

# The Application and Research of Big Data Fusion in Inter-City Transportation Integration in Zhujiang-Xijiang Economic Belt

Wei He<sup>1, a</sup>, Ping He<sup>2</sup>, Haiyong Jing<sup>3</sup>

<sup>1</sup>College of Computer Science and Information Engineering, Guangxi Normal University, Guilin 541004, China;

<sup>2</sup>Shangqiu Medical College, Shangqiu, 476000, China;

<sup>3</sup>Shangqiu Central Hospital, Shangqiu, 476000, China.

<sup>a</sup>hhwwboy@163.com

**Abstract.** The opening and development of the Zhujiang-Xijiang Economic Belt still faces many difficulties and challenges. Among them, the imperfections in the connectivity of major infrastructure such as transportation are a major issue that restricts development. Therefore, it is imperative to develop inter-city transportation in the Zhujiang-Xijiang economic belt. This paper studies the data fusion of traffic big data, and provides decision support for the development of inter-city transportation construction through targeted fitting of traffic data and other related data in the Zhujiang-Xijiang economic belt.

**Keywords:** Zhujiang- Xijiang River Economic Belt, Big Data, Transportation Integration.

## 大数据融合在珠江-西江经济带城际交通一体化中的应用与研究

何威<sup>1</sup> 何平<sup>2</sup> 经海永<sup>3</sup>

<sup>1</sup>西师范大学计信学院, 桂林, 541004, 中国;

<sup>2</sup>丘医学高等专科学校, 商丘, 476000, 中国;

<sup>3</sup>丘市中心医院, 商丘, 476000, 中国

**摘要:** 珠江-西江经济带开放发展仍面临诸多困难和挑战, 其中交通等重大基础设施互联互通不完善, 是制约发展的一大问题。所以发展珠江-西江经济带城际交通, 是当务之急。本文对交通大数据数据融合展开研究, 通过对珠江-西江经济带交通数据及其它相关数据的针对性拟合, 为城际交通建设发展等提供决策支持。

**关键词:** 珠江-西江经济带, 大数据, 交通一体化

### 1. 前言

2014年10月13日, 广东和广西自治区为推进珠江三角洲生态交通综合建设和西江生态走廊建设, 签署了珠江西江经济带实施联合行动计划。这表明, 广东和广西将共同推进经济带的综合运输体系。本文正是基于这一大数据时代的趋势, 分析了“西江珠江经济区”城际交通的整合, 并利用各种技术获取有价值的的数据资源进行智能数据分析和建模, 为决策提供依据。

### 2. 国内外研究现状

传统上来说大数据的特点包括: 规模大、种类多、价值密度低、速度快, 这四大特点, 交通大数据亦是如此。如何快速有效的处理海量数据, 去除数据干扰和无效数据, 从中挖掘出有价值的信息, 为城际交通建设提供决策支持, 是本文研究的重点。在当前大数据的背景下, 特别是社交网络的发展, 各种传感器的应用使得交通数据的收集变得异常多样化, 它也使得

数据维度非常高。高维数据不仅增加了数据存储空间和运算时间复杂度，而且导致了数据处理中的维数灾难和其他问题。在机器学习和数据挖掘的分类任务中，经常需要处理大量的高维数据，其中不相关或冗余的属性将严重影响分类性能。因此，在如何在高维数据中进行有效的属性选择是进一步处理前的关键步骤。

鉴于大数据在交通一体化方向的应用，国内外研究和应用有很多方面，如段宗涛，结合数据处理和管理，从多个方面分析了道路交通数据处理与管理的关键技术。通过分析，得出了交通数据的关键问题、道路交通数据研究的技术基础和道路交通数据的可能应用前景，通过这一系列梳理，明确了道路交通大数据研究的关键技术及其相关应用[3]。王孝之等基于海量城际交通数据，采用数据融合算法，结合 Rank Logit 模型，研究了广州与武汉的城际交通模式划分。在此基础上，描述了武汉广州城际交通在高铁运营后的变化，并对可能的未来形势、城市规划、交通规划和运营商进行了敏感性分析。它在营销方式和促销方式上起到了积极的作用。对今后高铁等新的交通方式的运行起到了积极的作用[4]。

国外交通一体化发展的总体思路是以现代信息技术为基础，以交通系统优化为主线，通过多种运输方式实现高效无缝衔接，实现一体化高效协调。充分发挥交通资源优势，充分发挥多种交通方式的优势，满足各地区、各地区的交通需求。促进社会和经济的全面发展。在文献[5]中，作者对纽约郊区城市的流动数据进行了分析，指出其研究成果有助于城市规划师了解城市动态。基于 GPS 数据，文献[6]对车辆的详细轨迹进行跟踪，研究城市居民出行的频繁模式、交通密集区的预测方法和交通拥堵情况。本文研究了智能交通的数据处理体系，其创新之处在于：运用大数据、数据挖掘的相关技术，对城际交通一体化建设提供强大的支持，体现了科学性和先进性。在技术层面，面对数据信息源的多样性和信息量的不断增加，在进行大数据多源多维数据关联分析及知识挖掘时，通过高维大数据数据融合，建立数学模型，形成创新的高维大数据知识发现理论框架和应用体系。

### 3. 研究价值

#### 3.1. 大数据在交通体系的应用价值

随着社会经济的快速发展，城市机动车辆的数量大大增加。在这一背景下，珠江西江经济带的城际交通加速，城市道路系统的平衡已经被打破。传统的交通管理信息系统难以满足交通拥堵等复杂的交通需求。根据长尾理论，大数据聚集海量数据一起聚合离散数据需求，形成数据的长尾，以满足传统需求，如交通需求难以实现。尽管人数不多、份额较少，但聚合效应加长了数据的长尾，这些个体的需求数据会在数据的长尾下形成庞大的需求市场，为满足人群的需求以及个性化服务奠定了数据基础[11]。可见，用大数据管理交通是交通管理模式的一种变革，阻碍传统交通的瓶颈也可通过大数据的分析有效的解决[8]。

主要有以下几点：第一，大数据可以跨行政区域限制。行政区划的划分导致各地方政府利益最大化，难以控制其他行政区域内公共交通基础设施和公交线路的建设。而通过数据的共享可以解决此类问题。二是大数据具有集中化管理的优势。我国大部分城市的交通运输管理主体分散在不同的部门。这种分散造成了公共交通管理的碎片化，如交通信息的分散、单一的信息内容等。而通过集中统一化管理，可形成管理上的统筹。第三，大数据对公共资源的分配更合理。传统的交通部门权责界定未厘清，专业分工的细化也促使公共交通管理部门职能重叠，因而在运营上浪费大量人力、物力。大数据能辅助人们制定出较好的统筹与协调解决方案，在各个交通部门之间合理配置交通职能，针对有关道路问题进行合理信息资源配置。第四，大数据可以打破交通发展的不平衡。大数据科技将技术与理论制定相结合，将信息技术应用于公共交通，从决策层面提高资源利用率，减少对诸如土地等外部资源的依赖。

利用大数据对公共交通进行分析，运输系统的运行效率更高。在大数据中，消耗的计算量随着数据库占用更多的数据而减少。换句话说，在公共交通工具的分配过程中，分配成本会随着大数据的聚集而减少。这种有效的配置可以提高车辆的有效行驶里程，提高运输效率，例如，根据美国洛杉矶研究所的研究，车辆运输减少 46%~84%可以提供相同或更好的运输服务。

另一方面,流量的智能管理可以使静态空闲数据立即智能化,并且面向用户的软件应用程序也可以将大量的数据转换成可理解的图形界面。不仅如此,大数据具有较高的预测能力,而且可以降低假阳性和假阴性的概率。它可以随时监控公共交通的动态特性。例如,在驾驶员离开之前,大数据管理系统将根据在前方道路导致交通拥堵的天气因素来判断备用路线以避免拥堵,并通过智能电话告知驾驶员[1]。

### 3.2 研究目的

在“西江-珠江”经济带城际交通一体化的建设中,利用各种技术获取有价值的资源,进行智能数据分析和建模,提供决策依据。

在设计研究项目时,充分考虑了交通大数据有:规模大、种类多、价值密度低、速度快,这四个大数据的特点。如何快速有效的处理海量数据,去除数据干扰和无效数据,从中挖掘出有价值的信息,为城际交通建设提供决策支持,是本文研究的重点。本文的难点首先在于数据的收集工作,即收集到足够多的真实数据;其次是在数据融合时,如何保证既有效的融合数据,找到数据的关联性,又不丢失重要信息。

结合数据挖掘的技术,使得大数据融合变得十分的可行。数据融合对城市交通一体化,有着非常积极的意义,可分为以下几点:

(1) 发现不同城市真实需求。通过数据融合技术,准确的发掘出不同城市交通数据、经济数据、人口数据等相关大数据中的有用信息,明确交通一体化建设需求。

(2) 通过辅助决策的预测。交通数据分析可以为交通管理、决策、规划和运行、服务和主动安全防范提供更有效的支持。

(3) 快速反应,实现闭环控制。通过物联网、云计算等一系列手段打造智能交通网络,通过数据分析快速指导行动。

(4) 合理分配资源。通过数据融合,针对性的数学建模,以及最优化分析,在成本控制范围内,找到效率与成本平衡的最优点,实现资源的高效利用。

### 3.3. 研究意义

大数据丰富了知识服务的内涵和手段,但也给知识服务带来了新的挑战。由此,构建交通大数据环境下,数据融合的基本框架模型是必要的。数据本身并不能产生价值。只有当存储的数据被深入地分析和执行后,才能创造巨大的价值已经被许多网络公司的经验所证实,采集、存储和深层次分析处理大规模数据集,采用基于统计学的机器学习和数据挖掘的大数据处理技术比传统的建立在关系型数据库平台上的商务智能技术更有效。其减弱了传统挖掘算法的输入数据是完全的假设,研究了含缺失数据的输入数据中缺失与已知数据的特征关联关系,创建了混合核填充和利用不完全信息进行缺失数据填充的理论和技巧。

在“西江-珠江”经济带城际交通一体化的建设中,利用各种技术获取有价值的资源,进行智能数据分析和建模,提供决策依据,是当今大数据时代下科学数据智能发展的一种体现。通过对“西江-珠江经济带”不同城市相关数据的收集,探讨多源/跨模态高维大数据学习理论,在城际交通一体化辅助建设中的应用,拟建立基于高维大数据数据融合的计算机辅助数据挖掘系统,以应用于西江-珠江经济圈城际交通一体化的建设。

## 4. 研究方法

本文拟对交通大数据数据融合展开研究,通过对珠江-西江经济带交通数据及其它相关数据的针对性拟合,为城际交通建设发展等提供决策支持。科学研究思路集中在珠江-西江经济带区域大数据管理的研究,力争提出新的统计、数据挖掘和机器学习理论,有效缓解以上问题,形成新的大数据集集成理论和应用体系,亟待以点为面辐射到全国。

大数据丰富了知识服务的内涵和手段,但也给知识服务带来了新的挑战。数据融合是在大数据环境下进行知识获取、组织和利用的有效途径。由此,构建交通大数据环境下,数据融合的基本框架模型是必要的。数据本身并不能产生价值。只有当存储的数据被深入地分析和执行后,才能创造巨大的价值已经被许多网络公司的经验所证实,采集、存储和深层次分析处



理大规模数据集，采用基于统计学的机器学习和数据挖掘的大数据处理技术比传统的建立在关系型数据库平台上的商务智能技术更有效[8]。

如何快速有效的处理海量数据，去除数据干扰和无效数据，从中挖掘出有价值的信息，为城际交通建设提供决策支持，是本文研究的重点。详细内容为：针对交通大数据的特点，对高维大数据环境下的稀疏学习和凸优化相关的理论、计算方法和实际应用等问题进行系统的探索和研究。具体包括：1) 探索选择性学习的特性以研究高维大数据知识发现的可扩展性和稳定性。即首先利用稀疏学习能导致不同稀疏结构和流形学习能保持数据的局部流形事实，设计信息损失尽可能少的选择性学习（即同时移除噪音样本和降低属性维数）模型，然后通过充分利用数据间的各种的相关性研究自变量和因变量均为高维的大数据知识发现问题。2) 揭示多源/跨源高维大数据的融合规律。利用凸优化方法解决叠加回归模型的优化问题，探讨把多源/跨模态数据投影到一个公共空间且它们的公共部分和私有部分尽可能分离的方法和模型。3) 通过对“西江-珠江经济带”不同城市相关数据的收集，探讨多源/跨模态高维大数据学习理论，在城际交通一体化辅助建设中的应用，拟建立基于高维大数据数据融合的计算机辅助数据挖掘系统，以应用于西江-珠江经济圈城际交通一体化的建设。

针对交通大数据中，数据的属性进行有意义的选择。属性选择的目标是从高维数据集中选择一个更为紧实更准确的子集作为数据的新属性集。所选择的子集具有比原始集合更低的维数，这使得后续的知识发现过程更高效，并且在属性减少之后去除冗余和噪声属性，从而实现更精确的分类。稀疏学习方法是人工智能、应用统计学和视觉认知的研究方向。它也是机器学习领域的最新研究热点之一。在属性选择方法中，成功地应用了 L2 范数的群范数稀疏正则化，并在各种实际应用中表现出良好的性能。本质上，自相似性是普遍存在的，也就是说，对象的一部分类似于它的其他部分。发现具有相似性质的对象可以近似地表示为其自身性质的线性组合，称为自表达。基于对象自相似性的特点，自表达特性适用于大多数高维数据，在机器学习和计算机视觉中有着广泛的应用。该应用采用特征的自表达属性方法，即每个属性具有其他属性的特征，并且更好地利用数据的内部特性来选择属性。基于以上考虑，本项目利用特征的自我表达能力，并结合稀疏学习，以达到属性选择的目的，并在分类上结合传统的支持向量机（SVM）分类方法进行分类，以达到了良好的应用效果[12]。

本文的难点首先在于数据的收集工作，即收集到足够多的真实数据；其次是在数据融合时，如何保证既有效的融合数据，找到数据的关联性，又不丢失重要信息。本文的创新之处在于：利用大数据技术和数据挖掘技术，对城际交通一体化建设提供强大的支持，体现了科学性和先进性。在技术层面，面对数据信息源的多样性和信息量的不断增加，在进行大数据多源多维数据关联分析及知识挖掘时，通过高维大数据数据融合，建立数学模型，形成创新的高维大数据知识发现理论框架和应用体系。

## 5. 结论

本文针对珠江-西江经济带的地域及经济发展不平衡的特点，结合国内外的先进研究成果，提出利用大数据融合的方法来对其城际交通一体化进行研究。创新性的提出了一种属性选择方法，可将其利用在交通数据的分析研究中，可为将来利用数据进行交通规划决策提供理论支持。在以后的研究中，拟使用迁移学习的方法，对珠江-西江经济带发达地区的交通规划和管理进行研究，对相对欠发达地区的交通建设提供数据上的支持。

## References

- [1]. Mulan J. Guangdong and Guangxi promote 'the implementation of the joint action plan for the Zhujiang River Xijiang River economic belt development' plan [EB/OL]. (2016-4-10) [http://www.gxrb.com.cn/html/2014-10/14/content\\_1037977.htm](http://www.gxrb.com.cn/html/2014-10/14/content_1037977.htm)

- [2]. Sun Z, Gu B. Where is the big traffic data from and how to use it [J]. *Research on Urban Rail Transit*, 2014, 17(11).
- [3]. Duan C, Zheng X, Li Y, et al. Road traffic data and its key technologies [J]. *microelectronics and computers*, 2015(6):85-89.
- [4]. Wang X. Intercity transportation mode selection based on Rank Logit and fusion data [D]. Dalian: Dalian University of Technology, 2012.
- [5]. Becker R A, Caceres R, Hanson K, et al. A Tale of One City: Using Cellular Network Data for Urban Planning[J]. *Pervasive Computing IEEE*, 2011, 10(4):18-26.
- [6]. Giannotti F, Nanni M, Pedreschi D, et al. Unveiling the complexity of human mobility by querying and mining massive trajectory data[J]. *Vldb Journal*, 2011, 20(5):695-719.
- [7]. Zhu Z. Research on urban traffic data fusion and traffic situation processing [D]. Shandong University, 2011.
- [8]. Chen M. Application of big data in public transport [J]. *book and intelligence*, 2012 (6): 22-28. Zhou Weigang, Yang Lianghuai, Pan Jian, et al. On the construction of intelligent traffic data processing platform [C]// *China Intelligent Transportation annual meeting*. 2013.
- [9]. Meng X, benevolent. Big data management: concepts, technology and challenges [J]. *computer research and development*, 2013, 50 (1): 146-169.
- [10]. Hitt M A. The Long Tail: Why the Future of Business Is Selling Less of More by Chris Anderson[J]. *Journal of Product Innovation Management*, 2010, 24(3):274-276.
- [11]. He W, Cheng X, Hu R, et al. Feature Self-representation Based Hypergraph Unsupervised Feature Selection via Low-rank Representation[J]. *Neurocomputing*, 2017, 253(C):127-134.