Semi-Supervised Learning of User-Preferred Travel Schedules

(Extended Abstract)

Amrudin Agovic[*†]
University of Minnesota
aagovic@cs.umn.edu

Maria Gini [*]
University of Minnesota
gini@cs.umn.edu

Arindam Banerjee [†]
University of Minnesota
banerjee@cs.umn.edu

ABSTRACT

We present a graph-based semi-supervised approach for learning user-preferred travel schedules. Assuming a setting in which a user provides a small number of labeled travel schedules, we classify schedules into desirable and non-desirable. This task is non-trivial since only a small number of labeled points is available. It is further complicated by the fact that each schedule is comprised of multiple components or aspects which are different in nature. For instance in our case arrival times are modeled by probability distributions to account for uncertainty, while other aspects such as waiting times are given by a feature vector. Each aspect can thought of as a different type of observation for the same schedule While existing label propagation approaches can exploit vast amounts of unlabeled data, they cannot handle multi-aspect data. We propose Multi-Aspect Label Propagation (MALP), a novel approach which extends label propagation to handle multiple types of observations.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms

1. INTRODUCTION

Finding the best combination of flight tickets, hotels and car rental places when going on a trip can be a very challenging task. There are various websites providing supposedly attractive tickets. The traveler has to go through a jungle of possible offers, frequently within the same price range, to find what best suits her needs. It is our goal to aid a traveler in finding the most preferred schedules for a trip. Given a small set of desirable schedules, and a small set of non-desirable schedules, our objective is to classify unseen schedules.

[*]Partial support provided by NSF under grant IIS-0414466.
[†]Partially supported by NASA grant NNX08AC36A and NSF grant IIS-0812183.

Cite as: Title (Short Paper), Author(s), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10–15, 2009, Budapest, Hungary, pp. XXX-XXX.

One of the biggest challenges in creating a model for classifying schedules is that each flight schedule is composed of multiple types of components. Within a travel schedule we might have several flights, waiting times and airlines, which are all different in nature. The conventional approach to represent observations using one feature vector is not appropriate, since in this case not all features are of the same type, some might be categorical, others numerical. We want a model which takes into account the fact that each schedule is comprised of multiple components, and which can also exploit the structure within these components. Throughout the rest of this paper we refer to schedule components as aspects.

As in many real-world problems, in our application the ratio of labeled to unlabeled data is rather small. Therefore building a reliable supervised model is extremely difficult. In semi-supervised learning the objective is to exploit both the structure of unlabeled data and labaled data when making predictions. The vast amount of unlabeled data and the belief that it contains useful structure motivates the use of semi-supervised learning. Over the recent years there has been significant interest in graph-based semi-supervised learning resulting in a number of successful algorithms [3, 4, 2, 1, 5]. In graph-based semi-supervised learning unlabeled data are exploited by constructing a similarity graph over the data. Labels are subsequently "propagated" over this graph. While these methods can work well even with only few labeled points, they cannot handle multiple types of observations.

We propose Multi-Aspect Label Propagation (MALP). Given a multi-aspect data set we compute a similarity graph within each aspect. A weight is assigned to each aspect by estimating how well it separates the training data. Labels are propagated using both within-aspect similarities and estimated between-aspect similarities.

2. REPRESENTATION OF SCHEDULES

We represent a flight schedule by six aspects

1. Flight to destination: Represented by two Gaussian distributions indicating arrival and departure times, while capturing probabilities of delay.
2. Flight from destination: Same as above.
3. Wait Time: Feature vector indicating amount of total waiting time and the number of stops in each direction.
4. Milage: Number of miles for the entire trip.
5. Airline: An indicator vector showing which airlines are used in this trip.
6. Price: The price for the entire trip.

These aspects have been selected based on the fact that they require different representation and the belief that some of them contain important information.

3. PROPOSED APPROACH - MALP

We assume a partially labeled multi-aspect data set. Let y_i denote the label associated with point i, where $y_i \in \{-1, 1\}$. We denote the value of aspect a at point i by $x_i^{(a)}$.

Compute Within Aspect Similarity: The first step of our algorithm is to compute within aspect similarity weights. Within each aspect a we construct a K-nearest neighbor graph $G_a = (V_a, E_a)$. We then from an affinity matrix $Z^{(a)}$ with

$$z_{ij}^{(a)} = \exp(-d(x_i^{(a)}, x_j^{(a)})/\sigma^2) \tag{1}$$

where d denotes a distance function. For aspects which are represented by probability distributions we use relative entropy, otherwise we use L2-distance. Weights are normalized to sum to one.

Compute Between Aspect Similarity: The relevance of each aspect is estimated using training data. Let A be the set of points where $y_i = +1$, and B the set of points where $y_i = -1$. We consider the set of all labeled points $L = A \cup B$. The weight for aspect a is computed as normalized distance between classes:

$$g^{(a)} = \frac{\sum_{i \in A, j \in B} d(x_i^{(a)}, x_j^{(a)})}{\sum_{i \in L, j \in L} d(x_i^{(a)}, x_j^{(a)})} \tag{2}$$

Aspects for which within-class distances are small and between-class distances are large will tend to have a higher weight. We normalize g such that $\sum_a g^{(a)} = 1$.

Construct Main Graph: We construct a graph $G = (V, E)$ which includes all aspect graphs. V is the union of all V_a, and E the union of all E_a. Assuming that we have N data points and M aspects, we now have $N * M$ nodes in our graph, M nodes for each data point. For each data point we add one additional node. Let S be the set of newly added nodes. We connect each node in S to the corresponding node in each aspect by using between aspect weights. The nodes in S are the only nodes connected to all aspects. They are used in obtaining a final prediction. Let W be the resulting weight matrix for G.

Propagate Labels:: We now use the constructed graph to propagate labels. Let f_0 denote the initial labels for all nodes in G. For unlabeled nodes we set f_{0j} to zero. It is our goal to learn a function $f : V \rightarrow \{-1, 1\}$ which assign labels to each node in the graph. We learn f by minimizing the following objective function:

$$\min_f f^t L f + \mu ||f - f_0||^2 \tag{3}$$

where $L = I - W$ is the graph Laplacian.The objective function can be equivalently rewritten as

$$\min_f \sum_{ij} w_{ij}(f_i - f_j)^2 + \mu \sum_i (f_i - f_{0i})^2 \tag{4}$$

The first term, imposes a smoothness assumption preferring functions which do not vary abruptly over the graph. As shown in [1] this objective function can be minimized by iterative updates of the form:

$$f^{t+1} = \alpha W f^t + (1 - \alpha) f_0 \tag{5}$$

where $\alpha = \frac{1}{\mu+1}$ is a chosen parameter.

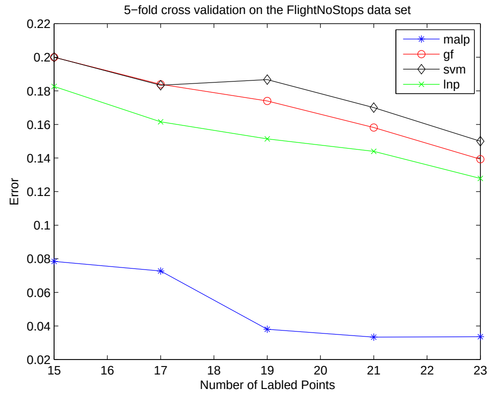

Figure 1: 5-fold cross validation results

4. EXPERIMENTS

Our results are based on a data set of 303 flight schedules from Minneapolis to Memphis, obtained using various websites. We compare our algorithm to Gaussian Fields (GF), Locally Linear Neighborhood Propagation (LNP) and Support Vector Machines (SVM). We created a labeling of the data set in which flight schedules with small waiting times are labeled as desired. For GF, LNP and SVMs we represent schedules as feature vectors, in which all aspect information is combined. In SVMs we use an RBF kernel. For GF and LNP L2-distance is used for constructing similarity graphs. Figure 1 shows the results of 5-fold cross validation. Conventional label propagation approaches (GF, LNP) appear to do better than SVM. This can be explained by the fact that the size of the training data is very small. Our proposed approach significantly outperforms the other methods. This does not come as a surprise since waiting time is modeled as an aspect.

5. CONCLUSIONS

While a more thorough evaluation of our method is needed, our preliminary results indicate that our approach has the potential to capture preferred schedules accurately. From an algorithmic perspective we have extended the applicability of Label Propagation to multi-aspect data sets.

6. REFERENCES

[1] F. Wang and C. Zhang. Label propagation through linear neighborhoods. In *ICML*, 2006.

[2] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf Learning with local and global consistency. In *NIPS*, 2003.

[3] O. Chapelle, B. Scholkopf, and A. Zien, editors. *Semi-Supervised Learning*. MIT Press, 2006.

[4] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *ICML*, 2003.

[5] M. Szummer and T. Jaakkola. Partially labeled classification with markov random walks. In *NIPS*, 2001.