

# Mining Sequential Association Rules for Traveler Context Prediction

Chad A. Williams\*  
Department of Computer  
Science  
University of Illinois at Chicago  
cwilliam@cs.uic.edu

Abolfazl (Kouros)  
Mohammadian  
Department of Civil  
Engineering  
University of Illinois at Chicago  
kouros@uic.edu

Peter C. Nelson  
Department of Computer  
Science  
University of Illinois at Chicago  
nelson@cs.uic.edu

Sean T. Doherty  
Department of Geography &  
Environmental Studies  
Wilfrid Laurier University  
sdoherty@wlu.ca

## ABSTRACT

Recent work has focused on creating models for generating traveler behavior for micro simulations. With the increase in hand held computers and GPS devices, there is likely to be an increasing demand for extending this idea to predicting an individual's future travel plans for devices such as a smart traveler's assistant. In this work, we introduce a technique based on sequential data mining for predicting multiple aspects of an individual's next activity using a combination of user history and their similarity to other travelers. The proposed technique is empirically shown to perform better than more traditional approaches to this problem.

## Categories and Subject Descriptors

I.5.2 [Pattern Recognition]: Design Methodology—Classifier design and evaluation

## General Terms

Traveler context

## Keywords

Sequential mining, travel patterns, activity prediction

## 1. INTRODUCTION

Modeling and predicting travel patterns has been a source of interest and a challenge for regional planners for a number of decades. In the past, data available for these studies has by in large been limited to aggregate traffic information and paper based travel surveys. With the increase in prevalence of hand held computers and GPS

---

\*This research was supported in part by the National Science Foundation IGERT program under Grant DGE-0549489.

devices, the detail and types of information that can be captured for travel histories are becoming far richer than ever before.

While these data sources share many characteristics with other studied travel activity logs, the additional information on decision processes and opportunities for further enhancement with GIS offer some unique challenges as well as opportunities in behavior prediction. In this work, we focus on techniques for extracting meaning from these augmented data sources. One of the challenges associated with understanding traveler behavior is its strong dependency on the transportation network. The strength of this relationship makes it nearly impossible to address behavior without also considering the overall *traveler context* in terms of location, mode of transit, what is being done, and the time of the activity to name a few aspects. A better understanding of how this context affects future behavior is likely to not only improve predictive models, but also provide additional insight into how choices are made among alternatives.

In this work, we explore the challenge of traveler prediction not as activity or location prediction, but as context prediction. Specifically we examine the problem of given a sequence of the traveler's prior context how well can the context of their next activity be predicted. Rather than focusing on prediction of common activity sequences based on the history of many travelers, we shift the paradigm to predicting the travel context of an individual based on their own travel history and the history of others. This subtle shift is to facilitate moving travel prediction from the realm of urban planning to enable personalized smart traveler assistant devices. The principle behind these devices are to provide intelligent services like multi-modal travel planning to support a user's travel needs [6, 22]. With this type of application, understanding the context of a traveler can help identify what type of information the users is interested in and help narrow potential options. Consider a scenario where a user wants to make a trip across town mid-day. If the assistant knew the user's likely mode of transit and accessibility thresholds, the system might be able to evaluate and suggest one transit route over another.

We propose a novel approach to predicting individual traveler behavior and associated context from a stream of their prior traveler

context. The proposed approach examines each step in the history and the predicted next step of the traveler as a set of attributes or characteristics describing the context of that step. As a whole, these steps form a sequence of sets describing the traveler’s context progressing through time, or an enhanced activity sequence. We introduce a technique for mining sequential associative rules from individual activity sequences and augment recall and thus overall performance by incorporating the patterns of other travelers through the traveler’s context.

## 2. MINING MULTI-VARIATE STREAMS

In this paper we introduce a method for mining temporal stream data based on constrained associative sequence mining introduced in Agrawal and Srikant [2]. Our contribution is an extension of these techniques for next step multi-variate prediction within a single discrete temporal stream. In this context, the multi-variate prediction refers to the prediction of the multiple variables or characteristics associated with the next traveler context simultaneously, as opposed to trying to classify each characteristic individually.

Association mining and sequential association mining has traditionally been applied in the business world for analyzing purchase patterns through market basket analysis. Since this technique was introduced in 1993 [1], associative pattern analysis has remained widely popular due in part to the readability of the found patterns and the intuitiveness of the strength of the rule [13]. The output of this type of sequential associative mining is a rule of the form:

$$\langle \{AB\} \{C\} \rangle \Rightarrow \{DE\}, \text{ with support} = 25\%, \text{ and} \\ \text{confidence} = 80\%$$

This rule can be read as: if the set  $\{AB\}$  is followed by the set  $\{C\}$  this implies the set  $\{DE\}$  will appear in the future. The strength of this pattern is that this full sequence,  $\langle \{AB\} \{C\} \{DE\} \rangle$ , occurred in 25% of the training examples (support); and for all training examples where  $\langle \{AB\} \{C\} \rangle$  appeared, the full sequence  $\langle \{AB\} \{C\} \{DE\} \rangle$  appeared 80% of the time (confidence). Thus, as this example shows, unlike some other model-based techniques for example neural networks, the model and patterns identified by this technique can be further analyzed and explained if desired. In addition, because the sequential patterns identified are only order dependent, meaning additional sets of items could occur between the identified sets without consequence as long as the sets are in that order, the technique is also more robust to noise in terms of still being able to identify the underlying pattern. From a traveler prediction perspective, this feature is likely to help in identifying the underlying patterns despite sporadic activities that may not be part of the regular routine of a traveler’s day.

Identifying patterns in traditional associative mining relies on multiple training sets for its primary constraint *support* [1]. With associative sequence mining, there is a similar dependency on multiple training sequences [2]. The implication of this when applied to the context of transportation is that for a travel or activity pattern to be significant, the pattern must be present across multiple travelers. While this constraint is likely a good guideline for predicting traveler patterns in general, if the goal is to predict the travel pattern of an individual, then patterns that are unique to that individual are likely significant for predicting future behavior of that individual even if they have little predictive value for the set of all travelers as a whole. In addition, these techniques are not well suited to lengthy sequences, as the distance between sets within a sequence is not accounted for. Applying this logic within the context of transporta-

tion, this would be equivalent to saying the likelihood of an event occurring is just as dependent on an activity that occurred 4 days ago as it is on the previous activity. While such relationships may exist, it seems reasonable to assert that activities that occurred in the traveler’s recent history are in general more likely to be better predictors of the immediate next activity.

Based on these perceived weaknesses of existing associative sequential mining techniques when applied to individual traveler prediction; we propose an alternative to the traditional constraints which we assert is better suited to this type of prediction problem. The purpose of the proposed technique is to identify frequent sequences within multi-variate temporal data, such as the many different characteristics describing the activity of a traveler, that can be used when patterns within a single stream are meaningful, and is more suited for lengthy sequences. Below we present an approach to constraining the problem of frequent sequence mining to allow existing associative sequential rule mining algorithms to work in this context.

### 2.1 Revised Constraint Definitions

An important difference between this new problem formulation and the typical data used with sequential association rule mining is the number of sequences. In traditional sequential association rule mining there are many sequences, each representing an instance of a specific occurrence of a sequence of item sets. As such, traditionally how frequently an item set or sequence of item sets occurs is defined in terms of the number of training sequences the set(s) occurs in, more formally support count of  $X$  is defined as:

$$X.\text{count} = \frac{\# \text{ of training sequences containing the item set } X}{\text{total } \# \text{ of training sequences}},$$

and the definition of the support of a sequence of an arbitrary number of item sets  $X Y \dots Z$  is defined as:

$$\text{support}(\langle \{X\} \{Y\} \dots \{Z\} \rangle) = \frac{(X_i \cup Y_j \cup \dots \cup Z_k).\text{count}}{\text{total } \# \text{ of training sequences}},$$

where  $i < j < k$  (number of times the item sets appear in that same order). By contrast, since we want to identify patterns in a single training sequence, we are interested in how frequently a set of items occurs in terms of the total number of item sets in the user’s sequence. Thus, we define the count of an item set as:

$$X.\text{count}' = \frac{\# \text{ of item sets containing (the items of set } X)}{\text{total } \# \text{ of item sets in the training sequence}},$$

and the definition of the support of an arbitrarily long sequence of item sets,  $X Y \dots Z$ , is thus also defined in terms of item sets, but additional consideration is taken since we are trying to capture support of sequences of item sets within a single sequence. Thus, without additional constraints any item anywhere in the sequence after the first item set could be considered support for the sequence regardless of their distance apart in the sequence. We thus want to constrain the number of item sets considered after the first item set in the sequence in question. To address this we use the concept of a **support window**. A support window specifies the item sets after the first item set to consider when calculating support. Thus when calculating the support of any potential frequent subsequence within the original sequence, there is a support window of some specified length  $w$  that slides to the first item set considered. Using this concept we define the support of a sequence of an

arbitrary number of item sets  $X Y \dots Z$  to be defined as:

$$\text{support}'(\langle\{X\}\{Y\}\{\dots\}\{Z\}\rangle) = \frac{(\langle X_i \cup Y_j \cup \dots \cup Z_k \rangle).\text{count}'}{\text{total \# of support windows in the training sequence}}$$

where  $i < j < k \leq (i + w)$ . Thus the number of times the sequence occurs within the window constraints of all possible windows in the sequence. All other common multiple support sequential mining constraints designed to help limit the search space such as  $\lambda$  (relative minimum support), support difference constraint, and minimum support threshold are calculated with respect to this new support measure [17]. Likewise, when rules are mined from the extracted frequent item sets, confidence is defined in terms of this definition of support.

The results reported in this paper were obtained by applying these constraints to the Generalized Sequential Patterns (GSP) algorithm [21], although it should be noted that any multiple support sequential association rule mining algorithm such as PrefixSpan [19] could be used with this technique.

### 3. RELATED WORK

Techniques for constraining and mining sequence association rules have been extensively studied since associative mining and later sequential mining techniques were introduced by Agrawal *et al.* [1, 2]. The sequential mining techniques introduced in that work also explored a similar concept to the sliding windows discussed here, but this idea was only explored in terms of defining support between sequences rather than as a mechanism for discovering patterns within a single sequence [2]. Zaki introduced length, width, and maximum gap constraints to reduce data mining time. In Harms *et al.* and Harms and Deogun an algorithm was introduced for mining frequent episodes from multiple sequences using time lag constraints and separating antecedent and consequent constraints [11, 12]. All of these works have focused on extracting and applying the associative rules uniformly across sequence series in prediction. In addition, all of these works tend to focus on predicting a single next item in a given sequence. Other work has examined set prediction as constrained sequence mining, but focused on traditional sequential association rules [10]. Work on label sequential rules were introduced and examined in the context of text mining for their benefits at being able to discern the context and applicability of rules [18, 14].

Much of the related transportation specific research falls into two general categories: micro-simulation, and individual travel prediction. In the area of micro-simulation, related work has primarily focused on using activity survey data for generating simulated activity schedules or verifying simulations [20, 15]. Recent work has examined using mental maps and cognitive learning for improving choice models through observations during micro-simulations [3]. However all of this work focuses on simulating behavior of artificial travelers. Other work has focused on predicting next location of individuals based on GPS traces [4]. Liao *et al.* extended this idea and examined this problem as an unlabeled activity model for predicting the next location [16]. Our work differs from these in that like the GPS based techniques, it focuses on individual behavior, but our goal is more similar to micro-simulation models, in that we are interested in modeling a more feature rich set describing the reason for the behavior beyond the locations.

## 4. EXPERIMENTS AND DISCUSSION

### 4.1 Evaluation Metrics

For measuring prediction performance, we use the information retrieval metrics of precision and recall [5]. The basic definition of recall and precision can be written as:

$$\text{precision} = \frac{\# \text{ true positives}}{\# \text{ true positives} + \# \text{ false positives}}$$

$$\text{recall} = \frac{\# \text{ true positives}}{\# \text{ true positives} + \# \text{ false negatives}}$$

For the purpose of this study, since we are primarily interested in the correctness of the attribute value (if the attribute appeared at all). Thus *# true positives* is the number of attribute values predicted correctly; *# false positives* is the number of attribute values incorrectly predicted where the attribute did appear in the time step, and *# false negatives* is the number of attributes no value was predicted for, but some value for the attribute appeared in the time step. Since these two measures are often associated with a trade-off of one for the other, we also examine a combined metric the F-measure [23] which can be calculated as:

$$\text{F-measure} = \frac{(2 \cdot \text{precision} \cdot \text{recall})}{\text{precision} + \text{recall}}$$

We use this metric to compare the balanced performance of the algorithms.

### 4.2 Methods

#### 4.2.1 Data

In this work we use a dataset derived from the Computerized Household Activity Scheduling Elicitor (CHASE) survey to evaluate the proposed constrained rule mining problem [8]. CHASE was conducted in Toronto between 2002-2003 and consists of travel activity information for 426 adults in 271 households. This survey focused on observed activity and travel patterns as well as the planning process associated with these activities captured over a 7 day period for each individual. The data was collected via a computerized scheduling program that allows survey participants to self-report on their actual activities as well as their on-going scheduling process. The user is asked to add, update, and delete their travel activity plans as their week progresses. The final state of the scheduled activities is recorded as the observed activity sequence similar to the activity sequences that are recorded in traditional surveys based on activity diaries. Additional details on the survey methods, sample characteristics, and data quality analysis can be found in [8, 7].

The CHASE survey captured a wide range of in-home and out-of-home activity types. These activities are captured and organized by a high-level classification as well as a more detailed activity type. In this work the high level activities examined are organized into ten classifications (Night sleep - other needs, Social, Meals, Work/School, Household Obligations, Drop-off/Pick-up, Shopping, Services, Active recreation, and Entertainment), which are further broken down into 52 specific activity types (Grocery shopping, Internet shopping ... exercise, active sports, etc.).

In addition to the activity type, detail on the involved persons was also captured in the survey. For each activity, subjects were asked to indicate other people that were directly involved with the activity from a pull-down list of names. The list of names was initially populated via interview and the relationship with that person was noted. As the survey progressed, participants could add new people to the list by specifying their name and relationship (son, daughter, spouse, relative, friend, etc.). Parents of young children were asked

to specify which of their children were “under their care”, to obtain further detail about this participation. This information was used to further differentiate the three main activity types from above into five categories used in the most recent versions of tour models [24], including:

1. Individual mandatory activities (i.e. not conducted with other household members, but may be with other non-household members)
2. Joint maintenance activities (i.e. conducted with other household members)
3. Joint discretionary activities (i.e. conducted with other household members)
4. Allocated maintenance activities (i.e. not conducted with other household members, but may be with other non-household members)
5. Individual discretionary activities (i.e. not conducted with other household members, but may be with other non-household members)

Almost all activities fell into this categorization, except a portion of individual mandatory activities conducted with other household members.

The data is structured such that each travel activity corresponds to an event in the person’s travel stream with a set of attributes corresponding to that event that capture information about the activity which we have referred to as the traveler context. The validity of using this type of activity data to create traditional multi-user tour based scheduling models has been analyzed in recent work [9]. For this paper we have selected a subset of these attributes (18 of approximately 200) and only activities that are adults only, current acts, good data, no driving acts, and no doing the survey acts. These 18 attributes listed below describe the portion of the traveler context examined in this work and all have discrete values:

- Activity Group
- Activity Specific Type
- Activity group (3 categories related to if tour was mandatory)
- Activity group (5 categories related to involvement of others in tour)
- Activity group (8 categories for tour analysis)
- Total number of children under care in the household that are involved
- Location id
- Location, observed (in or out-of-home)
- City
- Tour type: At-home (AH) or home-based (HB)
- Observed sequence of the AH or HB tour (1st AH tour of the day, 2nd AH tour etc)
- Observed sequence of the activity in the tour (1st act in the tour, 2nd act in the tour, ... )

- Duration flexibility binary indicator
- Time flexibility indicator
- Spatial flexibility indicator
- Whom the activity is conducted with
- Whom the activity is normally conducted with
- Interpersonal flexibility

These attributes were selected as they represent a mix of information about the type of activity, the location, relative time the activity took place (# tour of the day), and the person’s involved. Thus, the dataset can be thought of as a stream or sequence of events with the set of attribute values at each event being highly related. For this dataset, there are 41,312 sets of traveler context, with the average adult’s traveler context sequence being just over 92 sets in length. The CHASE dataset was selected in part due to the strength of the relationship between attributes both intra-event and inter-event; as well as the combination of non-trivial sequence length and enough streams to make the dataset suitable for comparing multiple approaches to this problem.

### 4.3 Experiments and Evaluation

In the results below, we also present the results of Naive Bayes and a first order Markov model for the purpose of a comparative baseline. For both of these models, a classifier was built for each of the 18 attributes potentially present in each time step. The results presented below evaluate the set produced by this group of classifiers at predicting the next attribute value set at the next step in the series. The numbers presented are the average across the 18 attributes.

To ensure the significance of our results, we have employed a 10 times cross-folding validation methodology. Since sequential data is by definition order sensitive, care must be taken to ensure order is preserved as much as possible while still producing meaningful cross-folding. For this we employed a loop technique designed to keep as much of the training data as sequential as possible while still allowing the test set to immediately follow the training data that would have been seen just prior to the test data in the actual event. To illustrate this, consider a sequence  $S$  that would be split into parts  $S_1, S_2, \dots, S_5$ . The resulting test series would be:

$$\begin{aligned} \text{Train}_1 &= S_1 S_2 S_3 S_4 & \text{Test}_1 &= S_5, \\ \text{Train}_2 &= S_5 S_1 S_2 S_3 & \text{Test}_2 &= S_4, \\ \text{Train}_3 &= S_4 S_5 S_1 S_2 & \text{Test}_3 &= S_3, \\ & \dots & & \end{aligned}$$

All results shown are the average of the 10 runs.

In the first set of experiments, we examine the impact of minimum support on predictive quality using a support window of 3 and a window size of 1 for predictions. Figures 1 and 2 depicts the effect minimum support has on precision and recall respectively with the additional support window constraint. As these figures show, like traditional association rules, the higher the minimum support, the more precise the rules, but there is a tradeoff of lower recall. Figure 3 displays the F-measure of these same models and show that the lower minimum support of 10% combined with a confidence threshold of 66% provides the best balance of precision and recall.

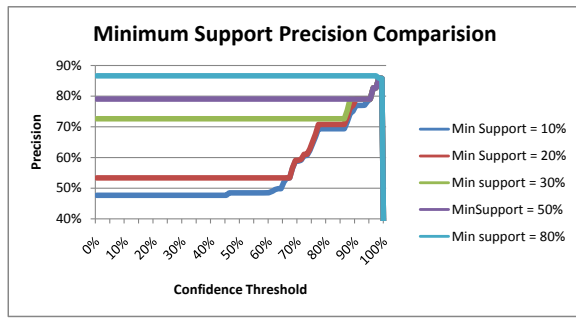


Figure 1: Comparison of precision for various minimum support.

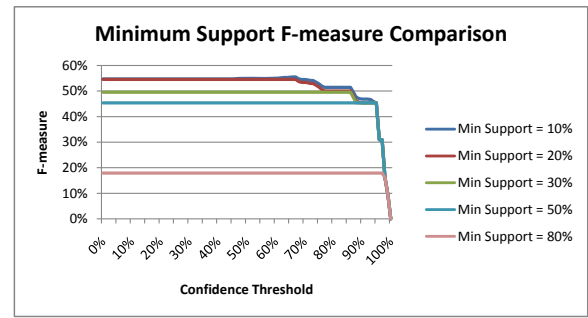


Figure 3: Comparison of F-measure for various minimum support.

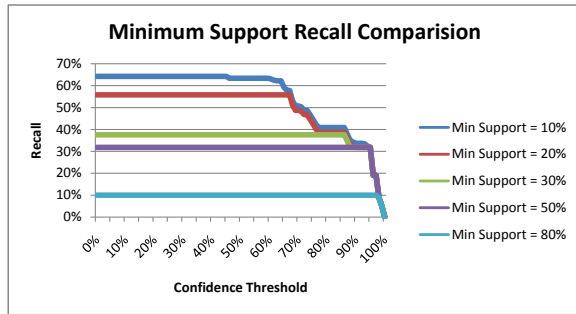


Figure 2: Comparison of recall for various minimum supports.

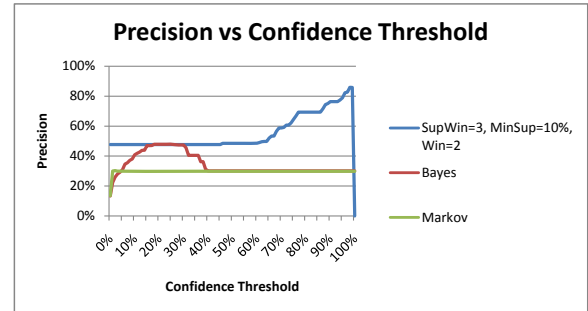


Figure 4: Comparison of precision for algorithms.

In the next set of experiments we compare the support window constrained algorithm, with two common algorithms as baselines. For the support window constrained algorithm we selected the best mix of support window, prediction window, and minimum support constraints based on F-measure for comparison; specifically a support window of 3, a prediction window of 2, and minimum support of 10%. Figures 4 and 5 depict how the proposed algorithm compares with each of these baselines with respect to precision and recall. As Figure 4 shows the proposed algorithm is as precise as either of the baselines across the entire confidence threshold range. A closer look at the recall shows that while Bayes has a higher recall at low confidence thresholds, the proposed algorithm has a better F-measure for the entire range as seen in Figure 6.

#### 4.4 Conclusion and Future Work

In this work we introduced a technique for constraining sequential associative mining algorithms so that individual streams of activity sequences can be mined effectively. As our results demonstrate, this technique shows significant promise for accurately predicting next step context compared to Bayesian and Markov approaches. This advantage is likely due to this type of technique being less susceptible to the noise in the form of unplanned or sporadic activities, since the technique allows common series of activities to be found even when there are often additional activities appearing in the sequences. For simplicity in the experiments above, the support window was specified by a fixed number of item sets; however it is easy to envision more complex notions of this same concept specific to transportation. For example, a more meaningful boundary in transportation may be within a day or within a tour. We plan to explore the benefit of a more complex transportation specific notion such as a tour for a support window in future work. As shown, an additional benefit of this technique is that the predictive model

is much more transparent than techniques such as neural networks, allowing additional insight to be gained from the models created. In future work, we intend to extend these ideas and enhance the way additional travel histories are selected for augmenting an individual's predictive model. Our hope is to identify techniques that can be used to extract what aspects of travel behavior can be reliable transferred across travelers and geographies.

#### 5. ACKNOWLEDGMENTS

The authors would like to thank the Social Sciences and Humanities Research Council of Canada for the collection of the CHASE activity data.

#### 6. REFERENCES

- [1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2):207–216, 1993.
- [2] R. Agrawal and R. Srikant. Mining sequential patterns. In P. S. Yu and A. S. P. Chen, editors, *Eleventh International Conference on Data Engineering*, pages 3–14, Taipei, Taiwan, 1995. IEEE Computer Society Press.
- [3] T. A. Arentze and H. J. P. Timmermans. Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. *Transportation*, 32(4):321 – 340, July 2005.
- [4] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Computing*, 7(5):275–286, 2003.
- [5] C. Cleverdon. Evaluation of tests of information retrieval systems. *Journal of Documentation*, 26:55–67, 1970.
- [6] J. Dillenburg, O. Wolfson, and P. Nelson. The intelligent travel assistant. In *Proceedings of the IEEE 5th International*

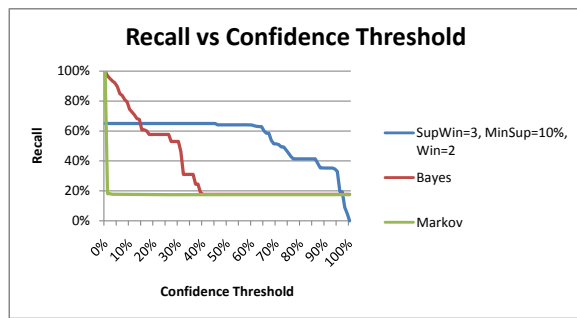


Figure 5: Comparison of recall for algorithms.

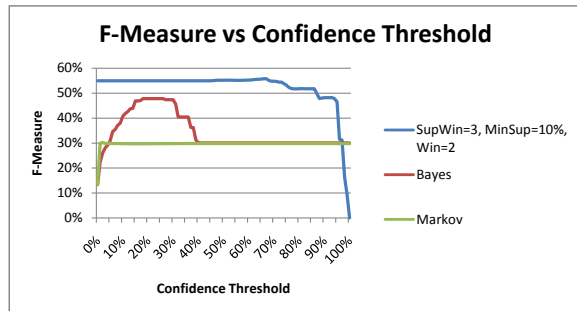


Figure 6: Comparison of F-measure for algorithms.

*Conference on Intelligent Transportation Systems*, pages 691–696, 2002.

- [7] S. Doherty, E. Nemeth, M. Roorda, and E. Miller. Design and assessment of the toronto area computerized household activity scheduling survey. *Journal of the Transportation Research Board*, 1894:140–149, 2004.
- [8] S. T. Doherty and E. J. Miller. A computerized household activity scheduling survey. *Transportation*, 27(1):75–97, February 2000.
- [9] S. T. Doherty and A. Mohammadian. The validity of using activity type to structure tour-based scheduling models. In *Proceedings of 86th Annual Meeting of the Transportation Research Board*, Washington D.C., January 2007.
- [10] M. N. Garofalakis, R. Rastogi, and K. Shim. Spirit: Sequential pattern mining with regular expression constraints. In *VLDB '99: Proceedings of the 25th International Conference on Very Large Data Bases*, pages 223–234, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.
- [11] S. K. Harms, J. Deogun, and T. Tadesse. Discovering sequential association rules with constraints and time lags in multiple sequences. In *Foundations of Intelligent Systems : 13th International Symposium, ISMIS 2002*, volume 2366, pages 373–376. Springer Berlin / Heidelberg, Lyon, France, 2002.
- [12] S. K. Harms and J. S. Deogun. Sequential association rule mining with time lags. *Journal of Intelligent Information Systems*, 22(1):7–22, 2004.
- [13] J. Hipp, U. Güntzer, and G. Nakhaeizadeh. Algorithms for association rule mining - a general survey and comparison. *SIGKDD Explorations Newsletter*, 2(1):58–64, 2000.
- [14] N. Jindal and B. Liu. Mining comparative sentences and relations. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, 2006.
- [15] M. S. Lee and M. G. McNally. On the structure of weekly activity/travel patterns. In *Transportation Research Part A: Policy and Practice*, volume 37, pages 823–839, 2003.
- [16] L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5-6):311–331, April 2007.
- [17] B. Liu, W. Hsu, and Y. Ma. Mining association rules with multiple minimum supports. In *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 337–341, New York, NY, USA, 1999. ACM Press.
- [18] B. Liu, M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 342–351, New York, NY, USA, 2005. ACM Press.
- [19] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. Mining sequential patterns by pattern-growth: the prefixspan approach. *Knowledge and Data Engineering, IEEE Transactions on*, 16:1424–1440, 2004.
- [20] O. Pribyl and K. G. Goulias. Simulation of daily activity patterns incorporating interactions within households: Algorithm overview and performance. In *Transportation Research Record*, volume 1926, pages 135–141, January 2005.
- [21] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In P. M. G. Apers, M. Bouzeghoub, and G. Gardarin, editors, *Proceedings of 5th Int. Conf. Extending Database Technology, EDBT*, volume 1057, pages 3–17. Springer-Verlag, 1996.
- [22] M. Torrens, P. Hertzog, P. Pu, and B. Faltings. Towards an intelligent mobile travel assistant. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, pages 1208–1209, New York, NY, USA, 2004. ACM Press.
- [23] C. van Rijsbergen. *Information Retrieval*. Butterworth, London, 1979.
- [24] P. Vovsha, M. Bradley, and J. Bowman. Activity-based travel forecasting models in the united states: Progress since 1995 and prospects for the future. In *EIRASS Conference on Progress in Activity-Based Analysis*, Maastricht, The Netherlands, 2004.