

**Redundancy Reduction for Prevalent Co-location Patterns**

**Abstract:**

Spatial co-location pattern mining is an interesting and important task in spatial data mining which discovers the subsets of spatial features frequently observed together in nearby geographic space. However, the traditional framework of mining prevalent co-location patterns produces numerous redundant co-location patterns, which makes it hard for users to understand or apply. To address this issue, in this paper we study the problem of reducing redundancy in a collection of prevalent co-location patterns by utilizing the spatial distribution information of co-location instances. We first introduce the concept of *semantic distance* between a co-location pattern and its super-patterns, and then define redundant co-locations by introducing the concept of *-covered*, where  (0≤≤1) is a coverage measure. We develop two algorithms RRclosed and RRnull to perform the redundancy reduction for prevalent co-location patterns. The former adopts the *post-mining* framework that is commonly used by existing redundancy reduction techniques, while the latter employs the *mine-and-reduce* framework that pushes redundancy reduction into the co-location mining process. Our performance studies on the synthetic and real-world data sets demonstrate that our method effectively reduces the size of the original collection of closed co-location patterns by about 50%. Furthermore, the RRnull method runs much faster than the related closed co-location pattern mining algorithm.

**Existing System:**

The traditional framework of spatial co-location pat-tern mining uses the frequencies of a set of spatial fea-tures participating in a co-location pattern to measure the prevalence (known as *participation index* [6], or *PI* for short) and requires a user-specified minimum PI threshold to find interesting co-location patterns.

The meaning of PI is that wherever a feature in a co-location pattern *c* is ob-served, all other features in *c* can be observed in its neighborhood with a probability of at least *PI*(*c*). Similar to the support metric in frequent itemset mining, the PI metric satisfies the *anti-monotonicity property*. That is, if a spatial co-location pattern is prevalent with respect to a threshold of PI, then all of its subsets will be discovered as prevalent co-location patterns. Traditional frameworks generate numerous redundant co-location patterns which jeopardize the usability of the technique, as it then de-mands great effort to discern or understand the discov-ered knowledge.

**Proposed System:**

This paper is an attempt to offer answers to these ques-tions. First, we propose a *semantic distance* metric between a co-location and its super-patterns, and show it is a sub-valid distance metric. Second, we define a concept of *-covered* to estimate the redundancy degree of co-locationpatterns. Lastly, we propose two algorithms: RRclosed, which follows existing redundancy reduction techniques to adopt the post-mining framework that reduces redun-dant co-locations from the set of CPCs; RRnull, which employs a *mine-and-reduce* framework to discover non-redundant results directly from the spatial data sets and runs much faster than the CPC mining algorithm, itself a very fast CPC mining method. Our performance study shows that the introduction of *-covered* can effec-tively reduce the number of CPCs.