

Discovering Indirect Associations in Crash Data Through Probe Attributes

Journal [Transportation Research Record: Journal of the Transportation Research Board](#)
Publisher Transportation Research Board of the National Academies
ISSN 0361-1981
Issue [Volume 2083 / 2008](#)
Category Safety and Human Performance
Pages 170-179
DOI 10.3141/2083-20
Online Date Wednesday, January 28, 2009

Authors

Anurag Pande¹, Mohamed Abdel-Aty¹

¹Department of Civil and Environmental Engineering, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816-2450

Abstract

Association analysis is a popular data-mining technique for detecting dependencies in transaction databases. The algorithms for discovery of association rules attempt to find the products that tend to sell together. For this purpose, one can treat crashes as individual transactions and crash characteristics as the "products" part of that transaction to discover patterns of the crash characteristics that tend to coexist. The a priori algorithm to search for association rules employs the lower bound on intersection (i.e., joint) frequency of two crash characteristics to ensure that only significant patterns are discovered. Hence, remarkable associations relating crash characteristics that have joint frequency lower than the specified threshold cannot be identified. The present study proposes to apply probe attributes and indirect associations to overcome this problem for crash data. Two sets of items, X and Y , are indirectly related if there exists another set P (i.e., probe attribute) such that, while sets X and Y rarely or never occur simultaneously, they are highly dependent on the common probe attribute P . Indirect associations may be particularly useful in the context of arterial crash patterns because several products are part of the same original nominal variable and have no intersection (i.e., zero joint frequency) in the market basket database. For example, the products rain and cloudy created from the same original variable weather have zero joint frequency. The probe attributes-based analysis revealed some remarkable arterial crash patterns for the SR-50 corridor (in central Florida), including an association between morning and afternoon peak-period crashes through the site location driveway access. It was also found that young drivers are prone to commit errors on arterial sections that are not level, straight, or both.