposes with other peoples are determined by forming groups on the basis of extraction of features. To perform this operation we have used iterative classification algorithm, in that first cotravel network is generated based on the constraints of the passengers record which is taken from the transportation system at the time of ticket booking then features are extracted i.e. basic features & network based features with the help of these features the grouping is formed then overlapping relations between the group passengers is also determined. After that we collectively performed iterative classification algorithm. Finally we get the travel purpose of the passengers travelling together in groups after applying collective iterative classification.

Keywords – Co-travel network, iterative classification, passenger group, travel purpose, collective inference.

I. INTRODUCTION

In today's era people generally wants to enjoy every moment in life so they are utilizing every second of life by entertaining themselves. There are several sources for enjoyment of people like traveling with friends, visiting to relatives, watching movie with friends, families etc.So this can be a different enjoyment sources. For example Japanese tourist are visited repeatedly to U.S. Japanese travelers have been one of the largest groups of travelers to U.S. There are different purposes of different travelers , while going to traveling they arrange get together for all relatives so their relatives meet to each other & give gifts as well as their any requirement to tourist so Japanese tourist visited repeatedly to overseas for giving return gifts to close friends/ relatives [1].So different people travel for different purposes some time may be single or may be with friends or in groups.

People traveling in groups to the same destination & especially having same purposes such as business, education, medical problem or sightseeing, or visit to relative [2]. So this can be a new research are in transportation system to work with so that the transportation system or even government transportation can also improve their services & policies & get as much as profit in their area. For example transportation system can provide or recommend meeting places for business groups & also recommend travel routes for tourist groups & even recommend popular & best hospitals for passengers who are traveling for medical purposes [2].

This research is applicable to any transportation system whether it may be a civil aviation or railway transportation or any government or private travel transportation system. The potential application of this research is that among all transportation system they can adjust their schedules according to their passenger groups. For example airline transportation system can change their flight according to passenger groups or railway transportation system can change in their traveling system by releasing extra trains while occasional seasons like festival holidays, summer vacations, or buses transportation system can releave special buses for students at the school or college timings. So in this way any transportation whether it may be a private or government transportation system can manage their schedule & improve their service system & get as much as profit & popularity in the world.

For implementing this application we should know the travel purpose of particular group. To develop this application we need to have the historical data of passengers in group to determine whether it is tourist group or a business group figure I. Gives an example of civil aviation passenger group that consist of eight passengers who book tickets and travel together[2].

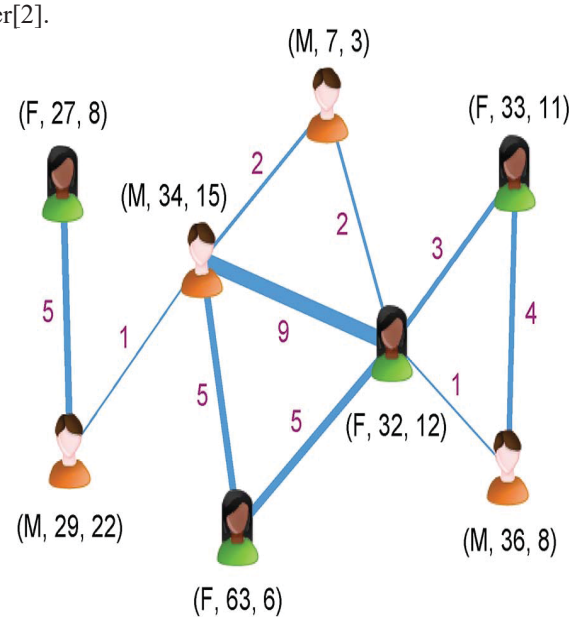

Figure 1 : Example of a civil aviation passenger group [2].

In above figure 1 the letter and numbers in bracket indicates the gender, age, and historical travel times of passenger, the number on edge indicates the historical cotravel times between two passengers (current cotravel is not counted)[2].Then, the problem is to infer the travel purpose of this passenger group.

Initially searching the purpose of passenger group is a classification problem and the solution is applying a traditional classifiers. The traditional classification is based on IID (independent identical distribution) which indicates that any of the two passenger groups is independent of each other. But in real world one person or number of people can be frequently appear in number of Different groups i.e. passenger groups might overlap with each other. And this results to overlaps among the different passenger groups makes no longer independent of each other. Traditional classifier doesn't consider this fact so it leads to inaccuracy. That why we need to simultaneously decide on all labels of passenger groups together & this is called collective inference.

Collective inference model uses auto correlation dependence relation [3], [4] between the variables of related entities which improves prediction. By estimating joint probability distribution over entire graph [5]-[8].
For implementing the collective inference and generating features for passenger groups we need to build relation between passengers and construct passenger social networks first[2]. Any transportation systems have reserved records of passengers who booked tickets together in their passenger information system (PNR).

Based on those records we can examine the passenger historical travel behavior and construct passenger social relation (here we call it as co-travel network). This type of social relationship beneficial to us for guessing the travel purpose of passenger. After constructing cotravel network, researchers propose a collective inference method based on the idea of iterative classification [9],[10],[11]. After that we use overlapping concept & finally collectively infer all the labels iterative way.

Researchers experiment on real data set of passenger travel records in the field of civil aviation in china & their proposed method can efficiently infer the travel purpose of passenger groups in cotravel network [2].

The rest of the paper contains literature survey which explains the necessary steps to implement this application as well as the methods, algorithms which are used while inferring the travel purposes of passenger groups.

II. LITERATURE SURVEY

This section explains important methods used for implementing this paper.

A. Collective inference -

Inferring the object collectively rather than classify them separately is called collective inference. Collective inference guess the joint distribution of the class label & attributes on the object & the class label of adjoining objects also. For example: this is equal as a particular topic of WebPages depends on those of adjoining pages [12].

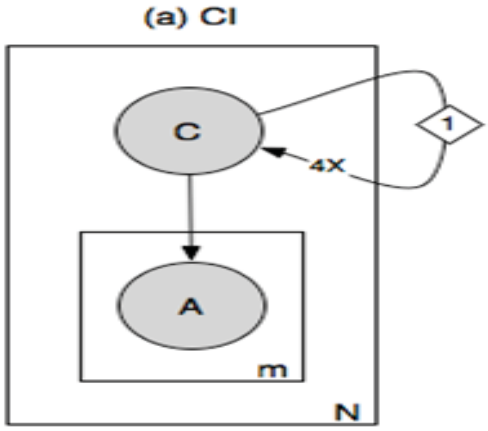

Figure 2 : Collective inference model [12].

The figure 2 states that the words on a webpage (the attributes Ai) depend only on the topic of that page (*C*) and are independent of the topic as well as depend upon neighbor four WebPages. Collective inference performs operation collectively on all data i.e.it find interlink depending relation from all depending WebPages.

Many studies indicate that collective inference can significantly reduce classification error as compared with traditional classifiers. Collective inferences can prowess relational autocorrelation [12] i.e. the value of a variable for one instance is highly correlated with the Value of the same variable on another instance [12]. Many studies of collective inference have reported large reductions in error when the method is applied.

B. Why collective model –

There are several reasons behind using collective inference. Some of them are discussing here.

1) Collective models can make use of known class label to improve inferences about unknown labels.(ex:known topics of WebPages)[12].

2) Collective inference improves the accuracy of relational models.

For performing collective classification there are four popular inference algorithms used these are as follows [10].

1) Iterative classification
2) Gibbs sampling
3) Loopy belief propagation.
4) Mean –field relaxation labeling.

In the area of research there are number of approaches for collective classification. At the higher level of abstraction these approaches can be divided into two different types [10].

1) Local conditional classifiers.

2) Global formulation.

Above four algorithms are categories into these two approaches. Under the local conditional classifier iterative classification algorithm (ICA) and Gibbs sampling are most commonly used for inference. Under the global formulation the remaining two are used. This research used iterative classification with collective inference concept so this survey focuses on ICA algorithm. For this application only (explain in further section) generalized ICA concept & algorithm is explained in next section.

C. Iterative classification-

In relational data, classification algorithm use knowledge about one object & with the help of that knowledge we can guess about related objects. If two objects are related, guessing something about one object can help to guess about the other & this is said to be "Iterative classification". Iterative classification gives a significant increase in classification accuracy as compared with a single pass approach.

Guesses are made at the initial stage & are given to the data so that the subsequent guesses can be possible [9].for instance, [10] shows an example how the ICA works, how they label to unobserved nodes with the help of observed nodes. The generalized algorithm for ICA is as follows for finding unobserved label from observed labels[9][10].

for each node *Yi* €*Y* do
  Compute a*i* using only *X* ∩ *Ni*
    Yi←*f*(a*i*)
end for
Repeat
  Generate ordering *O* over nodes in *Y*
for each node *Yi* *O* do
  Compute a*i* using current assignments to *Ni*
    yi ←*f*(a*i*)
end for
until all class labels have stabilized or a threshold number of iterations have elapsed.

Algorithm 1: Iterative classification algorithm [10].

The following section explains the important steps occurred for performing this implementation.

Table 1: Notation list [2].

Notation	Description	Notation	Description
G	Passenger group.	$gender_i$	Gender of passenger p_i.
P_g	Member set of passenger group g.	$seat_i$	Seat number of passenger p_i.
ori_g	Origin of passenger group g.	$mile_i$	Total historical travel mileage of passenger p_i.
dst_g	Destination of passenger group g.	$\#travel_i$	Total historical travel times of passenger p_i.
dis_g	Distance between ori_g and dst_g.	G=(V,E)	Co-travel network.
d_date_g	Departure date of passenger group g.	$G=(V_g,E_g)$	Corresponding subgraph of passenger group g in G.
r_date_g	Return date of passenger group g.	d_i	Degree of passenger p_i.
$\#male_g$	Number of male of passenger group g.	e_{ij}	Edge between passenger p_i and p_j in G.
P	Passenger.	w_{ij}	Weight of edge e_{ij}.
i,j,k	Index of passenger.	T_i	Set of neighbors of passenger p_i.
age_i	Age of passenger p_i.	$\#comp_g$	Number of components of subgraph G_g.

D. *Cotravel network construction* – A cotravel network is a graph $G = (V,E)$, where V is a node set, and each node $pi \in V$ represents a passenger; E is an edge set, and each edge $eij \in E$ indicates that passengers pi and pj have traveled together at least once[2].

As previously mentioned, any carrier service maintains a record of passengers (PNR) which shows all historical behavior of passenger. For such record we have to construct cotravel network, in which edges to nodes revel social relationship between passengers in the real world.
For clarity the notations used in algorithm is given in Table –1[2].

Given a set of travel records $S = \{g1, g2, \ldots, gn\}$, we simply construct a cotravel network $G = (V,E)$ by extracting relations from each passenger group, as outlined in Algorithm 2[2].

Input: $S=\{g_1,g_2,g_3\ldots.g_n\}$;

Output: G = {V,E};

for each passenger group $g \in S$

for each passenger pair (p_i,p_j) $(p_i, p_j \in P_g, p_i \neq p_j)$
if $p_i \notin V$ **then**

$V \leftarrow V \cup \{p_i\}$;

end if;

if $p_j \notin V$ **then**

$V \leftarrow V \cup \{p_j\}$;
end if;

if $e_{ij} \notin E$ **then**

$W_{ij} = 1$;

$E \leftarrow E \cup \{ e_{ij} \}$;

else

$W_{ij} = W_{ij} + 1$;

end if;

end for;

end for;

E. Generating features –

After constructing co-travel network we have to generate series of features for classification of passenger groups. The features are divided into two categories. One is basic features which are extracted from each passenger group from the PNR records. Other is network based features, which are generated referencing to cotravel network.

1. Basic features –

These are simple characteristics that can be directly generated from the passenger group without co travel network. Including holistic attributes of the current travel, demographic characteristics of group members, & historical travel static's of group members.

1.1 Holistic group attributes – Passenger group is limited a particular travel so the holistic attribute of the group are most basic features for inferring the travel purpose. Holistic attributes are summarized in Table –2.

1.2 Demographic characteristics – Demographics features are second types of basic features .These features are related to age of particular person because every group members having different age. The demographic features are summarized into Table 3 & formulization.

1.3 Historical characteristics – The historical feature plays an important role while inferring travel purposes of passenger group. These features are related to traveling history of passenger group. The historical features are summarized in Table– 4 & formulization.

2. Network based features –

As explained earlier basic feature are directly extracted from passenger PNR records without cotravel network. Whereas network based features are extracted with referencing cotravel network. Network based features are extracted with referencing to cotravel network. Network based feature are summarized in Table – 5 formulization.

Table –2: List of holistic group attribute [2].

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

Table 3: List of demographic characteristics [2].

Feature	Formulization	Feature	Formulization
average age	$avg\_age_g = \frac{\sum p_i \epsilon p_g age_i}{\|p_g\|}$	ratio of seniors	$senior\_ratio_g = \frac{\{p_i \| age_i > 60 p_i \epsilon p_g\}}{\|p_g\|}$
age variance	$Age\_var_g = \frac{\sqrt{\sum p_i \epsilon p_g (age_i - avg\_age_i)^2}}{\sqrt{\|p_g\|}}$	ratio of adults	$adult\_ratio_g = \frac{\{p_i \| 18 \geq age_i \geq 60 p_i \epsilon p_g\}}{\|p_g\|}$
Ratio of minors	$minor\_ratio_g = \frac{\{p_i \| age_i < 18 p_i \epsilon p_g\}}{\|p_g\|}$	ratio of male	$Male\ ratio_g = \frac{\#male_g}{\|p_g\|}$

Table 4 : List of historical characteristics[2].

Feature	Formulization
total mileage	$sum\_mile_g = \sum p_i \epsilon p_g mile_i$
Average mileage	$avg\_mile_g = \frac{sum\_mile_g}{\|p_g\|}$

total travel times	$\#sum\_travel_g = \sum_{p_i \in p_g} \#travel_i$
average travel times	$\#avg\_travel_g = \frac{\#sum\_mile_g}{\lvert p_g \rvert}$

Table 5 : Network based features[2].

Feature	**Formulization**
Edge density	$density_g = \frac{\lvert E_g \rvert}{\lvert V_g \rvert (\lvert V_g \rvert - 1)/2}$
Average weight of edge	$avg\_weight_g = \sum_{e_{ij} \in E_g} w_{ij}$
ratio of components	$comp\_ratio_g = \lvert \frac{\lvert V_g \rvert - \#comp_g}{\lvert V_g \rvert - 1} \rvert$
average clustering coefficient	$\#avg\_clus_g = \frac{\sum_{p_i \in V_g} \lvert \{ e_{jk} \mid p_j, p_k \in T_i, e_{jk} \in E_g \} \rvert}{\frac{d_i(d_i-1)}{2}} / \lvert V_g \rvert$
Average embeddedness	$avg\_emb_g = \sum_{p_i \in V_g} (\frac{1}{d_i} \sum_{p_j \in T_i} \frac{\lvert T_i \cap T_j \rvert}{\lvert T_i \cup T_j \rvert}) / \lvert V_g \rvert$

F. Iterative classification –

Iterative classification is explained earlier. Here the iterative classification algorithm is used for overlapping concept [2] (shown in figure II) because there may be different passengers which may exist in business group & with other group also .So this is very important fact that overlapping should be find out & collectively infer the travel purpose of all passenger group. The solution to this is use the iterative classification collectively so that no passenger should be left from overlapping & infers accurate travel purpose of passenger group By using iterative classification algorithm collectively we have removed constraint of IID [12]. The iterative classification algorithm is summarized in Table- VII algorithm-II [2] for the implementation of this application.

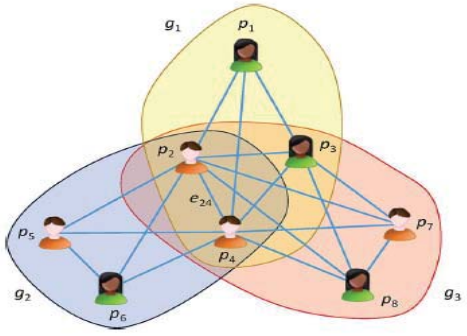

Figure 2: Example of overlapping of passenger group [2].

Algorithm II: Iterative classification algorithm [2].

Input: $S=\{g_1, g_2, g_3, \ldots g_n\}$; G = {V,E}; F:;
Output: $P(Y) = \{P(y_1), P(y_2), \ldots P(y_n)\}$;
for each edge $e_{ij} \in G$
for each possible label value $l \in L$

$P^{(0)}(y_{ij}= l) = 1/|L|$;
end for;
end for;
repeat
for each passenger group g ∈S
for each possible label value l ∈ L
$P^{(t)}(y_g= l) = \frac{\sum_{e_{ij} \in E_g} P^{(t-1)}(y_j=l)}{|E_g|}$;
$P^{(t)}(y_g = l) = f(F, P^{(t)}(y_g = l))$;
end for;
end for;
for each edge e_{ij} ∈ G
for each possible label value l ∈ L
$P^{(t)}(y_{ij}= l) = \frac{\sum_{g:E_g \in e_{ij}} P^{(t-1)}(y_g=l)}{w_{ij}}$;
end for;
end for;
until $|P^{(t)}(y_g = l) - P^{(t-1)}(y_g = l)| < \xi$;

IV.CONCLUSION

In this paper we have studied the problem of inferring the travel purpose of passenger group & the iterative classification algorithm solves this problem iteratively based on the concept of collective inference by first constructing cotravel network i.e. extracting social relationship between passengers then extracting the basic features & network based features for classification of passenger group. Then used the overlapping relations between passenger group labels & finally by using ICA algorithm to collectively infer all the labels. In this way by inferring the travel purpose of passenger group the govt /private carrier services can improve their service patterns & get as much as possible benefit.

REFERENCES

[1] 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, Jun. 2004.
[2] Youfang Lin, Huaiyu Wan, Rui Jiang, Zhihao Wu, and Xuguang Jia," Inferring the travel purposes of passenger groups for better understanding of passengers," 2014.
[3] 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.
[4] J. Neville and D. Jensen, "Leveraging relational autocorrelation with latent group models," in *Proc. 5th ICDM*, vol. 5, pp. 322–329, 2005.
[5] M. Jaeger, "Relational bayesian networks," in *Proc.13th UAI*, vol 13, pp. 266–273, 1997.
[6] B. Taskar, P. Abbeel, and D. Koller, "Discriminative probabilistic models for relational data," in *Proc.18th UAI*, vol.18, pp.485–492, 2002.
[7] P. Domingos and M. Richardson, "Markov logic: A unifying framework for statistical relational learning," in *Proc.ICML Workshop SRL*, pp. 49–54, 2004.
[8] J. Neville and D. Jensen, "Relational dependency networks,"*J. Mach. Learn.Res.*, vol. 8, pp. 653–692, 2007.
[9] J. Neville and D. Jensen, "Iterative classification in relational data," in *Proc.AAAI Workshop SRL*, vol.17, pp.13–20, 2000.
[10] P. Sen *et al.*, "Collective classification in network data," *AI Magazine*,vol. 29, no.3, pp.93–106, 2008.
[11] L. K. McDowell, K. M. Gupta, and D. W. Aha, "Cautious collective classification," *J. Mach. Learn. Res.*, vol. 10, pp. 2777–2836, 2009.
[12] D. Jensen, J. Neville, and B. Gallagher, "Why collective inference improves relational classification," in *Proc.10th SIGKDD*, vol.10, pp.593– 598, 2004.