摘要

       近数十年来，已经有许多行程时间预测的方法被提出；但这当中大多数集中在高速公路或者动脉网的行程时间预测上。城市网络实时行程时间预测难于实现的几个原因：城市网络的复杂性和路径选择问题，无法获得实时传感器数据，时空数据的覆盖面问题，缺乏对实时事件的考虑。本文提出一种基于知识库的实时行程时间预测模型，该模型在实时和以往的行程时间预测中利用数据挖掘技术从基于位置服务的原始数据中发现其中的交通模式，并将其转化为行程时间预测的预测规则。

1.介绍

       如今，行程时间信息在智能交通系统的多个领域发挥着重要作用，例如：先进的交通管理系统，先进出行信息系统，商用车辆运营和应急管理系统。另外，行程时间预测对于出行者，交通管理者和物流运营商同样是大有裨益的。对于出行者和物流供应商而言，准确的行程时间估计能够避开拥堵路段，减少运输成本提高服务质量。对于交通管理者而言，行程时间信息是交通信息管理的一个重要指标。此外，运用行程时间信息以凝聚交通流量并显著减少平常的交通拥堵，因为公众可能会将各种公共交通作为他们出行的首选。因此，实时行程时间预测是一个具有重要参考意义的交通指标。然而，城市网络的行程时间预测是高度随机和随时间变化的，这是由交通控制设备，突发事件，道路施工以及天气条件等原因所导致的交通运输需求随机起伏所决定的的。换句话说，行程时间受一系列的交通因素影响，包括限速，交通流量，路径选择以及所在的道路和交通设施（例如：交通信号。同样，它同时也受到非交通因素的影响，例如交通事故，天气，道路建设等等。绝大多数前人的研究都是基于交通因素，例如速度、流量和占有率来估计行程时间的，并没有考虑非交通因素。因此，之前的研究方法并不能很好的应用到实施交通环境中。

       Iroy and Kuwahara (2001) 研究发现交通拥堵的减缓程度取决于路网的复杂性。高速公路车流常常被当成是不间断流。城市网络车流则更为复杂，因为车辆在城市网络中不仅有排队延迟也信号延迟和转向延迟。因此城市网络的行程时间预测比高速公路、单一主干道的行程时间预测更具挑战性。此外，还需要解决路径选择等在城市网络行程时间预测中的问题。例如：行程时间预测模型应该决定当给定OD（起点-终点）对时，哪条路径应该作为推荐路径。近几十年来，也已经有许多行程时间预测模型被提出来了，但这些模型大多数是对高速公路行程时间的预测(Chien & Kuchipudi, 2003; Rice & van Zwet,2004; Wu, Ho, & Lee, 2004)。或者仅仅关注的是主干道路网络 (Jiang &Zhang, 2003; Lin, Kulkarni, & Mirchandani, 2004)。

       在过去，许多智能交通的研究和运输机构都使用双回路探测器获得交通数据，它能够获得交通流量（在一定时间内通过探测器的车流量数），速度，道路占有率并且很容易高速公路和城市干道的数据。 (Lin & Zito,2005)。如今，交通数据收集技术已经取得了很大的进步，为了提高交通管理的效率甚至采用了实时收集交通数据。Lin 和 Zito(2005)研究表明，交通信息收集和行程时间度量可以分成三类：基于现场的测量方法，基于探测车载设备的测量方法和基于传感器的测量方法。基于现场的测量方法通过自动车牌识别技术采集车牌字符和到达多个测量点的时间，匹配连续测量点的车牌字符，并通过到达不同测量点的时间估算行程时间。基于探测车的测量方法通过分析探测车队采集的原始数据进行行程时间预测。基于传感器的测量方法通过分析固定传感器例如:回路探测器，转换器即安装在干道上的无线信号标记所采集的原始数据进行行程时间预测。然后每种交通信息采集方法对于行程时间预测都存在缺点和局限性。例如，基于现场和基于传感器的行程时间预测方法都存在空间覆盖率问题，这是因为传感器和车辆自动识别设备都是固定受限的。基于探测车的行程时间预测方法(Chung, Sarvi, Murakami, Horiguchi, &Kuwahara, 2003; Nakata & Takeuchi, 2004; Yang, 2005) 存在成本，时空覆盖问题，这是因为城市交通网络探测车队的维护费用是相当高的。

      在以往的文献中许多行程时间预测的方法是基于历史数据进行分析的，这些方法可以分为如下几类(Lin & Zito, 2005)：回归分析方法（数学模型）(Wu et al., 2004)，时间序列估计方法，数据融合模型(Wen, Lee, &Cho, 2005)已经人工智能方法，例如神经网(Mark, Sadek, & Rizzo, 2004) 。Nakata 和 Takeuchi (2004)运用自回归模型和空间状态模型对时间序列数据进行建模据此预测出行时间 (Chung, 2003; Chung et al., 2003; Lin et al., 2004; Yang,2005)。卡尔曼滤波为许多研究提供了高效的计算方式（递归方式），因为它在几个方面有着非常强大的作用：它能估计过去、现在和未来的状态，甚至可以精确地预测模型所未知的现实。Wu et al. (2004)的研究中心用支持向量回归模型预测高速公路用户的出行时间。Bajwa, Chung, 和Kuwahara (2004)用模式匹配技术进行行程时间预测。交通模式类似于用历史交通模式去预测当前交通状态，用与当前交通情况最接近的历史模式预测现时交通情况。Chung et al. (2003)开发了一个OD估计方法能更准确估计拥堵情况下的交通流和交通量。此外，基于数据融合的出行时间预测模型还进一步整合了灰色理论(Takahashi,Izumi, 2003)和神经网络。Yang(2005) 开发了一种用于高速公路监测数据的数据处理和数据融合混合模型。

      然而，以往大多数研究只考虑了空间网络的静态模型，并且仅仅根据历史数据进行行程时间的预测，因而缺乏对实时事故和交通状态的考虑。换句话说，如果不考虑空间网络中的实时事件会影响预测结果。实时环境下的城市网络行程时间预测难以实现的原因主要有：（1）网络的复杂性，（2）路网中路径选择的问题，（3）实时传感器数据难以得倒或者成本过高，（4）基于传感器和探测车的行程时间预测存在时空数据覆盖面的问题，（5）缺乏事件响应机制而导致低预测准确性。为使行程时间预测系统更实用更准确，本文提出基于知识库的实时行程时间预测模型。利用独立知识库的优势，使得行程时间预测知识能够根据需求的变化动态适应并及时响应外部突发事件。这使得由某领域专家所总结的经验知识和从LBS应用程序中通过模式挖掘出来的知识能够应用到实施环境中。

       提出出行时间预测模型的基本思路是，某条选定的路线的行程时间可以通过估计该路线上的各个路段的行程时间以及延迟来测算，如公式（1）所示。各路段的行程时间可以通过实时预测和历史预测的线性组合来估算。起点（O），终点（D）和行程开始时间（t）是预测公式T(O, D, t)的输入参数。两个子函数Tc和Th分别是基于实时和历史交通数据的预测。相应的，Td 表示路线中经过十字路口的总延迟。两个控制变量α和β分别是实时（Tc）和历史（Th）预测函数的权重。两个预测因子的权重是由该领域的专家所确定的，这一点将在第三部分详述。

（1）

本文的目的是为城市网络提出一种实时行程时间预测模型，通过将实时结果和基于OD对的历史行程时间预测器进行线性组合，以此来预测行程时间。该模型使用LBS的原始数据，并结合地理信息系统将其转换成交通信息。通过集成历史回归数据、实时交通信息和其他实时外部信息来源进行行程时间预测。外部信息来源包括可能影响行程时间预测的实时信息，例如交通事故，道路建设和天气等。此外，由交通领域专家根据外部事件所确定的规则能动态调节历史和实时行程时间预测的组合权值，一次提高预测精度。例如，某OD对路径上当前交通事故的发生会触发一定的专家经验规则，并因此而提高实时信息对行程时间预测的权值。这个模型整合了数据挖掘和专家系统技术，从地理位置服务中挖掘交通模式并将其转换为行程时间预测的推断规则，因此既能处理交通因素问题也能处理非交通因素带来的影响。

本文提出的模型使用LBS应用程序的原始数据作为数据来源，并将安装有LBS应用程序的车辆当做交通浮动车。相较于传统的基于车辆的行程时间预测，这算是一种低成本高效益的方式，因为交通信息只需要从基于LBS的应用程序原始数据中获取。此外，基于地理位置服务的数据有时空覆盖率的优势。交通信息可以通过LBS应用全天候24小时实时动态更改。

下面的章节安排如下：第二章介绍LBS的相关概念，并解释如果从LBS服务中获取交通信息。基于知识库的行程时间预测将在第三章进行详细介绍。在第四章，我们基于台北城市网络实现了行程时间预测系统的原型，并以出租车调度系统作为LBS数据来源。实时数据和历史数据线性组合预测器的比较和评估也在本章进行了阐述。第五章进行研究总结和展望。

2.从LBS中获取交通信息