

交通大数据：治愈交通“病”的一剂良药

文/林友芳 编辑/本刊记者 陈佳

林友芳，北京交通大学计算机与信息技术学院副院长。



2015年两会上，“大数据”一词首次写入政府工作报告。大数据正在以势不可挡的趋势广泛渗透到各行各业之中，各行各业也因为与大数据的深入融合，迎来了发展的新契机，实现了更好的发展，也为解决各行业、各领域的现实问题提供了新的思路。

随着经济社会的迅速发展，交通拥堵、交通污染日益严重、交通事故频发等交通问题日益严峻，相较于严峻的形势，民众对于“智能出行”的要求却越来越高。在这样的情况下，大数据被视为解决交通领域诸多问题的一剂良药。

本刊深度对谈北京交通大学计算机与信息技术学院副院长林友芳，请他对交通领域的大数据进行深刻解读，以期破除交通领域长期存在的困境，同时为交通大数据的高效使用提出宝贵的建议。

交通大数据为交通问题的解决提供有效且低成本的支撑

《千人》：交通行业积累了大量的数据，目前，我国交通领域大数据利用和研究现状如何？

林友芳：首先，数据的利用，并不是一个新问题，人类对数据的利用已经有很长的历史。而大数据利用是

因为技术条件成熟，此时，人类有可能借助于机器快速有效地处理远超过以往规模的数据。比如，社会大众查询航班或列车的日志信息，因为数据量过大，存储与处理成本过高，这些日志往往就记录一小部分，或者记录一小段时间作为凭据。如今，随着硬件技术的发展和新型计算模式——云计算的出现，加之基于大规模廉价机群的分布式存储与并行计算架构越来越成熟，大数据处理有了现实可行性。此种现实可行性也给各行各业带来了新机会，交通行业就是其中之一。

目前，各交通行业参与者对交通大数据的利用问题，从思想认识的层面，我认为各交通行业的中高层管理者都已经认识到交通大数据可以为行业内问题的解决提供支撑。在行动方面，国家与交通行业单位也已经实施了不少行动计划，投入资源来存储、处理与利用数据，设立研究或应用项目，尝试着更好地利用数据，有些已经初显成效。

交通大数据相关的学术研究方面，在世界范围内，交通相关的大数据的研究也是热点应用研究问题，也产生了许多学术成果。我认为，交通大数据应用的成功，离不开大学或研究机构，做好产学研用的合作更有利交通大数据的有效利用。

《千人》：相较于其他领域的大数据，交通领域的大数据有何突出特点？

林友芳：交通大数据除了大数据所共同具有的几个“V”（Volume：大量、Velocity：高速、Variety：多样、Value：价值）以外，交通大数据最大的一个特点就是跟民生结合非常紧密。在大数据时代，几乎每个人都可以是交



结合人类的智慧，在虚拟世界中仿真对城市的改造方案，验证方案的有效性，为“城市病”的解决提供有效且低成本的支撑。



通大数据的贡献者。另一个特点就是分布广泛。交通大数据涉及的行业非常多，既包括传统的轨道、民航、道路、水运等交通行业，也涉及与交通有关的其他行业与政府部门，不同类型且大量的交通相关数据分布在不同的从业机构中。但是当下的交通大数据有一个缺点便是综合度不够，因此现在国家或相关部门都在倡导开展综合交通大数据的研究和应用。

《千人》：交通大数据为缓解甚至解决交通拥堵、交通污染、交通事故频繁发生等问题，提供了哪些新思路？

林友芳：应对一个问题，主要采用事前预测预防、事中快速干预和事后的反思调整三大类应对手段。解决交通问题，传统的手段主要根据人的理解、观察和分析，制定一些应对策略去解决存在的问题，有些手段是有效的，有些手段则收效不大，甚至可能起造成阻碍，这是因为没有找准问题产生的根源。找原因与解决问题是两个不可或缺环节。在一个复杂的交通系统中，如何能更准确地找到产生各种问题的原因，原因找准了才可能寻找到更恰当的解决问题的方案。

在当今这个时代，跟交通有关的

数据非常之多，利用大数据可以通过大量机器去学习与挖掘，帮助城市的管理者更准确、更有效地去寻找交通问题产生的各种根源。另一方面，有了大数据，在问题解决的过程中，也可以利用有关整个城市运行历史的大数据，通过虚拟现实技术，在虚拟世界里去仿真城市的活动，尝试提出不同的解决方案，结合人类的智慧，在虚拟世界中仿真对城市的改造方案，验证方案的有效性，为“城市病”的解决提供有效且低成本的支撑。

《千人》：数据可视化是基于大数据的热点应用，国内贵阳等城市正计划构建交通大数据应用平台，实现交通状况可视化，数据可视化将在交通领域将发挥哪些作用？

林友芳：数据可视化确实是大数据处理的热点应用之一。可视化问题就是信息展现问题，是大数据应用平台面向特定场景下的特定用户的信息展现环节。可视化问题本质是面向需要决策的人提供必要的信息的问题，是为人的决策提供必要支撑的一个重要手段。

不管在哪个领域，数据可视化都必须是服务于人们的场景决策。在交通领域也一样，交通状况的可视化主要起到的作用就是交通诱导。对于个人来说，有助于人们寻找一个更合理的出行道路。对整体交通运行来说，有助于通过交通诱导，使路网负载更加合理，缓解交通拥堵问题。

例如，不同形式的交通状况的可视化对于计划出行或正在路途中的人们来说，具有重要的参考意义。现在各个互联网电子地图平台都提供各种形式的交通路况可视化功能，为人们的出行提供决策支撑。许多大城市在一些主要道



路，也提供一些相对当前前进路径的相关方向的道路交通情况，辅助司机进行道路选择。

用好交通大数据，科学性思维举足轻重

《千人》：由您的研究方向出发，您认为胜任交通大数据的研究和应用需要具备哪些专业能力？

林友芳：大数据的研究与应用需要有交叉学科的知识，而某个行业的大数据的研究与应用又涉及具体行业领域的业务知识。

要开展交通大数据的研究，需要的专业能力至少应该包括：（1）对具体交通行业理解能力和经验,对行业应用场景中的应用需求把握的能力，理解客户的能力。（2）对大数据处理技术的把握与实施能力；（3）对数据挖

掘、机器学习、信号处理等理论和技术在交通行业应用的理解与实施能力；

（4）交叉学科知识的应用能力；（5）组织管理协调能力等等。

《千人》：大数据分析表面看是数据分析问题，实则包含一种科学的态度，需要做到足够的严谨，您认为交通大数据收集、整合与分析过程中的科学性如何保证？

林友芳：大数据处理过程和方法的科学性是保证结果准确性和可解释性的前提，也是确保结果能得到人们相信与并得到应用的重要前提。但是，保证大数据处理的科学性又是一个非常具有挑战的问题。

我认为首先是科研与开发人员的态度，做任何学问与工作都需要有科学的态度。严谨是从事科学研究最基本的要求。其次是数据，在数据收集，数

据的真实性、可靠性是最主要，也是科学性的最为基本的保证。

在数据整合过程中，科学性保证的依据主要与整合标准的科学性、整合方法的科学合理性有关。例如，我们在判定两个旅客之间的关系强度时，可以将两个旅客一起同乘飞机的次数当成一个基本依据。但是，对于两个报团参加旅行的陌生人来说，可能很短时间内一次旅行就共同飞了5个航段。但是对某夫妻来说，可能过去两年仅三次一起坐过飞机。这两组人之间的关系强度，到底哪对更强呢？显然，简单整合共同飞行次数用来标定两者关系的强度就可能会带来谬误。然而，在我们研究的方法中，可以通过算法根据不完整历史行程数据智能推断出人们从出发到回家的完整旅程。这样一来，对前面陌生人共同旅程，通过智能算法可以推断出他们只共同出现在大的旅行团中，进行过一次完整的旅游。而后面那对夫妻的行程，

算法可能会推断出两人一起完成了三次出游。通过这样的智能整合，关系强度强弱标定的科学性就更有可能是得到了。有了科学性的保证，用户才有信心去用整合过的数据。

对于大数据处理中的科学性，许多实验都具有大量学者推出各种指标，用于验证新提出的方法的科学性与有效性。比如大数据领域的分类问题，就有许许多多有关样本、过程、结果检验等方面的各种方法、指标与判断标准等。当然，对于大数据领域出现的一些新方法，某些方法从结果上看有一定的效果，部分结果也能得到实验或现实案例的检测，有一定的实用性，但是可能没有严谨的理论或原理能证明其科学性。对于这样的问题，或许未来新的研究者提出的理论或模型能够证明其科学性，或者未来研究可能会证明该方法只适用于特殊场景，并不具有普适性或科学性。

《千人》：有观点认为交通行业，要用好大数据，既不是着急去学习大数据的Hadoop技术，也不是去购置大数据硬件，而是要普及大数据思维，掌握大数据应用的本领。对此，您如何看待？

林友芳：我认为要用好大数据，意识与技术都是重要的，当然思想是更重要的，有了正确的思想认识才会有正确的行动。对大数据思维也没有统一的定义，总体上应该是在企业或组织机构的运作过程要参照历史上所形成的数据，通过大数据技术手段服务于后续业务运作的优化。大数据思维就是要求从业者有充分的通过高效利用大数据完成各项任务意识。

在大数据思维方面，首先要明确

一点：要做好大数据的工作，企业或组织机构的上层领导必须重视。领导不重视，不准确理解这项工作的重要性与必要性，相关工作就很难得到相应的资源与支持。同时，大数据思维还需要强调的一个方面就是让中下层的管理者有这个意识，并有意识地培养发现与规划特定、有意义的大数据应用场景的能力，特别是针对企业业务场景中的痛点或关键问题，尝试采用大数据思维去优化业务或解决问题。大数据在许多场合是有益的，但并非是万能的，也自然不是面面俱到的，一定要去针对企业的痛点问题或关键问题去寻找基于大数据的优化或解决方案。

另外，技术问题是实现方法的问题，仅有思想或想法没有技术和平台或没有选择对正确的技术和平台，也是无法达到效果或实现目的的。从技术角度而言，Hadoop并不是大数据的全部，只是一种重要云计算平台技术而已，大数据技术还涉及机器学习、认知计算、大内存计算、数据与知识工程等领域的理论、理念、技术、方法论或平台工具。

大数据是支撑智能交通系统的素材

《千人》：智能交通得以发展的契机有哪些？如何定义大数据在其中的地位？

林友芳：首先，社会需求是主因，普通民众希望交通服务业相关的系统能更具有智能性；第二是整个社会科学技术特别是信息技术、人工智能技术的发展，为智能交通的发展奠定坚实的基础；第三就是政府层面的重视，例如我国政府就智能交通的发展，推出许多重大的建设与科研项目，促进领域的发

展；最后是目前各个交通行业信息化程度越来越高。这些因素的共同作用，将使交通行业智能化程度得以迅速提高。

大数据在智能交通的发展中有重要作用。一个具有智能的系统要求这样的系统必须具有知识、学习能力和根据知识进行正确的决策并采取相应响应的能力。而大数据是交通系统运行过程的反映，是智能交通系统的学习素材，是系统当前状态和历史状态的反映，是系统知识的重要来源之一。人们也可以根据交通系统的运行大数据，采用大数据分析挖掘的手段，掌握系统的状态，在交通系统的规划、建设、改造、运行、维护和服务提供中，做出更为明智的决策，实现交通系统智能性的目的。

《千人》：大数据在智能交通产业未来的发展中将发挥哪些作用？您如何看待其市场前景？

林友芳：数据在智能交通产业未来将发挥中将发挥重要的作用。人类的智能首先表现为人具有根据材料学习，然后得到知识的能力，其次表现在人类具有根据自己的知识或当前场景中的现状做出决策并采取行动的能力。任何一个交通系统要具有智能性，首先必须具有一定的学习能力，而交通大数据是从不同的角度对现实交通运行状况的记录的结果，这些数据有助于人们加深对交通系统的运行状况的理解。

通过适当的手段，也可以设计智能的算法去从大规模数据中加工出信息，学习到知识，使系统积累的知识越来越多，使人与机器共同组成的系统潜在能做更为有效、更为准确或更为合理的决策。

任何一个民生行业的运转，离不开安全、效率与服务三大主题，交通行

业同样如此，交通大数据的利用，对交通行业最大意义就是更好地确保各领域交通安全有效的运行，并使从业者得到合理的回报，使出行者得到优质的服务。这其中即有宏观问题，也有中观问题，更有大量的微观问题，都可以使用交通大数据的支撑实现优化或得到解决。

可以说，人类的追求是无止境的，而未来的时代，大数据技术使得我们可以更快、更有效地理解我们人类自身的行为与需求，从而促进更智能、更人性化、更安全的服务环境的诞生，随之而来的就是相关行业的发展，交通行业同样如此。

交通大数据有助于交通系统理解人们的出行需求、出行习惯，大数据技术对交通相关产业的促进将会是巨大的，体现于各类交通大数据的生命周期的各类环节。

例如，交通大数据对于各行业后台信息系统的改造的需求，促进信息产业的发展。对数据感知和采集的需求，带来各类新颖有效的感知设备、通信手段、数据采集手段的产业的升级换代。对出行者使用习惯大数据的分析，促进相应的交通工具或部件的优化设计与制造产业的发展等等。

《千人》：大数据和智能交通结合有哪些最新的实际应用？

林友芳：我们有一些新的一些研究进展，例如，利用移动用户位置移动信令数据推导城市主干道路的交通状况的方法，再如，利用基于查询大数据的民航航线需求指数的变化来发现异常波动，如何根据发现的异常去推测未来潜在的社会事件，以及预测未来航线需求的一些尝试。

民航业不能为了大数据而大数据

《千人》：民航大数据是您的研究方向之一，民航安全问题和民航客机延误是民航领域较为突出的两个问题，请问大数据在保证民航安全和解决民航客机延误中如何发挥作用？

林友芳：大数据在保证民航安全上将会发挥重要的作用。例如，在客运运输安全方向，我们提出的基于旅客行为大数据构建的旅客社会网络，其实对潜在安全防控对象的识别和布控就具有支撑作用。再如，在民航信息系统安全运行方面，民航信息中心的安全稳定运行，对民航业务安全运行直接相关，我们正在为中国航信实施基于日志大数据的安全解决方案，该方案未来对数据中心的各类业务就会有很好的支撑作用。

在解决民航客机延误方面，其实大数据也应该可以发挥更多的作用。例如，民航空中运行大数据、全国机场运行状态、资源状态大数据、航空公司航班计划数据、航线环境气象数据、空管数据等各种数据的融合，有助于提高航空公司与机场航班协同调度智能性与效率，从而减少航班延误。这方面我国民航业已经在不断地进行尝试，并取得了一些成效。

同时，利用历史民航各类运行大数据，通过合理的分析挖掘，有利于业务专家分析延误原因，从而采取适当的措施来提升效率。另外，利用民航客运大数据，未来还可以通过虚拟环境来仿真整个复杂系统的运行，通过生成各种可能的虚拟的调度方案来仿真，在模拟真实环境的虚拟世界里对各类方案进行反复预演和评估，来制定更为有效的调度策略、为各类管理改进提供决策支撑。

《千人》：身处大数据时代，有一些企业因为害怕落后被时代淘汰，盲目的投资，为了大数据而大数据，从而浪费企业宝贵的资源。据此，您认为民航业应该如何更好地利用大数据？从哪几个方面发力？

林友芳：投资需谨慎，不能为了大数据而大数据。正如前文所提及的，对企业来说，大数据必须要针对关键问题或痛点，针对关键应用场景，寻找可能可行的大数据支撑方案。

对民航业来说，同样也是这样，要更好地利用好大数据，其实需要解决的问题也不少。民航业是一个很大的行业，本身范围也非常大，涉及的企业与单位非常之多，不同的企业与用户的诉求不同。

大数据思维很重要，不仅得有大数据思维，还得有正确的大数据思维，也得意识到大数据解决不了所有的问题。思想意识层面是一个重要问题，要在企业内宣传、培训与传播正确的大数据思维。另外，要开展好应用需求场景分析，找到行业或企业中痛点或关键问题，寻找或建立与业务运行相关的大数据，研究相关的解决方案，通过技术手段加以实现，仿真实验有效果以后可以部署应用，用来解决实际中的问题。

要利用大数据，其实需要发力的环节可能会很多，每个企业的情况不一样，有可能会包括思维认识宣导、人才引进、队伍培养、外部研究力量引入、产学研结合、典型场景识别、应用产品的规划、投资回报分析、恰当的近期规划、恰当的软硬件平台构建、业务链条优化改进规划设计、以及与大数据处理的贯通、外部数据资源与外协力量的整合等等。