Анализ общесетевых закономерностей задержек железнодорожных перевозок с использованием Байесовского сетевого обучения

Mehmet Baran Ulak c, Anil Yazici a, Yun Zhang b

Ключевые слова:

Transit delays - ?

Crowdsourced data -?

Delay pattern identification - ?

Bayesian network learning - ?

Network dependency metrics -?

**Предисловие:**

Задержки в железнодорожном сообщении, как правило, рассматриваются с точки зрения своевременности выполнения рейсов или проблем на отдельных остановках. При таком подходе к масштабированию остановок игнорируется тот факт, что задержки также вызываются и закрепляются факторами, охватывающими всю сеть (например, узкими местами, возникающими из-за совместного использования путей несколькими транзитными линиями). Целью данной статьи является разработка сетевой модели и показателей, которые могут количественно определять зависимость задержек между остановками транзитной сети и выявлять локальные источники проблем в масштабах всей сети. Для этой цели используется байесовское сетевое обучение (на стыке машинного обучения и network science). На основе вычисленных байесовских сетей (BNs) были разработаны сетевые показатели (индуцирующие и восприимчивые) для количественной оценки влияния задержек на остановках в масштабах всей сети. Для реализации предложенной структуры были учтены задержки на Лонг-Айленде. Данные о железнодорожных перевозках (LIRR) были собраны с помощью краудсорсингового приложения OnTime для сбора информации о транзите в режиме реального времени. Разработанная модель BN была протестирована путем перекрестной проверки, дала многообещающие результаты по точности, успешно выявила проблемные остановки на основе отчетов LIRR и предоставила дальнейшее изучение влияния сети. Модель BN и разработанные показатели были дополнительно протестированы с использованием естественного эксперимента, то есть исследования "до" и "после", посвященного недавно завершенному проекту расширения железнодорожных путей в LIRR. Полученные результаты свидетельствуют о том, что BN learning может успешно выявлять сетевые зависимости и указывать железнодорожные соединения/коридоры, которые являются наилучшими кандидатами для последующих инвестиций в улучшение. В целом, разработанные показатели позволяют количественно оценить зависимость задержек между остановками и могут быть использованы разработчиками политики и практиками для принятия решений об инвестициях и улучшениях.

**1. Введение**