Filtering Frequent Spatial Patterns with Qualitative Spatial Reasoning

Vania Bogorny [a] Bart Moelans [a] Luis Otavio Alvares [ab]

[a] *Theoretical Computer Science Group, Hasselt University & Transnational University of Limburg, Belgium*
[b] *Instituto de Informática – Universidade Federal do Rio Grande do Sul, Brazil*

The full version of this paper appeared in the 2007 IEEE Workshop Proceedings of the 23rd International Conference on Data Engineering (ICDE2007)[6]

1 Introduction

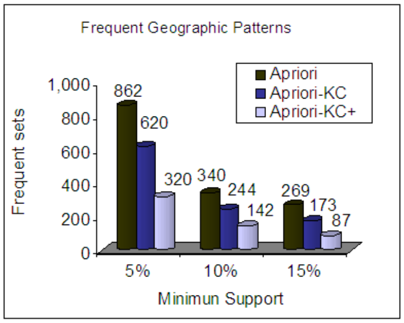
The huge amount of patterns generated by frequent pattern mining algorithms has been extensively addressed in the last few years. Different measures have been proposed to evaluate how interesting association rules are. However, according to [3] it is difficult to come up with a single metric that quantifies the "interestingness" or "goodness" of an association rule. In most approaches, non-interesting rules are eliminated during the rule generation, i.e., a posteriori, when frequent itemsets have already been generated.

In spatial frequent pattern mining the number of non-interesting association rules can increase even further than in transactional pattern mining. Geographic data have semantic dependencies and spatial properties which in many cases are well known and non-interesting for data mining.

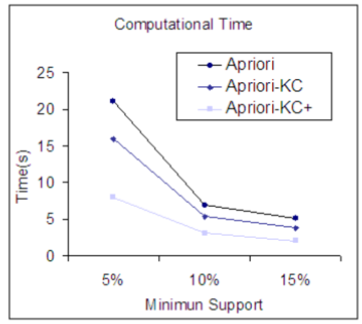
In the geographic domain, not only well known geographic dependencies (e.g., $is\_gasStation \rightarrow touches\_street$) generate a large number of patterns without novel and useful knowledge. Different spatial predicates may contain the same geographic object type. In transactional frequent pattern mining items have binary values, and do either participate or not in a transaction. For instance, the item $milk$ is either present or not in a transaction t. In spatial frequent pattern mining the same spatial object (item) may have different qualitative spatial relationships with the target feature (transaction), and by consequence participate more then once in the same transaction t. For instance, a city C may *contain* an instance i_1 of river, be *crossed by* an instance i_2 of river, or even *touch* an instance i_3 of river. Different spatial relationships with the same geographic object type will generate associations such as $contains\_River \rightarrow touches\_River$ when data are considered at general granularity levels [8]. It is well known that a city does not touch a river because it also contains a river. Such kind of rule is non-interesting for most applications. An interesting association rule would be the combination of any of these two predicates with a different geographic object type or some non-spatial attribute. For instance: $contains\_River \rightarrow WaterPollution = high$ or $touches\_River \rightarrow exportationRate = high$.

In [5], Apriori-KC is proposed. This method made some changes on Apriori [1] to eliminate well known geographic patterns using background knowledge. In [4] this method is extended with an additional step where not only frequent itemsets are reduced, but input space is reduced as much as possible, since this is still the most efficient way for pruning frequent patterns. In [7] the closed frequent pattern mining approach is applied to the geographic domain, eliminating both well known dependencies and redundant frequent itemsets. In this paper we extended the method presented in [5] to reduce the number of non-interesting spatial patterns using qualitative spatial reasoning. Not only the data is considered, but its semantics is taken into account, and frequent patterns that contain the same feature type are removed a priori.

The method proposed in this paper is more effective and efficient then most other methods which perform pruning after the rule generation, since our method explores the anti-monotone constraint of Apriori [1] and prunes non-interesting patterns during the frequent set generation. Other transactional pattern mining



(a) Frequent Geographic Pattern with Apriori, Apriori-KC, and Apriori-KC+



(b) Computational Time to Generate Frequent Geographic Patterns with Apriori, Apriori-KC, and Apriori-KC+

Figure 1: Experiments on real data

approaches such as [2, 9] remove redundant frequent patterns and association rules exploiting support and confidence constraints, while our method is independent of such thresholds. In these approaches "redundant" rules are eliminated by frequent itemset pruning, but "non-interesting" and "meaningless" rules are still generated. In Figure 1(a) it is clear that there is a reduction of the number of frequent patterns, and in Figure 1(b) the efficiency of the method is confirmed.

Acknowledgements

This research has been funded by the Brazilian agency CAPES, the European Union (FP6-IST-FET programme, Project n. FP6-14915, GeoPKDD: Geographic Privacy-Aware Knowledge Discovery and Delivery), and by the Research Foundation Flanders (FWO-Vlaanderen), Research Project G.0344.05.

References

[1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In *VLDB*, pages 487–499, 1994.

[2] Yves Bastide, Nicolas Pasquier, Rafik Taouil, Gerd Stumme, and Lotfi Lakhal. Mining minimal non-redundant association rules using frequent closed itemsets. In *CL '00: Proceedings of the First International Conference on Computational Logic*, pages 972–986, London, UK, 2000. Springer-Verlag.

[3] Roberto J. Bayardo and Rakesh Agrawal. Mining the most interesting rules. In *KDD*, pages 145–154, 1999.

[4] Vania Bogorny, Sandro Camargo, Paulo Engel, and Luis Otavio Alvares. Mining frequent geographic patterns with knowledge constraints. In *ACM-GIS*, pages 139–146. ACM, 2006.

[5] Vania Bogorny, Sandro Camargo, Paulo Engel, and Luis Otavio Alvares. Towards elimination of well known patterns in spatial association rule mining. In *IS*, pages 532–537. IEEE Computer Society, 2006.

[6] Vania Bogorny, Bart Moelans, and Luis Otavio Alvares. Filtering frequent spatial patterns with qualitative spatial reasoning. In *Proceedings of the 22nd International Conference on Data Engineering Workshops, ICDE 2007, 15–16 & 20 April 2007, Istanbul, Turkey*, pages 527–535, 2007.

[7] Vania Bogorny, Joao Valiati, Sandro Camargo, Paulo Engel, Bart Kuijpers, and Luis Otavio Alvares. Mining maximal generalized frequent geographic patterns with knowledge constraints. In *ICDM*, pages 813–817. IEEE Computer Society, 2006.

[8] Jiawei Han. Mining knowledge at multiple concept levels. In *CIKM*, pages 19–24. ACM, 1995.

[9] Mohammed J. Zaki. Generating non-redundant association rules. In *KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 34–43, New York, NY, USA, 2000. ACM Press.