



The 7th International Conference on Ambient Systems, Networks and Technologies
(ANT 2016)

Application of Principal Component Analysis for Outlier Detection in Heterogeneous Traffic Data

Pritam Saha^{a*}, Nabanita Roy^a, Deotima Mukherjee^a, Ashoke Kumar Sarkar^b

^aIndian Institute of Engineering Science and Technology, Shibpur, Howrah-711103, India

^bBirla Institute of Technology and Science Pilani, Rajasthan-333031, India

Abstract

Level-of-service (LOS) measures of two-lane highways exhibit incompatibility if the prevailing traffic is heterogeneous in character. Thus, such traffic warrants development of LOS criteria on the basis of compatible measures which capture its characteristics. The present paper has suggested the use of percent speed-reduction and percent slower vehicles, as the measures of performance, while defining LOS criteria. Defining such criteria is basically a classification problem and clustering could be applied as an effective technique for its solution. However, heterogeneity in the traffic mix results in the presence of significant proportion of outliers in the data set, which can distort the results and render into misleading or useless outcomes. The study considers principal component analysis to be an efficient technique in detecting outliers from the data set and accordingly applies it on the proposed LOS measures. An iterative process, adopted for removing outliers, indicates that significant proportion of outliers comprises of non-motorized traffic data; this accordingly ensures reliability of the data set. The study concluded the unfeasibility of LOS assessment of the entire traffic, considering both motorized and non-motorized modes, with respect to a common scale.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Conference Program Chairs

Keywords: Two-lane-highways; level-of-service; performance measures; heterogeneous traffic data; outlier detection; Principal Component Analysis

* Corresponding author. Tel.:91-9831942049.

E-mail address: saha.pritam@gmail.com

1. Introduction

Highway Capacity Manual (HCM)¹ introduced the concept of level-of-service (LOS) for assessing the performance of traffic and provides criteria for it at varied operating conditions. Defining such criteria is basically a classification problem and clustering could be applied as an effective technique for its solution. Clustering is basically an unsupervised approach that classifies the observed data set, under normal and extreme variations, based on certain distance measures. The different vehicular and driver's characteristics in heterogeneous traffic, however, results few observations to considerably differ from the remainders. These are termed as outliers of the dataset, which consequently affect the data clustering. Several researchers have explained the concepts of applying principal component analysis (PCA) as a robust method in detecting these outliers in multivariate settings. The present study, thus, demonstrates the use of PCA as a pre-processing statistical tool to secure a reliable data for clustering in order to define LOS criteria.

The assessment of LOS, on two-lane highways, is a complicated issue because of their unique operational characteristics. HCM provides a method for assessing performance of such highways based on three measures: percent time-spent-following (PTSF), average travel speed (ATS) and percent free flow speed (PFFS). However, a number of researchers have shown that the method is not totally compatible with the highways having heterogeneous traffic with large speed differential; this results in frequent car-following interaction and formation of platoons. Slower vehicles, thus, cause impedance to faster vehicles and compel them to disobey lane discipline and take considerable amount of risk while passing. Keeping this fact in view, the present study identified two major attributes that are responsible for platooning; they are the variation of free-speed characteristics of different types of vehicles: percent speed-reduction (PSR) and, the limiting speed that would differentiate a slower vehicle from other vehicles plying on the highway: percent slower vehicles (PSV).

A real mixed-traffic situation with various kinds of modes, both motorized and non-motorized, is observed on a rural highway, when it approaches a city or a town. The present study, consequently, selected a two-lane highway section, on a National Highway in India, close to a city. The PCA was considered to be a very promising technique in making a perceptible contribution to the detection of outliers from the observed data set and was accordingly applied on the proposed LOS measures. The hypothesis of the study was 'percent speed-reduction should increase with a simultaneous increase of the proportion of slower vehicles in the traffic stream'. Disagreement, however, indicates inconsistency in the observed data points, thus compelling them to be treated as outliers. Therefore, the present study aimed at investigating the statistical plots of PCA to detect such anomalous sample data. The outliers thus detected, would represent those vehicles, the speed of which is insensitive to the presence of slower vehicles in the traffic mix.

2. Review of literature

Assessment of traffic performance on highways is imperative, while taking decisions of investments made at different stages of design life. The second edition of the *Highway Capacity Manual* (HCM)² introduced the concept of level of service (LOS) with the aim of expressing highway performances. The HCM LOS measures, however, have undergone significant changes with the advent of new generation vehicles¹⁻⁴. Introduction of such vehicles considerably affects traffic performance, particularly on two-lane roads, because of both directional movement of traffic; this attributes to the formation of frequent platoons and consequent delay. There have been a number of researchers who suggested use of different LOS measures considering speed, platooning and passing⁵⁻⁷ and also, investigated the relationship of these measures with flow parameters⁸⁻¹⁰. However, studies indicate that the majority of the measures exhibit incompatibility when the traffic is heterogeneous in character and speed differential is quite

Download English Version:

<https://daneshyari.com/en/article/485345>

Download Persian Version:

<https://daneshyari.com/article/485345>

[Daneshyari.com](https://daneshyari.com)