# Spatio–temporal Rule Mining: Issues and Techniques

Győző Gidófalvi[1] and Torben Bach Pedersen[2]

[1] Geomatic aps — center for geoinformatik
gyg@geomatic.dk
[2] Aalborg University — Department of Computer Science
tbp@cs.aau.dk

**Abstract.** Recent advances in communication and information technology, such as the increasing accuracy of GPS technology and the miniaturization of wireless communication devices pave the road for Location–Based Services (LBS). To achieve high quality for such services, spatio–temporal data mining techniques are needed. In this paper, we describe experiences with spatio–temporal rule mining in a Danish data mining company. First, a number of real world spatio–temporal data sets are described, leading to a taxonomy of spatio–temporal data. Second, the paper describes a general methodology that transforms the spatio–temporal rule mining task to the traditional market basket analysis task and applies it to the described data sets, enabling traditional association rule mining methods to discover spatio–temporal rules for LBS. Finally, unique issues in spatio–temporal rule mining are identified and discussed.

## 1 Introduction

Several trends in hardware technologies such as display devices and wireless communication combine to enable the deployment of mobile, Location–based Services (LBS). Perhaps most importantly, global positioning systems (GPS) are becoming increasingly available and accurate. In the coming years, we will witness very large quantities of wirelessly Internet–worked objects that are location–enabled and capable of movement to varying degrees. These objects include consumers using GPRS and GPS enabled mobile–phone terminals and personal digital assistants, tourists carrying on–line and position–aware cameras and wrist watches, vehicles with computing and navigation equipment, etc.

These developments pave the way to a range of qualitatively new types of Internet–based services [8]. These types of services, which either make little sense or are of limited interest in the context of fixed–location, desktop computing, include: traffic coordination and management, way–finding, location–aware advertising, integrated information services, e.g., tourist services.

A single generic scenario may be envisioned for these location–based services. Moving service users disclose their positional information to services, which use this and other information to provide specific functionality. To customize the

A Min Tjoa and J. Trujillo (Eds.): DaWaK 2005, LNCS 3589, pp. 275–284, 2005.

interactions between the services and users, data mining techniques can be applied to discover interesting knowledge about the behaviour of users. For example, groups of users can be identified exhibiting similar behaviour. These groups can be characterized based on various attributes of the group members or the requested services. Sequences of service requests can also be analyzed to discover regularities in such sequences. Later these regularities can be exploited to make intelligent predictions about user's future behaviour given the requests the user made in the past. In addition, this knowledge can also be used for delayed modification of the services, and for longer–term strategic decision making [9].

An intuitively easy to understand representation of this knowledge is in terms of rules. A *rule* is an implication of the form $A \Rightarrow B$, where $A$ and $B$ are sets of attributes. The idea of mining association rules and the subproblem of mining frequent itemset was introduced by Agrawal et al. for the analysis of market basket data [1]. Informally, the task of mining frequent itemsets can be defined as finding all sets of items that co–occur in user purchases more than a user–defined number of times. The number of times items in an itemset co-occur in user purchases is defined to be the *support* of the itemset. Once the set of high–support, so called *frequent* itemsets have been identified, the task of mining association rules can be defined as finding disjoint subsets $A$ and $B$ of each frequent itemset such that the conditional probability of items in $B$ given the items in $A$ is higher than a user–defined threshold. The conditional probability of $B$ given $A$ is referred to as the *confidence* of the rule $A \Rightarrow B$. Given that coffee and cream are frequently purchased together, a high–confidence rule might be that "60% of the people who buy coffee also buy cream." Association rule mining is an active research area. For a detailed review the reader is referred to [5].

Spatio–temporal (ST) rules can be either *explicit* or *implicit*. Explicit ST rules have a pronounced ST component. Implicit ST rules encode dependencies between entities that are defined by spatial (north–of, within, close–to,. . . ) and/or temporal (after, before, during,. . . ) predicates. An example of an explicit ST rule is: "Businessmen drink coffee at noon in the pedestrian street district." An example of an implicit ST rule is: "Middle–aged single men often co–occur in space and time with younger women." In this paper, we describe our experiences with ST rule mining in the Danish spatial data mining company, Geomatic.

The task of finding ST rules is challenging because of the high cardinality of the two added dimensions: space and time. Additionally, straight-forward application of association rule mining methods cannot always extract all the interesting knowledge in ST data. For example, consider the previous implicit ST rule example, which extracts knowledge about entities (people) with different attributes (gender, age) that interact in space and time. Such interaction will not be detected when association rule mining is applied in straight-forward manner. This creates a need to explore the special properties of ST data in relation to rule mining, which is the focus of this paper.

The contributions of the paper are as follows. First, a number of real world ST data sets are described, and a taxonomy for ST data is derived. Second, having the taxonomy, the described data sets, and and the desirable LBSes in mind, a general methodology is devised that projects the ST rule mining task to traditional market basket analysis. The proposed method can in many cases efficiently eliminate the above mentioned explosion of the search space, and allows for the discovery of both implicit and explicit ST rules. Third, the projection method is applied to a number of different type of ST data such that traditional association rule mining methods are able to find ST rules which are useful for LBSes. Fourth, as a natural extension to the proposed method, spatio–temporally restricted mining is described, which in some cases allows for further quantitative and qualitative mining improvements. Finally, a number of issues in ST rule mining are identified, which point to possible future research directions.

Despite the abundance of ST data, the number of algorithms that mine such data is small. Since the pioneering work of [2], association rule mining methods were extended to the spatial [3,4,6,11], and later to the temporal dimension [12]. Other than in [13,15], there has been no attempts to handle the combination of the two dimensions. In [15] an efficient depth–first search style algorithm is given to discover ST sequential patterns in weather data. The method does not fully explore the spatial dimension as no spatial component is present in the rules, and no general spatial predicate defines the dependencies between the entities. In [13], a bottom–up, level–wise, and a faster top–down mining algorithm is presented to discover ST periodic patterns in ST trajectories. While the technique can naturally be applied to discover ST event sequences, the patterns found are only within a single event sequence.

The remainder of the paper is organized as follows. Section 2 introduces a number of real world ST data sets, along with a taxonomy of ST data. In Section 3, a general methodology is introduced that projects the ST rule mining task to the traditional market basket analysis or frequent itemset mining task. The proposed problem projection method is also applied to the example data sets such that traditional association rule mining methods are able to discover ST rules for LBSes. Finally, Sections 4 and 5 identify unique issues in ST rule mining, conclude, and point to future work.

## 2 Spatio–temporal Data

Data is obtained by measuring some attributes of an entity/phenomena. When these attributes depend on the place and time the measurements are taken, we refer to it as ST data. Hence such ST measurements not only include the measured attribute values about the entity or phenomena, but also two special attribute values: a location value, *where* the measurement was taken, and a time value, *when* the measurement was taken. Disregarding these attributes, the non–ST rule "Businessmen drink coffee" would result in annoying advertisements sent to businessmen who are in the middle of an important meeting.

***Examples of ST Data Sets.*** The first ST data set comes from the "Space, Time, and Man" (STM) project [14]—a multi–disciplinary project at Aalborg University. In the STM project activities of thousands of individuals are continuously registered through GPS–enabled mobile phones, referred to as mobile terminals. These mobile terminals, integrated with various GIS services, are used to determine close–by services such as shops. Based on this information in certain time intervals the individual is prompted to select from the set of available services, which s/he currently might be using. Upon this selection, answers to subsequent questions can provide a more detailed information about the nature of the used service. Some of the attributes collected include: location and time attributes, demographic user attributes, and attributes about the services used. This data set will be referred to as STM in the following.

The second ST data set is a result of a project carried out by the Greater Copenhagen Development Council (Hovedstadens Udviklings Råd (HUR)). The HUR project involves a number of city busses each equipped with a GPS receiver, a laptop, and infrared sensors for counting the passengers getting on and off at each bus stop. While the busses are running, their GPS positions are continuously sampled to obtain detailed location information. The next big project of HUR will be to employ chip cards as payment for the travel. Each passenger must have an individual chip card that is read when getting on and off the bus. In this way an individual payment dependent on the person and the length of the travel can be obtained. The data recorded from the chip cards can provide valuable passenger information. When analyzed, the data can reveal general travel patterns that can be used for suggesting new and better bus routes. The chip cards also reveal individual travel patterns which can be used to provide a customized LBS that suggests which bus to take, taking capacities and correct delays into account. In the following, the datasets from the first and second projects of HUR will be referred to as HUR1 and HUR2, respectively.

The third ST data set is the publicly available INFATI data set [7], which comes from the intelligent speed adaptation (INtelligent FArtTIlpasning (INFATI)) project conducted by the Traffic Research Group at Aalborg University. This data set records cars moving around in the road network of Aalborg, Denmark over a period of several months. During this period, periodically the location and speeds of the cars are sampled and matched to corresponding speed limits. This data set is interesting, as it captures the movement of private cars on a day–to–day basis, i.e., the daily activity patterns of the drivers. Additional information about the project can be found in [10]. This data set will be referred to as INFATI in the following.

Finally, the last example data set comes from the Danish Meteorology Institute (DMI) and records at fixed time intervals atmospheric measurements like temperature, humidity, and pressure for Denmark for 5 km grid cells. This data set is unique in that unlike the other datasets it does not capture ST characteristics of moving objects, but nonetheless is ST. This data set will be referred to as DMI in the following.

***A Taxonomy of ST Data.*** Data mining in the ST domain is yet largely unexplored. There does not even exist any generally accepted taxonomy of ST data. To analyze such data it is important to establish a taxonomy.

Perhaps the most important criterion for this categorization is whether the measured entities are *mobile* or *immobile.* The ST data in the DMI data set is immobile in the sense that the temperature or the amount of sunshine does not move from one location to the other, but rather, as a continuous phenomenon, changes its attribute value over time at a given location. On the other hand, the observed entities in the other four datasets are rather mobile.

Another important criterion for categorization is whether the attribute values of the measured entities are *static* or *dynamic*. There are many examples of static attributes values but perhaps one that all entities possess is a unique identifier. Dynamic attributes values change over time. This change can be slow and gradual, like in the case of the age of an observed entity, or swift and abrupt, like in the case of an activity performed by the observed entity, which starts at a particular time and last for a well-specified time interval only.

## 3 Spatio–temporal Baskets

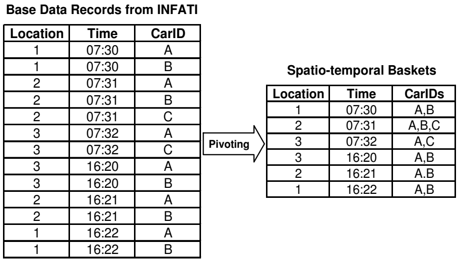
Following the methodology of market basket analysis, to extract ST rules for a given data set, one needs to define ST *items* and *baskets*. This task is important, since any possible knowledge that one can extract using association rule mining methods will be about the possible dependencies of the items within the baskets.

***Mobile Entities with Static and Dynamic Attributes.*** Consider the STM data; it is mobile in nature and has several static and dynamic attributes. Base data contains the identity and some demographic attributes of the user, and the activity performed by user at a particular location and time. Further attributes of the locations where the activity is performed are also available. By applying association rule mining on this base data one can find possible dependencies between the activities of the users, the demographics of the users, the characteristics of the locations there the activities are performed, and the location and time of the activities. Since the location and time attributes are items in the baskets one may find {Strøget,noon,businessman,café} as a frequent itemset and from it the association rule {Strøget,noon,businessman} ⇒ {café}. Strøget being a famous pedestrian street district in central Copenhagen in Denmark, this rule clearly has both a spatial and temporal component and can be used to advertise special deals of a café shop on Strøget to businessmen who are in the area around noon.

In the INFATI data set, a record in the base data contains a location, a time, a driver identifier, and the current speed of the car along with the maximum allowed speed at the particular location. The possible knowledge one can discover by applying association rule mining on the base data is where and when drivers or a particular driver occur(s) and/or speed(s) frequently. However, one may in a sense pivot this table of base data records such that each new row represents an ST region and records the car identifiers that happen to be in that region. Applying association rule mining on these ST baskets one may find which cars co–occur frequently in space and time. Such knowledge can be used to aid intelligent rideshare services. It can also be valuable information for constructing traffic flow models and for discovering travel patterns. While the possible knowledge discovered may be valuable for certain applications, the extracted rules are not clearly ST, i.e.: there is no *explicit* ST component in them. In fact the same set of cars may frequently co–occur at several ST regions which may be scattered in space and time. Nonetheless, it can be argued that since the "co–occurrence" between the items in the ST baskets is actually an ST predicate in itself, the extracted rules are *implicitly* ST.

An alternative to this approach might be to restrict the mining of the ST baskets to larger ST regions. While this may seem useless at first, since the baskets themselves already define more fine–grained ST regions, it has several advantages. First, it allows the attachment of an explicit ST component to each extracted rule. Second, it enhances the quality of the extracted rules. Finally, it significantly speeds up the mining process, as no two itemsets from different regions are combined and tried as a candidate. Figure 1 shows the process of pivoting of some example records abstracted from the INFATI data set. Figure 2 shows the process and results of spatio–temporally restricted and unrestricted mining of the ST baskets. In this example the shown frequent itemsets are based on an absolute minimum support of 2 in both cases, however in the restricted case specifying a relative minimum support would yield more meaningful results. Naturally the adjective "relative" refers to the number of baskets in each of the ST regions. Figure 2 also shows the above mentioned qualitative differences in the result obtained from spatio–temporally restricted vs. unrestricted mining. While the frequent co–occurrence of cars A and B, and cars A and C are detected by unrestricted mining, the information that cars A and B are approximately equally likely to co-occur in area A1 in the morning as in the afternoon, and that cars A and C only co–occur in area A1 in the morning is missed.

Similar pivoting techniques based on other attributes can also reveal interesting information. Consider the data set in HUR2 and the task of finding frequently traveled routes originating from a given ST region. In the HUR2 data set a record is generated every time a user starts and finishes using a transportation service. This record contains the identifier of the user, the transportation line used, and the location and time of the usage. For simplicity assume that a trip is defined to last at most 2 hours. As a first step of the mining, one can retrieve all the



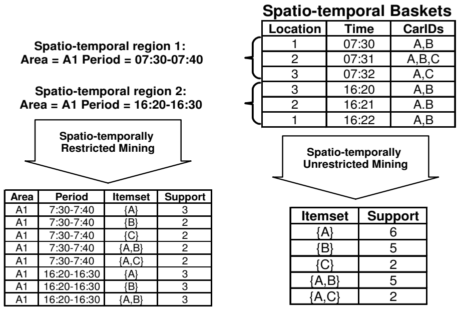
Base Data Records from INFATI

| Location | Time | CarID |
|---|---|---|
| 1 | 07:30 | A |
| 1 | 07:30 | B |
| 2 | 07:31 | A |
| 2 | 07:31 | B |
| 2 | 07:31 | C |
| 3 | 07:32 | A |
| 3 | 07:32 | C |
| 3 | 16:20 | A |
| 3 | 16:20 | B |
| 2 | 16:21 | A |
| 2 | 16:21 | B |
| 1 | 16:22 | A |
| 1 | 16:22 | B |

Spatio-temporal Baskets

| Location | Time | CarIDs |
|---|---|---|
| 1 | 07:30 | A,B |
| 2 | 07:31 | A,B,C |
| 3 | 07:32 | A,C |
| 3 | 16:20 | A,B |
| 2 | 16:21 | A,B |
| 1 | 16:22 | A,B |

**Fig. 1.** Process of pivoting to obtain ST baskets from INFATI base data



Spatio-temporal Baskets

| Location | Time | CarIDs |
|---|---|---|
| 1 | 07:30 | A,B |
| 2 | 07:31 | A,B,C |
| 3 | 07:32 | A,C |
| 3 | 16:20 | A,B |
| 2 | 16:21 | A.B |
| 1 | 16:22 | A,B |

Spatio-temporal region 1:
Area = A1 Period = 07:30-07:40
Spatio-temporal region 2:
Area = A1 Period = 16:20-16:30
Spatio-temporally Restricted Mining
Spatio-temporally Unrestricted Mining

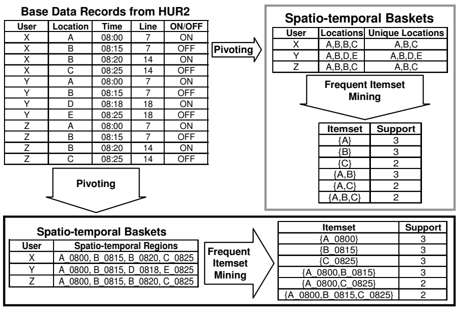
| Area | Period | Itemset | Support |
|---|---|---|---|
| A1 | 7:30-7:40 | {A} | 3 |
| A1 | 7:30-7:40 | {B} | 2 |
| A1 | 7:30-7:40 | {C} | 2 |
| A1 | 7:30-7:40 | {A,B} | 2 |
| A1 | 7:30-7:40 | {A,C} | 2 |
| A1 | 16:20-16:30 | {A} | 3 |
| A1 | 16:20-16:30 | {B} | 3 |
| A1 | 16:20-16:30 | {A,B} | 3 |

| Itemset | Support |
|---|---|
| {A} | 6 |
| {B} | 5 |
| {C} | 2 |
| {A,B} | 5 |
| {A,C} | 2 |

**Fig. 2.** Process and results of spatio–temporally restricted vs. unrestricted mining of ST baskets

records that fall within the ST region of the origin. Following, one can retrieve all the records within 2 hours of the users that belonged to the first set. By pivoting on the user–identifiers, one can derive ST baskets that contain locations where the user generated a record by making use of a transportation service. Applying association rule mining to the so–derived ST baskets one may find frequently traveled routes originating from a specific ST region. The pivoting process for obtaining such ST baskets and the results of mining such baskets is illustrated in a simple example in the light bordered box of Figure 3. Naturally, the frequent itemset mining is only applied to the "Unique Locations" column of the ST baskets. As before the minimum support is set to 2. Considering the spatial relation between the locations one might consider altering the bus routes to better meet customer needs. For example, if locations A and C are close by on the road network, but no bus line exists with a suitable schedule between A and C, then in light of the evidence, i.e., support of A,B,C is 2, such a line can be added. Note that while the discovered frequent location sets do not encode any temporal relation between the locations, one can achieve this by simply placing ST regions into the ST baskets as items. The pivoting process and the results of mining are shown in the dark bordered box of Figure 3. The discovered ST itemsets can help in adjusting timetables of busses to best meet customer needs.

***Immobile Entities with Static and Dynamic Attributes.*** So far the examples considered datasets that are mobile and have either static, dynamic, or both types of attribute values. Now consider an immobile ST data with mostly dynamic attribute values, as the DMI data set. The base data can be viewed as transactions in a relational table with a timestamp, a location identifier and some atmospheric measurements like temperature, humidity, and pressure. Considering the geographical locations A, B, C, and D depicted in Figure 4, we might be interested in trends like, when the temperature in regions A and B is high and the pressure in regions A and C is low, then at the same time the humidity in region D is medium. By applying something similar to the pivoting techniques above, we can extract such information as follows. For each record concatenate the location identifiers with the atmospheric measurements. Then, for each dis-



Base Data Records from HUR2

| User | Location | Time | Line | ON/OFF |
|---|---|---|---|---|
| X | A | 08:00 | 7 | ON |
| X | B | 08:15 | 7 | OFF |
| X | B | 08:20 | 14 | ON |
| X | C | 08:25 | 14 | OFF |
| Y | A | 08:00 | 7 | ON |
| Y | B | 08:15 | 7 | OFF |
| Y | D | 08:18 | 18 | ON |
| Y | E | 08:25 | 18 | OFF |
| Z | A | 08:00 | 7 | ON |
| Z | B | 08:15 | 7 | OFF |
| Z | B | 08:20 | 14 | ON |
| Z | C | 08:25 | 14 | OFF |

Spatio-temporal Baskets

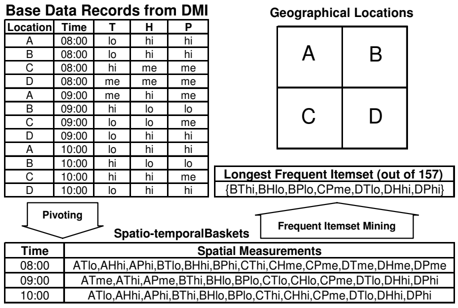
| User | Locations | Unique Locations |
|---|---|---|
| X | A,B,B,C | A,B,C |
| Y | A,B,D,E | A,B,D,E |
| Z | A,B,B,C | A,B,C |

| Itemset | Support |
|---|---|
| {A} | 3 |
| {B} | 3 |
| {C} | 2 |
| {A,B} | 3 |
| {A,C} | 2 |
| {A,B,C} | 2 |

Spatio-temporal Baskets

| User | Spatio-temporal Regions |
|---|---|
| X | A_0800, B_0815, B_0820, C_0825 |
| Y | A_0800, B_0815, D_0818, E_0825 |
| Z | A_0800, B_0815, B_0820, C_0825 |

| Itemset | Support |
|---|---|
| {A_0800} | 3 |
| {B_0815} | 3 |
| {C_0825} | 3 |
| {A_0800,B_0815} | 3 |
| {A_0800,C_0825} | 2 |
| {A_0800,B_0815,C_0825} | 2 |



Base Data Records from DMI

| Location | Time | T | H | P |
|---|---|---|---|---|
| A | 08:00 | lo | hi | hi |
| B | 08:00 | lo | hi | hi |
| C | 08:00 | hi | me | me |
| D | 08:00 | me | me | me |
| A | 09:00 | me | hi | me |
| B | 09:00 | hi | lo | lo |
| C | 09:00 | lo | lo | me |
| D | 09:00 | lo | hi | hi |
| A | 10:00 | lo | hi | hi |
| B | 10:00 | hi | lo | lo |
| C | 10:00 | hi | hi | me |
| D | 10:00 | lo | hi | hi |

| Longest Frequent Itemset (out of 157) |
|---|
| {BThi,BHlo,BPlo,CPme,DTlo,DHhi,DPhi} |

Spatio-temporalBaskets

| Time | Spatial Measurements |
|---|---|
| 08:00 | ATlo,AHhi,APhi,BTlo,BHhi,BPhi,CThi,CHme,CPme,DTme,DHme,DPme |
| 09:00 | ATme,AThi,APme,BThi,BHlo,BPlo,CTlo,CHlo,CPme,DTlo,DHhi,DPhi |
| 10:00 | ATlo,AHhi,APhi,BThi,BHlo,BPlo,CThi,CHhi,CPme,DTlo,DHhi,DPhi |

**Fig. 3.** ST baskets and frequent itemset mining for HUR2

**Fig. 4.** ST baskets and frequent itemset mining of DMI

tinct time interval when measurements are taken, put all concatenated values, each of which is composed of a location identifier and an atmospheric measurement, into a single, long ST basket. By performing association mining on the derived ST baskets one can obtain the desired knowledge.

As an illustrative example, depicted in Figure 4, consider the four neighboring cells A, B, C, and D and the corresponding measurements of temperature (T), humidity (H), and pressure (P) at three different times. Items in the ST baskets are derived by concatenating a location identifier followed by an attribute symbol and an attribute value. Hence, the item 'ATlo' in the ST basket at time '08:00' encodes the fact that at '08:00' at location 'A' the temperature ('T') was low ('lo'). Notice that the extracted knowledge refers to specific locations. If one is interested in obtaining knowledge about the inter–dependencies of these attributes relative (in space) to one another, for each base data record at each distinct time interval when measurements are taken, an ST basket can be constructed that encodes measurements from neighboring cells only. So, for example considering the immediate 8 neighbors of a cell and assuming three different attributes the number of items in each basket is $3 + 8 \times 3 = 27$. Considering a five–by–five relative neighborhood centered around a cell the number of items in each basket is 75, and the number of possible itemsets, given three possible attribute values for each of the attributes is $3^{75} \approx 6.1 \times 10^{34}$. To reduce complexity, top–down and bottom–up mining can occur at different spatial and temporal granularities.

While in the above examples the type of ST data that was analyzed and the type of ST knowledge that was extracted is quite different the underlying problem transformation method—referred to as *pivoting*—is the same. In general, one is given base records with two sets of attributes $A$ and $B$, which are selected by a data mining expert and can contain either spatial, temporal and/or ordinary attributes. Pivoting is then performed by grouping all the base records based on the $A$–attribute values and assigning the $B$–attribute values of base records in the same group to a single basket. Bellow, attributes in $A$ are referred to as *pivoting* attributes or *predicates*, and attributes in $B$ are referred to as *pivoted* attributes or *items*. Depending on the type of the pivoting attributes and the type of the pivoted attributes the obtained baskets can be either *ordinary*, *spatial*, *temporal*, or *ST* baskets. Table 1 shows the different types of baskets as a function of the different types of predicates used to construct the baskets and the different types of items placed in the baskets. The symbols s, t, st, i, and b in the table are used to abbreviate the terms 'spatial', 'temporal', 'spatio–temporal', 'items', and 'baskets' respectively.

In the "co-occurrence" mining task, which was earlier illustrated on the INFATI data, the concept of restricted mining is introduced. This restriction is possible due to a side effect of the pivoting technique. When a particular basket is constructed, the basket is assigned the value of the pivoting attribute as an implicit label. When this implicit basket label contains a spatial, temporal, or ST component, restricting the mining to a particular spatial, temporal, or ST subregion becomes a natural possibility. It is clear that not all basket types can

**Table 1.** Types of baskets as a function of predicate type and item type

| pred/item type | s–i | t–i | st–i | ordinary–i |
|---|---|---|---|---|
| s–predicate | s–b | **st–b** | | s–b |
| t–predicate | **st–b** | t–b | | t–b |
| st–predicate | **st–b** | **st–b** | **st–b** | **st–b** |
| other–predicate | s–b | t–b | **st–b** | ordinary–b |

**Table 2.** Possible mining types of different types of baskets

| basket/mining type | s–r | t–r | st–r | unr |
|---|---|---|---|---|
| s–basket | X | | | X |
| t–basket | | X | | X |
| st–basket | X | X | X | X |
| other–basket | | | | X |

be mined using spatial, temporal, or ST restrictions. Table 2 shows for each basket type the type of restrictions for mining that are possible. The symbols s, t, st, r, and unr in the table are used to abbreviate the terms 'spatial', 'temporal', 'spatio–temporal', 'restricted', and 'unrestricted' respectively.

## 4 Issues in Spatio–temporal Rule Mining

The proposed pivoting method naturally brings up questions about feasibility and efficiency. In cases where the pivoted attributes include spatial and/or temporal components, the number of items in the baskets is expected to be large. Thus, the number and length of frequent itemsets or rules is expected to grow. Bottom–up, level–wise algorithms are expected to suffer from excessive candidate generation, thus top–down mining methods seem more feasible. Furthermore, due to the presence of very long patterns, the extraction of all frequent patterns has limited use for analysis. In such cases closed or maximal frequent itemsets can be mined.

Useful patterns for LBSes are expected to be present only in ST subregions, hence spatio–temporally restricted rule mining will not only make the proposed method computationally more feasible, but will also increase the quality of the result. Finding and merging patterns in close–by ST subregions is also expected to improve efficiency of the proposed method and the quality of results.

Placing concatenated location and time attribute values about individual entities as items into an ST basket allows traditional association rule mining methods to extract ST rules that represent ST event sequences. ST event sequences can have numerous applications, for example an intelligent ridesharing application, which finds common routes for a set of commuters and suggests rideshare possibilities to them. Such an application poses a new requirement on the discovered itemsets, namely, they primarily need to be "long" rather than frequent (only a few people will share a given ride, but preferably for a long distance). This has the following implications and consequences. First, all subsets of frequent and long itemsets are also frequent, but not necessarily long and of interest. Second, due to the low support requirement a traditional association rule mining algorithm, disregarding the length requirement, would explore an excessive number of itemsets, which are frequent but can never be part of a long and frequent itemset. Hence, simply filtering out "short" itemsets after the mining process is inefficient and infeasible. New mining methods are needed that efficiently use the length criterion during the mining process.

## 5 Conclusion and Future Work

Motivated by the need for ST rule mining methods, this paper established a taxonomy for ST data. A general problem transformation method was introduced, called pivoting, which when applied to ST datasets allows traditional association rule mining methods to discover ST rules. Pivoting was applied to a number of ST datasets allowing the extraction of both explicit and implicit ST rules useful for LBSes. Finally, some unique issues in ST rule mining were identified, pointing out possible research directions.

In future work, we will devise and empirically evaluate algorithms for both general and spatio–temporally restricted mining, and more specialized types of mining such as the ridesharing suggestions. Especially, algorithms that take advantage of the above–mentioned "long rather than frequent" property of rideshare rules will be interesting to explore.

## References

1. R. Agrawal, T. Imilienski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proc. of SIGMOD*, pp. 207–216, 1993.
2. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. of VLDB*, pp. 487–499, 1994.
3. M. Ester, H.-P. Kriegel, and J. Sander. Spatial Data Mining: A Database Approach. In Proc. of SSD, pp. 47–66, 1997.
4. M. Ester, A. Frommelt, H.-P. Kriegel, and J. Sander. Algorithms for Characterization and Trend Detection in Spatial Databases. In *Proc. of KDD*, pp. 44–50, 1998.
5. B. Goethals. Survey on frequent pattern mining. Online at: `citeseer.ist.psu.edu/goethals03survey.html`
6. J. Han, K. Koperski, N. Stefanovic. GeoMiner: A System Prototype for Spatial Data Mining. In *Proc. of SIGMOD*, pp. 553–556, 1997.
7. INFATI. The INFATI Project Web Site: `www.infati.dk/uk`
8. C. S. Jensen. Research Challenges in Location-Enabled M-Services. In *Proc. of MDM*, pp. 3–7, 2003.
9. C. S. Jensen, A. Kligys, T. B. Pedersen, and I. Timko. Multidimensional Data Modeling for Location-Based Services. *VLDB Journal*, 13(1):1–21, 2004.
10. C. S. Jensen, H. Lahrmann, S. Pakalnis, and S. Runge. (2004) The INFATI data. Time Center TR-79, `www.cs.aau.dk/TimeCenter`.
11. K. Koperski and J. Han. Discovery of Spatial Association Rules in Geographic Information Databases. In *Proc. of SSD*, pp. 47–66, 1995.
12. Y. Li, X. S. Wang, and S. Jajodia. Discovering Temporal Patterns in Multiple Granularities. In *Proc. of TSDM*, pp. 5–19, 2000.
13. N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, Indexing, and Querying Historical Spatiotemporal Data. In *Proc. of KDD*, pp. 236–245, 2004.
14. STM. Space, Time Man Project Web Site: `www.plan.aau.dk/~hhh/`
15. I. Tsoukatos and D. Gunopulos. Efficient Mining of Spatiotemporal Patterns. In *Proc. of SSTD*, pp. 425–442, 2001.