

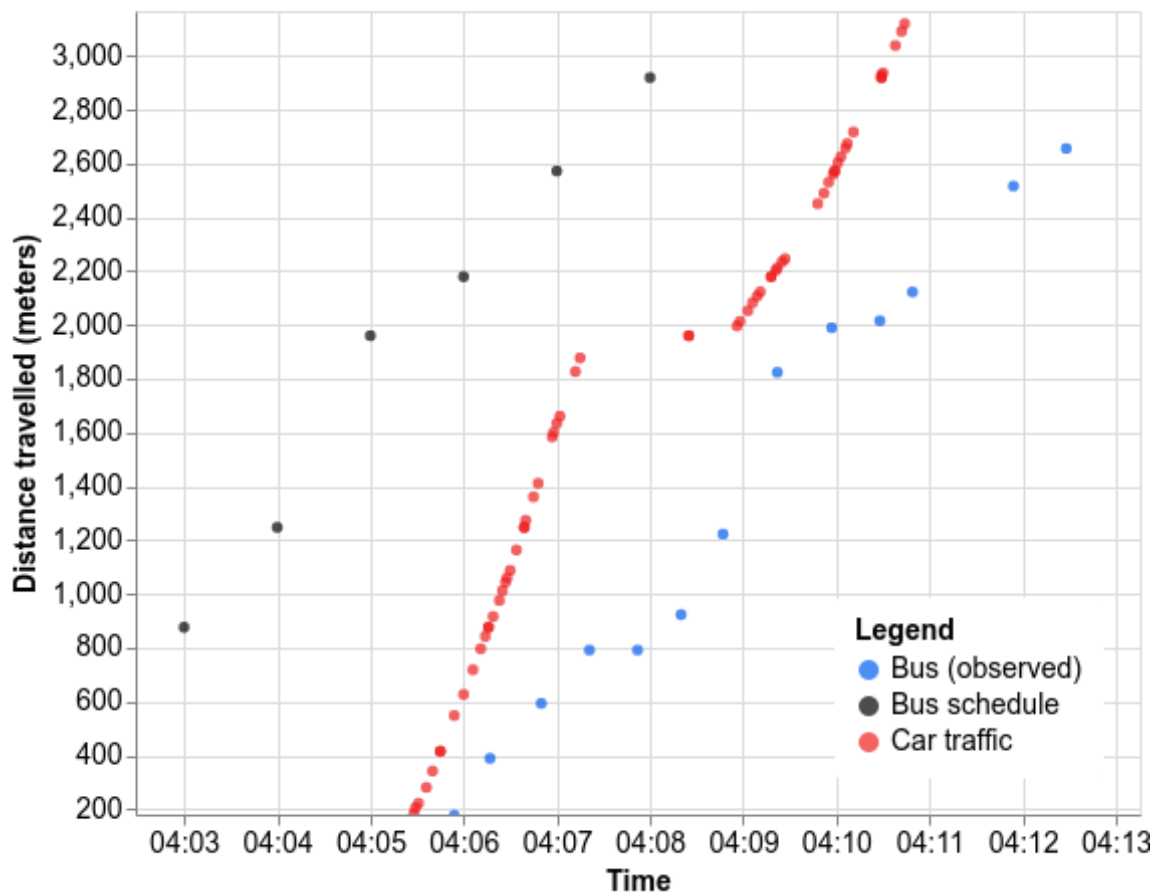
谷歌用机器学习预测公交延误

在这个世界上，每天都有几亿的人依赖公共交通，而在这几亿人中，又有超过半数的人是通过公交车来完成出行的。随着城市节奏不断加快，人们对出行体验的要求也越来越高。尤其对于那些总是被堵在路上的公交乘客而言，他们对延误深恶痛绝，无时无刻不想知道究竟什么时候会产生怎样的延误。不过由于技术和资源的限制，许多机构无法对公交部门提供实时数据，但是谷歌地图（Google Map）却可以做到，通过强大的技术手段对实时交通状况进行提取与预测。

从亚特兰大（Atlanta）到萨格勒布（Zagreb），从伊斯坦布尔（Istanbul）到马尼拉（Manila），谷歌地图已经为全世界数百个城市提供了实时的公交数据，以及预测公交延误的服务。这项功能已经为超过 6 千万的人保障公交出行的准确性。三周前，该公交延误预测系统首次在印度公布，实际上，该系统由机器学习模型驱动，为了更好的预测公交出行时间，该模型对实时路况、公交线网数据、站点数据等多种数据进行整合与处理。

一、模型初始化

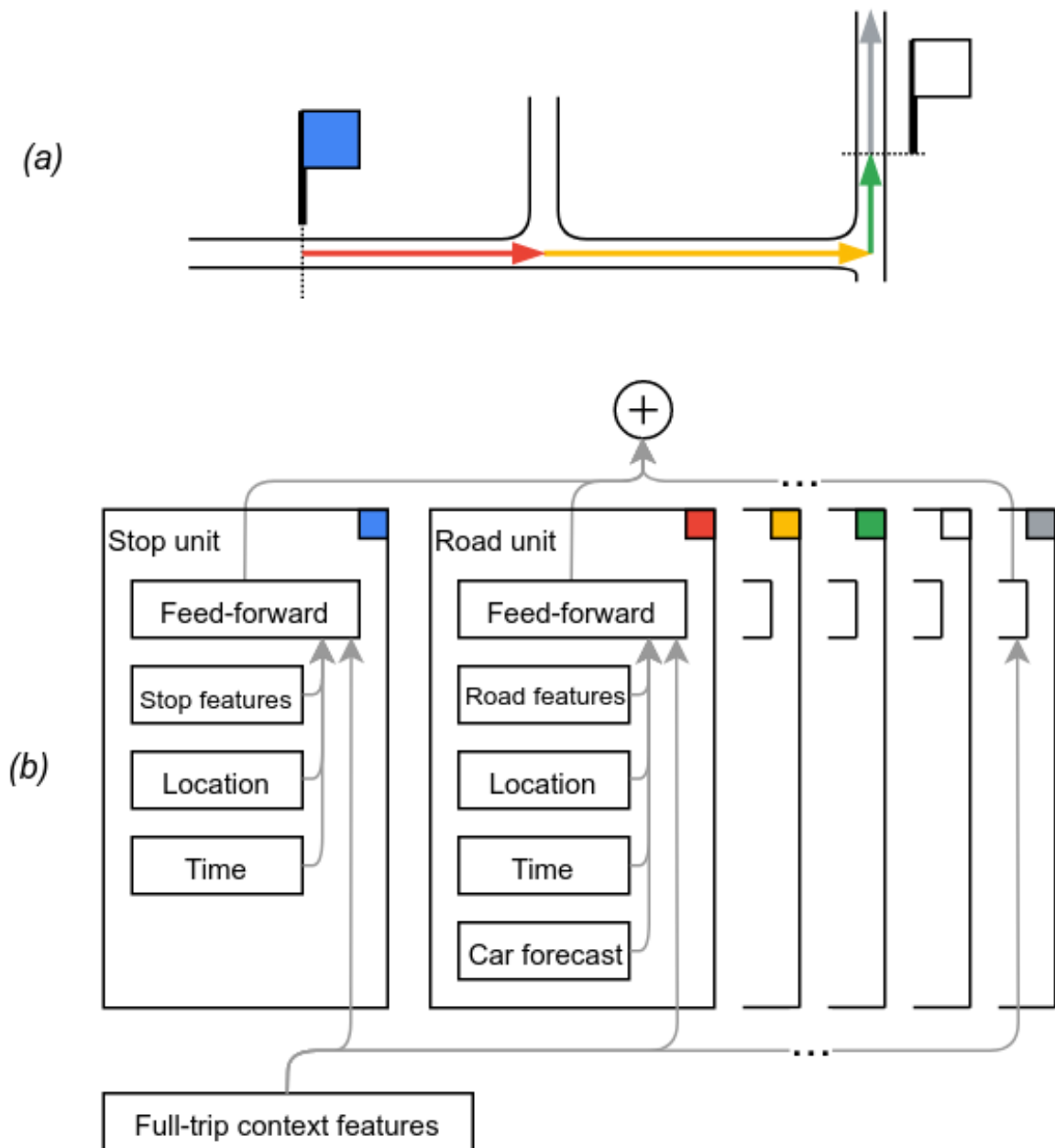
我们从一些受访用户那里了解到，许多城市的交通局并没有实时的预测数据，因此他们采用了一种巧妙的方式来估算公交延误：使用 Google Map 的“开车路线”（Driving directions）功能。然而，公交车的预测要比大型车辆预测困难度多，无法确定的停站上下客时间、是否有专用路权（如公交专用道），甚至公交车的转向和掉头所耗费的时间等等，处处充满了不确定性。就好比如下这个悉尼的案例，如图所示，为某周三下午的公交出行时间距离图，可以看出，公交车辆的实际路径（蓝色）总会比时刻表滞后。而且，也可以看出社会车辆（红色）的行驶状况的确会对公交车辆运行造成影响，例如在 2000 米处社会车辆（红色）的减速行为传递给了 800 米处的公交车辆，从而造成了公交车的长时间停车（低速行驶）。



图一、公交车辆轨迹与路况示意图

为了更好的推广我们的模型，我们把一些从官方渠道获得的实时公交地理信息数据当作训练集，并与同期线路上的路况车速进行校准。该模型被分为一系列的时间轴分段，每个街区或者站点都有相对应的时间轴分段，而每一个分段都有各自的预测持续时间，并与公交车辆的时间轴相对应。由于反馈频率低、公交车速过快或街区车站间距较短等因素，或许会导致一对相邻的观测点间存在多个分段。

就像最近取得重大突破的语音识别、机器翻译一样，我们这样的架构也非常适用于神经网络模型。当然，我们的模型更加但，每一个时间轴分段将会独立的预测其持续时间，并最终将各分段预测的时间求和之后输出。与许多序列模型不同的是，我们的模型既不需要组合输出，也不需要时间轴序列传递状态。序列结构也有很多独特之处，例如，我们可以通过序列模型来训练单个时间轴的持续时间预测模型；还可以优化我们的“线性系统”，在这个系统中，观测到的每个轨迹会将其总持续时间分配给它所经历的多个时间轴。



图二、在蓝色车站开始模拟公交出行 (a), 模型 (b) 将从蓝色车站开始到白色车站为止所经历的所有路段的延误进行求和。

二、在哪建模？

在训练我们的模型时，除了考虑到交通延误以外，我们还考虑了公交线网的细节，以及沿线的交通组织、渠化及信号控制等各类因素。在我们的模型中，即使是一个规模很小的街区，我们也会将路况预测转化到对应路段上的公交车速。下图清晰的将预测路况与公交车速的比值用不同颜色进行展示。越红，则越慢，当然，红色区域内很可能存在公交车站。当我们从街景图中来看图中绿色段得知，该路段设有一条右转的公交专用道。澳大利亚是左侧通行的国家，因此公交车辆的右转是受信号控制的，也比左转更困难一些。倘若一个模型不考虑这些细节问题，那么这个模型结果的可靠性便值得商榷了。



图三、预测结果及部分路段街景图。

为了捕捉街道、社区以及城市的独特属性，我们在模型中对不同大小的区域进行分层学习，用不同的时间轴分段与地理位置（道路或站点的精确位置）来表示模型中嵌入数据的综合。我们首先通过对不同区域的细粒度特征逐步进行严格的惩罚训练，并以此作为特征筛选结果。这就确保了在例如公交驾驶行为的百米影响区，这样的极度复杂的环境下细粒度特征仍可以被计算在内，而在开阔的城郊地区，这种细粒度特征就没有那么重要了。

在训练时，我们还模拟了数据缺失情况下的情况。在每一批训练集中，我们随机抽取一部分样本，并按一定的比例将样本中的某些地理特征信息去除。根据缺失信息的不同，样本大致分为三类，一类是有完整的与实际公交线路和道路情况保持一致的样本，一类仅保留了街区和城市等级的信息，而另一些样本中，几乎没有任何地理特征信息。训练这些不同类别的样本，会使模型的对未知环境的适应

性更强。同时，为了使模型更具有普适性，我们便收集了许多匿名的与公交出行相关的推论结果以扩充我们的训练数据集，这些推论也与谷歌地图中的商务热点时间段、停车位短缺时间段等类似特征属于同一个数据集。然而，公交线路千万条，道路情况千万种，即使我们已经做了这样那样的训练，但我们的模型也一定要进行更大范围的推广。

三、学习当地节奏

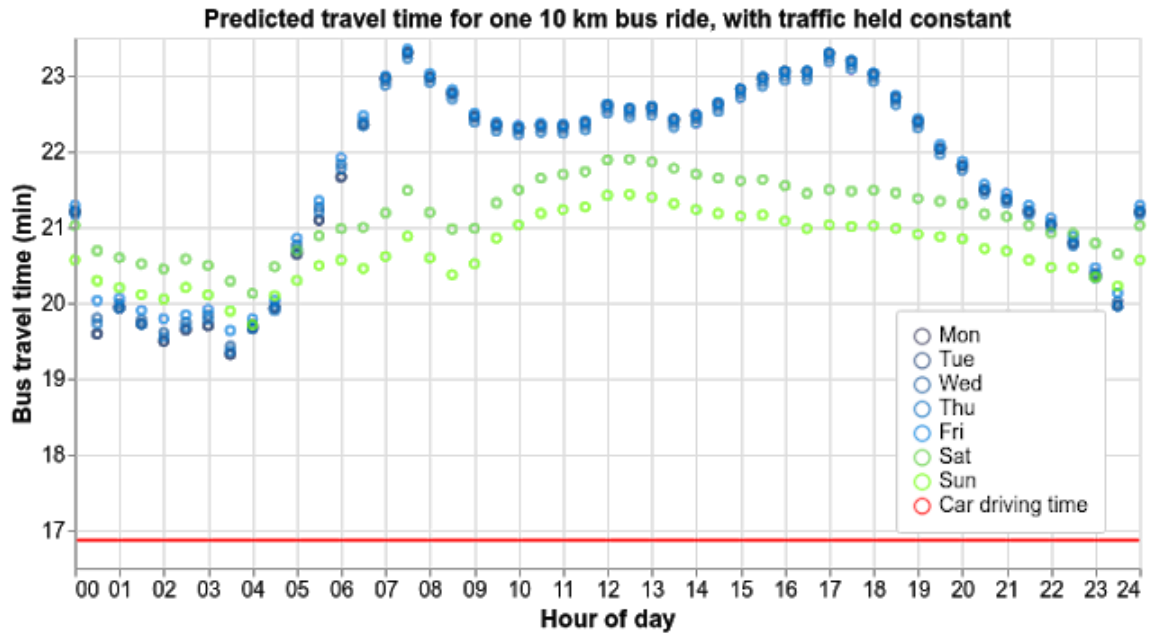
不同的城市，不同的社区，有着不同律动，因此我们将这些位置的表示形式与时间信号结合起来。公交车运行与时间之间有着密切且复杂的依赖性，就好比周二下午 6:30-6:45 之间出现的有趣的差异现象，在这个时间段内，一些地方刚刚经历过晚高峰的喧嚣，一些地方安静的仿佛深夜一般，一些地方却又正是忙时人群熙熙攘攘。我们的模型也学习了一天或一周中不同时间信号的嵌套，将地理特征信息整合起来，以便捕捉那些无法通过路况观测得到的地区性差异，例如高峰期间公交车站的候车人数。

我们嵌入不同的四维向量指定给一天中的不同时间。与大多数神经网络结构不同，四维向量是无法可视化的，因此我们下图中将降维过程示意，让读者更容易理解，理解在三维世界中的时间。时间时具有周期性的，因此我们在模型中将时间进行“循环”设置。但这个循环并不是一个平整时钟表面，是有很多弯曲的，我们的模型也学习这个特点，并让神经元形成了简单的规则，将“午夜”、“傍午”等概念剥离出来，因为这些不会对公交行为产生过大的变化。另一方面，街区和城市两种尺度下的夜间通勤模式差异很大，在每天下午 4 点到 9 点之间，这种现象产生了更复杂的“褶皱”模式，这也使城市高峰时间预测存在更大难度。



图四、我们可以假设四维空间中时间，就好比图中的表盘。在下午4点-9点以及上午7点-9点位置相关的时间窗口更加复杂，而例如凌晨2点-5点等这些特殊的时间段则会更平整许多。（这是艺术家 Will Cassella 的构想，应用 *textures.com* 的纹理以及 *hdrihaven* 的高动态范围图像）

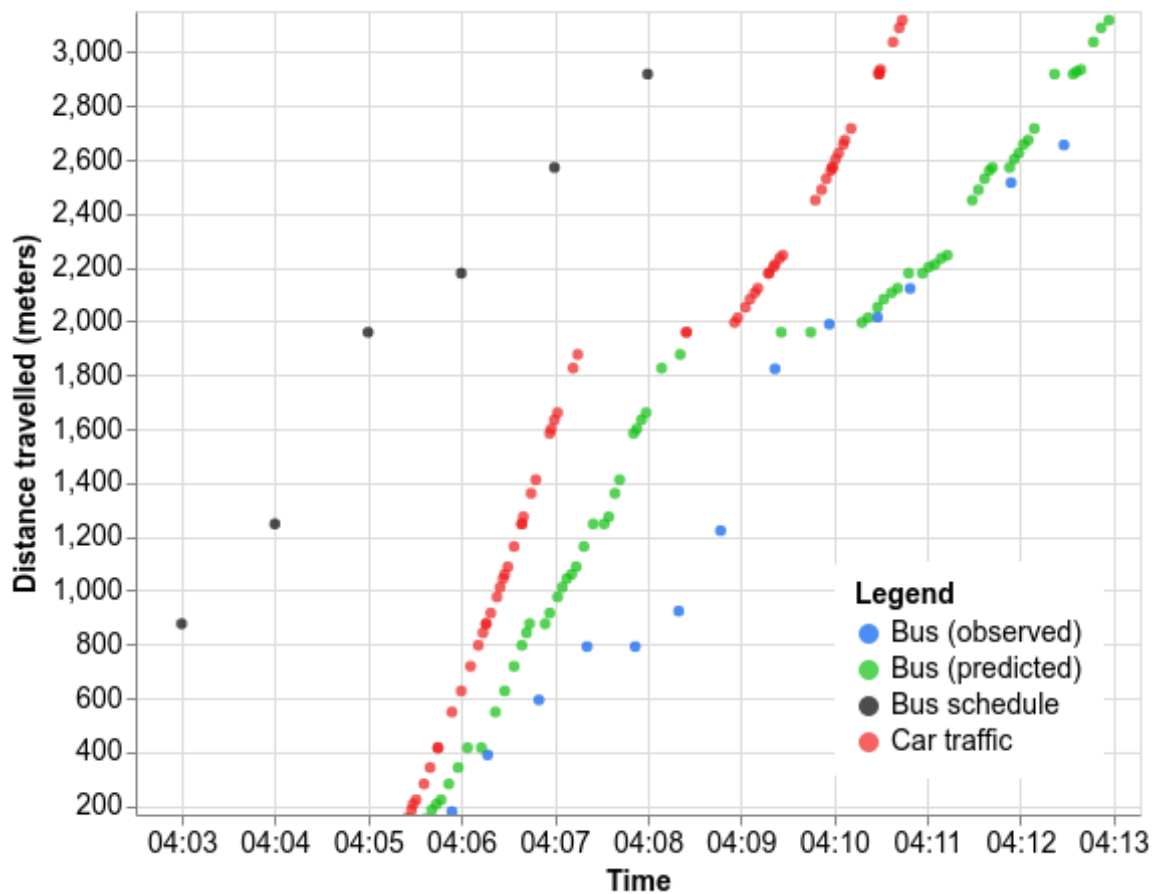
纵使我们一致保持车速不变，只要与这些其他信号一起考虑，这种时间表征仍让我们的预测难上加难。为了更好的解释这个复杂性，我们在下面进行案例演示。这是这张图采集了一辆公交车在一周内横穿 New Jersey（新泽西州）行驶 10 公里的行程时间，这是在午餐时间以及周末的高峰期间进行采集的数据。



图五、相同交通状况下的公交车行程时间预测（10km）

四、整合信息

随着模型的全面训练，让我们简单看一下我们的模型从悉尼公交数据上学到了什么。如果我们根据当天路况来运行模型，将会得到下图所示绿色点的预测结果，然而，这并不能代表一切。我们可以从图中看见，社会车辆在 800 米处发生停车，持续了 10 秒时间，但公交车却至少停了 31 秒。如果我们把目光集中在这公交车实际运行的一分半内，我们可以清楚的发现，公交车实际运行状态与公交时刻表或路况之间会有更多的细微差别。



图六、在有监督的深度学习下，公交预测行程时间与观测时间路况及时刻表之间的关系。

五、 预知行程

到目前为止，在模型中我们还有一件事情没有交代清楚，那就是公交时刻表本身。截止目前为止，在诸多官方部门的针对时刻表的相关实验中，仍然无法显著改善我们的预测结果。在很多城市，复杂多变的交通状况时刻威胁着时刻表的地位与可信度。诚然，我们从数据中可以看到，时刻表并非总那么不准，这可能由于公交部门交通模式进行了深入的研究。

我们将始终以用户体验为主，继续尝试利用更多的时刻表约束以及其他信号来驱动模型运作，以获得更准确的预测结果，为用户出行提供更可靠、更便捷的服务。我们希望在您的旅途中也能对您有所帮助。衷心的祝您旅途愉快！

致谢

这项工作是 James Cook、Alex Fabrikant、Ivan Kuznetsov 和 Fangzhou Xu 在 Google Research 中的共同努力，以及 Anthony Bertuca、Julian Gibbons、Thierry Le Bouleng_、Cayden Meyer、Anatoli Plotnikov 和 Ivan Volosyuk 在谷歌地图中的共同努力结果。我们感谢 Senaka Buthpitiya、Da Cheng Juan、Reuben Kan、Ramesh Nagarajan、Andrew Tomkins 和更大的公交团队对模型嵌入时间的支持和有益的讨论。感谢 Will Cassella 对模型嵌入时间的启发性重构想象。我们也十分感激我们的合作机构能够提供系统训练所需的宝贵的公交数据。

在谷歌地图上随时了解当地的巴士和长途列车时刻表

公共交通是数百外印度人的生命线。目前为止，谷歌地图已为超过 10 亿的用户提供导航服务，不论他们在什么地方，都可以用谷歌地图来探索他们的世界。除了提供基础的导航功能以外，我们还首次为印度当地提供全新的更可靠的功能与出行体验。

今天，我们很高兴能够在印度宣布三项由谷歌地图新开发的功能，这些功能能够让人们的公共交通更加高效，能够让不同的公共交通方式之间换乘更加便利：

- 印度十城的公交出行时间预测；
- 印度铁路列车实时状态信息；
- 多模式公共交通通勤整合提议。

一、 简化公交出行

通过对采集的实时交通数据及时刻表的计算，使得出行时间预测及延误的准确性大大提高，正因如此，谷歌地图可以根据实时的交通状况来告诉您公交的出行时间。

在当前交通状况的条件下预测公交出行时间，这是在印度推出的首个此类产品，在德里、班加罗尔、孟买、海德拉巴、普纳、勒克瑙、钦奈、迈索尔、哥印拜陀和苏拉特（Delhi, Bangalore, Mumbai, Hyderabad, Pune, Lucknow, Chennai, Mysore, Coimbatore, and Surat）同时推出。

当使用此功能时，只需输入起点与终点，而后单击公交选项即可。随后，公交出行时间预测值使用以不同颜色展示出来。（绿色为准点运行，红色为发生延误）

二、 长途列车实时信息

这项功能是与谷歌去年收购的“我的火车去哪了”应用程序（Where is My Train APP）一合作开发，可以通过实时状态的显示，告知人们列车的预计到达时间。这项操作也很简单，搜索您的起点与终点，便可查看可选列车班次，在这个列表中，便可以直接的获得各列车班次的实时状态与延误信息。

三、 与滴滴车相整合的混合出行模式

将滴滴车与其他公共交通方式相结合是一件令人兴奋的事情。在 Android 版的谷歌地图软件中，我们会通过公共交通标签显示出最佳的换乘选择，在什么时候、在何车站去换乘滴滴车，需要耗时多久。同时，还可以看到乘坐滴滴车的预估价格以及公交线路的发车时间。这项功能最初将在德里和班加罗尔（Delhi and Bangalore）进行应用，不久也将推广到其他城市。

在谷歌地图上找到一个座位并及时获得公交信息

原文/ Taylah Hasaballah、Anthony Bertuca

在平静怡和的日子里，最佳的度日方式之一，便是乘坐公交车缓缓绕城一周。这不仅性价比最高的休闲方式，也是也可以让你远离工业社会，静静的坐着读几页自己喜欢的书籍。然而，正当你享受之时，意外的延误与拥挤会瞬间将您悠闲的旅程打乱。

今天，为了让您在旅途中能够保持舒适的出行环境，谷歌地图又推出了两个新功能，来帮助您规划出行。

一、 公交实时延误

重要的会议、朋友的约会、医生的预约，当你面对这样的事情，通常第一件事就是查看公交时刻表，以确保能够按时到达。然而，不幸的是，公交时刻表并不能总是如实反应出当前的路况，从而让你在迟到时会有很多不必要的急迫感产生。

为了解决这一严重问题，谷歌地图推出了实时公交状态反馈功能，这些实时信息我们是无法直接从当地部门获得的。现在，你可以看到你要乘坐的公交车是否会晚点，晚点多久，以及根据现状路况更准确的出行时间。同时，你还可以再地图上看到发生延误的具体位置，这样，你就可以在上车之前预先知道都会发生什么了。

二、 车内拥挤度预测

没有什么比像蒸桑拿一样的坐在火车上被挤的像鱼罐头一样更让人难受的了。因此，我们引入了基于历史数据的公交拥挤程度预测，这样就可以知道你要做的公交车、火车或者地铁的拥挤程度了。在这样的功能帮助下，你完全可以做出一个更明智的决定，挤上车？还是再等一会？说不定你还能够抢到一个座位！