

Data Mining to Improve Traffic Forecasting by Recognizing Anomalies

James Howard
Colorado School of Mines
1500 Illinois St
Golden, Colorado 80401
jahoward@mines.edu

William Hoff
Colorado School of Mines
1500 Illinois St
Golden, Colorado 80401
whoff@mines.edu

ABSTRACT

Accurate traffic forecasting is of great interest for commercial, security, and efficiency applications. By traffic, we mean the movement of vehicles along a road network, the movement of people in a building, or similar data derived from the actions of a group of people. Traditional forecasting methods use statistical models learned from historical data. However, the accuracy of these models fail during the presence of anomalies. In such cases, the forecast can deviate significantly from historical averages. If anomalies can be observed multiple times, however, then a system can be trained to recognize the anomaly when it is happening and generate a more accurate forecast.

In this paper, we present a system that automatically discovers anomalies in time series data, and can recognize their occurrence in subsequent data. We then incorporate these modeled anomalies with common forecasting models to significantly improve short-term forecasting accuracy. We demonstrate improved short term forecasting accuracy on three datasets: a publicly available vehicle traffic dataset, and two building occupancy datasets, derived from a sensor network of motion sensors.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous;
D.2.8 [Software Engineering]: Metrics | *complexity measures, performance measures*

General Terms

Theory

Keywords

ACM proceedings, L^AT_EX, text tagging

1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Knowledge Discovery and Data Mining 2014, New York, New York USA
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

The *proceedings* are the records of a conference. ACM seeks to give these conference by-products a uniform, high-quality appearance. To do this, ACM has some rigid requirements for the format of the proceedings documents: there is a specified format (balanced double columns), a specified set of fonts (Arial or Helvetica and Times Roman) in certain specified sizes (for instance, 9 point for body copy), a specified live area (18 × 23.5 cm [7" × 9.25"]) centered on the page, specified size of margins (1.9 cm [0.75"]) top, (2.54 cm [1"]) bottom and (1.9 cm [0.75"]) left and right; specified column width (8.45 cm [3.33"]) and gutter size (.83 cm [0.33"]).

The good news is, with only a handful of manual settings¹, the L^AT_EX document class `le` handles all of this for you.

The remainder of this document is concerned with showing, in the context of an "actual" document, the L^AT_EX commands specifically available for denoting the structure of a proceedings paper, rather than with giving rigorous descriptions or explanations of such commands.

2. THE BODY OF THE PAPER

Typically, the body of a paper is organized into a hierarchical structure, with numbered or unnumbered headings for sections, subsections, sub-subsections, and even smaller sections. The command `\section` that precedes the text of a section

