基于墨卡托投影与交互式投票的地图匹配算法

摘要：

关键词

# 引言

## 研究背景及意义

充电桩推荐算法最大限度地减少了的排队时间与充电时间，缓解了EV 续航里程不足问题，尽可能保

持可接受的用户满意度且有助于电网的负载均衡，提升用户的使用体验，并且还能使一些在线服务获取更 高的利润。但目前的推荐算法不能直接适用于电动车充电场景，因为汽车的功能特性（地理位置实时更改） 以及充电桩固有的属性（不存在具体的分类与标签）。这使得目前应用于充电桩的推荐算法存在以下问题：

(1) 不能实时根据地理位置推荐位置合适的且满足用户偏好的充电桩；

(2) 虽然充电桩数量在不断增加，但远不及电动汽车的增速，且大数据显示公共充电桩的利用率很低；

(3) 目前很多类似的推荐算法只能做到离线训练，这导致推荐列表不能及时根据采集到的特征值进行 改变；

针对以上不足，本课题结合了神经网络与协同过滤算法，同时考虑了电动汽车的地理范围广特性以及充电桩无明显分类与标签特性，来增加算法的实时性与精确度，在满足地理位置条件的情况下，减少因电 动汽车充电导致的电网波动问题，尽可能靠近用户的偏好，提升用户交互体验。

## 国内外研究现状

随着电力系统的不断发展，智能电网的概念已经在试点的基础上进行大规模研究，以提高电力系统的

整体效率[1]。可再生能源在电网中的渗透率逐年上升，与此同时，电动汽车（EV）的相关技术日益成熟， EV的市场保有量不断增，进而引发了一系列问题[3]，如大量的EV 集中充电对现有电网的负荷能力提出 了很大的挑战，这会导致电网中无序的电力波动[4]；大量EV 在每天充电高峰期充电不便，且容易发生冲 突[5]；虽然充电桩的数量在逐渐增长，但大数据显示公共充电桩的利用率不足 15%[6]。为解决上述问题， 一些商业公司提出了分时定价策略来缓解充电压力，但依然不能从根本上解决问题[7]。为此，推荐算法逐 渐应用到了充电桩的推荐中，通过机器学习的方式记录用户习惯以及结合车辆地理位置与电网需求，充电 桩推荐算法可以为用户提供更加精准、更加符合用户习惯充电服务，且有效缓解充电桩利用率不足、电网 波动大等问题[8][9][10]。 国内充电设施运营平台主要有E 充网、爱充网等充电服务平台，利用互联网模式，通过自建充电桩或

与合作建充电桩，与电网需求响应平台数据相结合，细化各区域充电桩供电时段与价格，构造电动汽车充 电服务运营网络[11][12]；国外典型的充电服务系统有德国的莱茵集团和美国的 ChargePoint 等，甚至起步 比国内早，美国的ChargePoint 充电服务已处于相对完善的阶段，几乎覆盖了全美国的 90%的公共充电设 施[13]。

如何帮助电动汽车用户（EVU）快速、安全的充电又不影响充电桩的使用便捷性得到了业界广泛关注

[14]。由于充电桩不同于音乐、电影和图书，没有充电桩的具体分类以及直接评分数据[15]，只能根据用 户与充电桩的历史交互信息进行分析，所以基于内容的推荐算法不适用充电桩的推荐，因此，业界大部分基于协同过滤算法进行研究[16]。LiYujing 等人[17]提出了结合充电站的历史运营数据，采用深度学习方 法对到达充电站的车辆数量进行实时预测优化，并结合用户的历史充电数据，采用协同过滤算法得到用户 推荐列表。同时作者还综合考虑用户选择偏好、行驶距离和时间，为用户提供个性化的收费导航服务。 YingweiZhao 等人[18]针对目前数据利用率低、各因素联动程度低、多种收费引导策略计算能力不足的现 状，提出了一种基于协同过滤的引导式智能收费推荐方法。通过计算边缘推荐，可以准确高效地实现该方 案，并实现个性化。卜凡鹏[19]等提出一种基于协同过滤的充电桩推荐算法，通过电动汽车用户对充电桩 的交互行为进行分析，并与电网需求引导平台数据相结合对充电桩的供电时间进行分时定价，建立充电桩 推荐模型，来引导用户使用峰谷电价规避用电高峰期，同时提高充电桩的利用率。

基于协同过滤算法进行研究[16]。LiYujing 等人[17]提出了结合充电站的历史运营数据，采用深度学习方 法对到达充电站的车辆数量进行实时预测优化，并结合用户的历史充电数据，采用协同过滤算法得到用户 推荐列表。同时作者还综合考虑用户选择偏好、行驶距离和时间，为用户提供个性化的收费导航服务。 YingweiZhao 等人[18]针对目前数据利用率低、各因素联动程度低、多种收费引导策略计算能力不足的现 状，提出了一种基于协同过滤的引导式智能收费推荐方法。通过计算边缘推荐，可以准确高效地实现该方 案，并实现个性化。卜凡鹏[19]等提出一种基于协同过滤的充电桩推荐算法，通过电动汽车用户对充电桩 的交互行为进行分析，并与电网需求引导平台数据相结合对充电桩的供电时间进行分时定价，建立充电桩 推荐模型，来引导用户使用峰谷电价规避用电高峰期，同时提高充电桩的利用率。

# 相关工作

# 地图匹配算法MIVMM

随着GPS设备的广泛应用以及网约车的兴起，丰富的GPS轨迹数据现已成为智慧交通系统以及其相关服务的重要数据源。但GPS信号易受到电磁干扰或功率限制的影响，产生大量噪声数据，且目前搭载在车辆上的GPS设备大部分为低采样率设备（采样间隔大于60秒），低采样率设备产生的数据中连续采样点之间的联系较低，且存在噪声数据的影响。因此，在计算汽车行驶时间的估计时不能直接使用原始数据，需要进行地图匹配算法筛选出正确数据。地图匹配的目的是将原始GPS采样点与路网中的道路进行匹配，过滤错误数据，得到准确的路网数据，进而实现准确度较高的路网相关分析。

## 相关定义

定义1：路网（Road Network）：将现实道路网建模为一个有向图，路网例子如图1，其中节点表示路段的集合，边表示路口（在中的连通性）的集合，即如果车辆可以从道路直接行驶到道。

定义2：道路（Road）：道路是路网中最基本的组成单元之一，通常被描述为有向边，但在本文中被描述为一个节点，每条道路都有其固有属性，如车道数、限速、长度等。

定义3：路口（Intersection）：路口是路网中另一个基本组成单元之一，在本文中它被描述为路网中的边，用来表示道路间的关系，每个节点连接着多条路，且存在固有属性，如经纬度、连接道路数等。

定义4：路段（Segment）：大部分道路都是弯曲的，他们由多个路段（Segment）组成，每个路段是两个点连接的直线，在路网匹配过程中将道路划分为多个路段能够提高匹配的精准度。

定义5：路径（Path）路径是一个道路序列，，，其中为采样点个数，和分别表示路段的起点和终点。

定义6：GPS轨迹（GPS Trajectory）：GPS轨迹是由GPS采样点形成的一组序列，其中为GPS采样点个数，每个采样点用经纬度表示，每条GPS轨迹都有其时间戳、采样频率（连续采样点的时间间隔）等信息。

定义7：局部最优路径（Local Optimal Path）：对于在GPS轨迹中的任一采样点的每个候选点，，其中为的候选点个数，假设通过候选点的最优路径是GPS轨迹的真实路径。

## MIVMM算法模型

目前大多数车辆配备的GPS设备所记录的数据包括经纬度、时间戳以及车辆行驶方向与瞬时速度等。但由于存在外部信号干扰或其他因素干扰，会导致GPS设备所记录的信息产生大量噪声，尤其瞬时方向与瞬时速度并不能十分准确的反应真实情况，在我们所提出的地图匹配算法MIVMM中不会直接使用这些数据，而是首先对数据进行初步筛选，过滤掉其中十分确定为噪声的数据，然后再通过提取数据与数据间的关系信息，进行地图匹配过程。

图

具体的，所提出的基于墨卡托投影与交互式投票的地图匹配算法MIVMM包括以下五个阶段：数据预处理、候选点的匹配、位置上下文分析、相互影响建模以及交互式投票。图显示了MIVMM算法的整体框架。其中第一阶段的数据预处理使用卡尔曼滤波法对原始数据进行粗略过滤，筛选掉数据中大部分的错误数据；第二阶段的候选点匹配使用基于KD树以及墨卡托投影的KNN算法进行全范围的查询，获取第一阶段筛选出的采样点的候选点；第三阶段的位置上下文分析中我们利用数据中位置信息与路网数据以及约束分析为下一阶段提供基础凭证与数据；第四阶段的相互影响建模包括静态评分矩阵以及距离评分矩阵两部分的综合计算，然后得到加权评分矩阵，再根据加权评分矩阵得到每个候选点的局部最优路径，构建由候选点以及各个候选点的相关数据组成的候选图；第五阶段的交互式投票根据第三阶段构建的候选图以及第四阶段产生的局部最优路径数据进行投票，与传统方法不同的是我们采用对每相邻采样点之间的边进行投票，相邻采样点之间仅选出一个评分最高的候选边作为最终匹配边，进而得到最终的最优匹配路径。

## 卡尔曼滤波

## 候选点匹配

如何快速准确的确定每个采样点的候选点对于地图匹配是至关重要的，对于每个GPS采样点，，都有候选点，，其中为采样点个数，为候选点个数。本文使用KNN匹配算法进行候选路段的选择。在选出候选路段后进而选择对应的候选点。但原始KNN匹配算法使用的是欧几里得距离，但GPS设备一般记录的是球面坐标系，如WGS84坐标系。所以要对原始KNN算法做出改变。

墨卡托投影（Mercator）是角度不变的投影方式，可以实现将球面坐标系转换为平面坐标系，又称为等角正切圆柱投影。假设圆柱与地球赤道相切，取本初子午线与赤道的交点的投影为坐标原点，赤道的投影为轴，本初子午线的投影为坐标系的轴，这样构成的平面直角坐标系称为墨卡托平面直角坐标系。取地球椭球体的长轴为，短轴为，根据等角条件，墨卡托投影公式为：

其中为经度，东经取正值，西经取负值。为纬度，北纬取正值，南纬取负值。为地球椭球体第一偏心率。由上述可知，墨卡托投影将经纬度坐标映射为平面坐标系。

墨卡托投影的等角特性保证了对象的形状不会变形，同时保证了方向和相互位置的准确性。Web墨卡托投影与常规墨卡托投影有一个重要区别：Web墨卡托投影把地球假设为球体而不是椭球体。这种假设主要是为了计算简单和实现方便。理论精度差别在0.33%之内。当比例尺很大，地物很详细时，其差别可以忽略不计。

考虑把地球椭球体假设为正球体，取地球半径，有，第一偏心率。因此上述墨卡托投影公式可简化为：

由此即可将经纬度坐标转换为KNN算法能够使用的平面坐标。

为了快速的得到候选路段，KNN算法基于KD树的k近邻算法来获取候选路段，将静态路网数据建模为KD树，然后通过KD树找出距离采样点最近的候选路段。KD树例子如下图：

KD-Tree图

KD-Tree例子说明

KNN算法使用采样点与候选路段之间的垂直距离以及车辆方向与候选路段方向之间的夹角来确定候选路段，其公式为：

其中，为时间表，与表示权重，表示车辆采样点与候选路段之间的垂直距离的贡献度：

其中为车辆位置，为车辆在候选路段上的投影位置，为位置误差的标准差。车辆行驶方向与候选路段间的方向夹角贡献度：

其中表示车辆的行驶方向，表示候选路段的方向，表示车辆方向与候选路段方向的夹角。车辆位置与候选路段之间的垂直距离使用海伦公式计算，其中分别为三角形的半周长以及3条边长。

在选出候选路段后，还需要确定候选路段上的候选点。车辆采样点与候选路段之间存在两种关系：一是呈钝角三角形关系，如下图(a)所示，这种情况候选点只能选择距离采样点最近的候选路段的端点；另一种关系是呈直角或者锐角三角形关系，如下图(b)这种关系候选点在候选路段的端点之间。

图(a,b)

假设候选路段的两个端点分别为与，在平面坐标系中的直线方程为 。首先计算垂直于此候选路段且过采样点的直线方程，再计算候选路段与此垂线的交点，判断此交点是否满足以及，若不满足则为第一种情况，即候选路段与采样点之间呈钝角三角形关系，候选点为距离采样点最近的端点；若满足则为第二种情况，即候选路段与采样点之间呈直角或锐角三角形关系，候选点为采样点在候选路段上的垂点。

## 位置上下文分析

GPS点的测量误差满足高斯分布，其观测概率为：。其中是采样点的候选点，是从候选到采样点的欧几里德距离。从候选点到候选点的空间分析函数定义为：

其中是过渡概率。目的是测量两个连续候选点之间的最短路径和直路径的相似性。定义为：

其中是采样点到采样点的欧几里得距离，是从候选到的最短路径的长度。在没有交通事故的情况下，车速一般接近道路限速。因此，适当的时间分析函数应在道路限速附近获得最大值。时间分析函数定义为：

其中是从候选点到的最短路径的加权速度限制，是沿着候选点至之间最短路径行驶的车辆的平均速度。

在地图中，道路通常被分为几个级别，每个级别都有不同的道路限速。高架路的水平高于地面道路，即高架路的限速大于地面道路。因此，我们使用道路水平因子来模拟车辆在高限速道路上停留的趋势。的道路水平系数（RLF）定义为：

其中和表示和所在道路的速度限制。综上，路径的综合权重函数为：

位置上下文的分析结合路网拓扑结构，可以实现对3.4节中生成的候选点进行筛选，减少后续的计算负担。我们从三个方面对所有候选点进行分析：首先，在真实的路网结构中，两个连续采样点的候选点之间可能没有可达道路；其次，根据大量实验结论，错误匹配路径权重值总是小于，而正确匹配路径的权重值总是大于；最后，路网结构不是平面的，可能有立交桥或城市环岛等立体结构，而投影的结果是在平面坐标系中，因此，可能会将真正属于被遮挡道路的候选点投影到了其上面的道路上。结合以上三点分析，为了过滤错误数据减少后续计算负担，我们提出了以下三个约束：

约束1：相邻采样点的各个候选点之间的连通性需符合真实路网结构，可在路网结构中查询到最短路径，即匹配路径在路网结构中真实存在且能够在合理时间内可到达。

约束2：设两个相邻匹配点的候选点之间的权重的阈值。当路径的权重小于时则被视为错误匹配路径，此路径不参与后续计算。

约束3：可以根据不通道路拥有不通的限速这一特性，通过车辆的行驶速度与道路限速的关系进行判断。设为车辆沿着候选点到候选点之间最短路径行驶的平均速度，为所经过的所有道路的加权平均限速。当，时，其中，路径被视为错误的匹配路径。

约束1保证了后续参与运算的数据都符合实际路网结构，不会出现不存在的路径。约束2保证了不会出现下述情况：如果错误的候选路径的累积权重大于前几个正确候选路径的权重，即便在后续候选中权值逐渐减小，但最终累积权重有概率大于正确路径的权重，导致生成错误的结果。约束3由于考虑了路网的三维结构，通过限速与车辆行驶速度的关系提高了匹配候选路径的准确性。以上三点约束能够在参与后续计算之前过滤掉大部分的错误路径，减少运算时间。

## 候选点相互影响建模

连续的GPS采样点之间会相互影响，如下图所示，从到…….。显然，越是靠近该点，对该点的影响就越大。我们以上一节提出的位置上下文分析为建模基础，采用空间分析以及时间分析两个方面来衡量两个连续候选点之间的影响。在计算各个采样点对应的各个候选点之间相互影响的权重之前，我们构建了一个静态评分矩阵：

其中，此静态评分矩阵中的每个项目仅考虑两个连续点的信息，表示候选点的权重。并不反映它们之间的相互影响。因此，我们可以将其构建为静态矩阵，以减少计算相互影响的权重时的计算负担。我们将采样点间的相互影响建模为维距离权重矩阵，定义如下：

其中，和是和之间的欧几里德距离，是相对于道路网的参数。此矩阵给出了所有其他点到的距离影响的权重。对于受距离影响的权重可计算为：

在得到每个采样点的权重矩阵后，我们根据上一节筛选出的候选点创建候候选图，其中表示各个采样点的候选点集合，表示前一个采样点的候选点到当前采样点的候选点之间的路径集合，通过约束1可知只有当到存在实际可到达路径时才会存在边。图中的每个节点都拥有观测概率属性，即表示该候选点与其采样点的观测概率，每条边都拥有累积权重与投票数属性，用于后续投票计数。

图

## 交互式投票

局部最优路径可以反映最终匹配路径所经过的路径的可能性。如果路径（用表示）被更多的路径所包含，则路径称为最终路径一部分的概率就越大；如果路径没有被任何局部最优路径所包含，则最终匹配路径不包括该路径P。

对于采样点的每个候选点，其中为采样点个数，为设置的候选点个数，遍历每个候选点，当一个候选点被遍历时，我们假设该候选点就是最终匹配阶级过中正确的匹配点，找到一条通过点的概率最大的路径作为局部最优路径。我们用下图（a）真实路网为例，来演示查找局部最优解的过程。在该图中，有4个采样点以及其对应的候选点，下图（b）表示对于的候选点在候选图中累积的权重数据。在这个例子中，我们假设是正确的点，并设置路径和的权重为。这意味着路径必须通过候选点。累计权重。是的权重，受和的距离影响。在计算累积权重后，我们得到了所有候选点的。从可以得到局部最优候选点为，因为。从中我们可以得到上一个局部最优候选点。同样，我们得到了局部最优路径为，如下图 (b)中的虚线所示。

在找到每个候选点的局部最优路径后，我们得到了一组局部最优路径。通过对每一个点的局部最优路径进行投票，选出投票数最多的一条路径作为最终匹配的路径。具体的：遍历第四阶段生成的候选图，对图的所有边进行投票，最后筛选出每两个相邻采样点中投票最多的子路径作为最终匹配子路径，最后，所有子路径拼接成最终的最优匹配路径。

算法

# 实验

在本节中，我们使用真实出租车运营公司提供的全市出租车位置数据来评估MIVMM算法的性能。首先准备实验数据与设置实验方法，然后对MIVMM算法进行评估，并且将本算法与其他地图匹配算法进行多维度的比较，直观清晰的展示各个算法的性能表现。

## 数据

# 数据来源标注

我们采用葡萄牙波尔图市从2018年6月到2018年9月某出租车运营公司提供的开放数据作为评估MIVMM地图匹配算法的基础数据。该数据包括4000000条真实出租车轨迹数据，10000000个轨迹点。在这些数据中，首先需要过滤掉出租车停止所产生的GPS数据，只考虑车辆在行驶过程中产生的数据。在我们的数据中，包括采样时间戳、WGS84坐标经纬度、采样点的瞬时方向与瞬时速度等有效数据，采样间隔在10秒~30秒之间。为了全面评估MIVMM算法的性能，我们选取了三类产生在不同类型道路上的轨迹数据，包括高速公路、城市主干道以及次干道或支路。每种类型的道路随机选出100条轨迹数据进行实验来评估算法性能。在准确率方面的评估，我们根据个人经验，对所有选出的实验数据进行人工点位标注，以确保得到最准确的车辆位置，进而保障准确率的真实性。

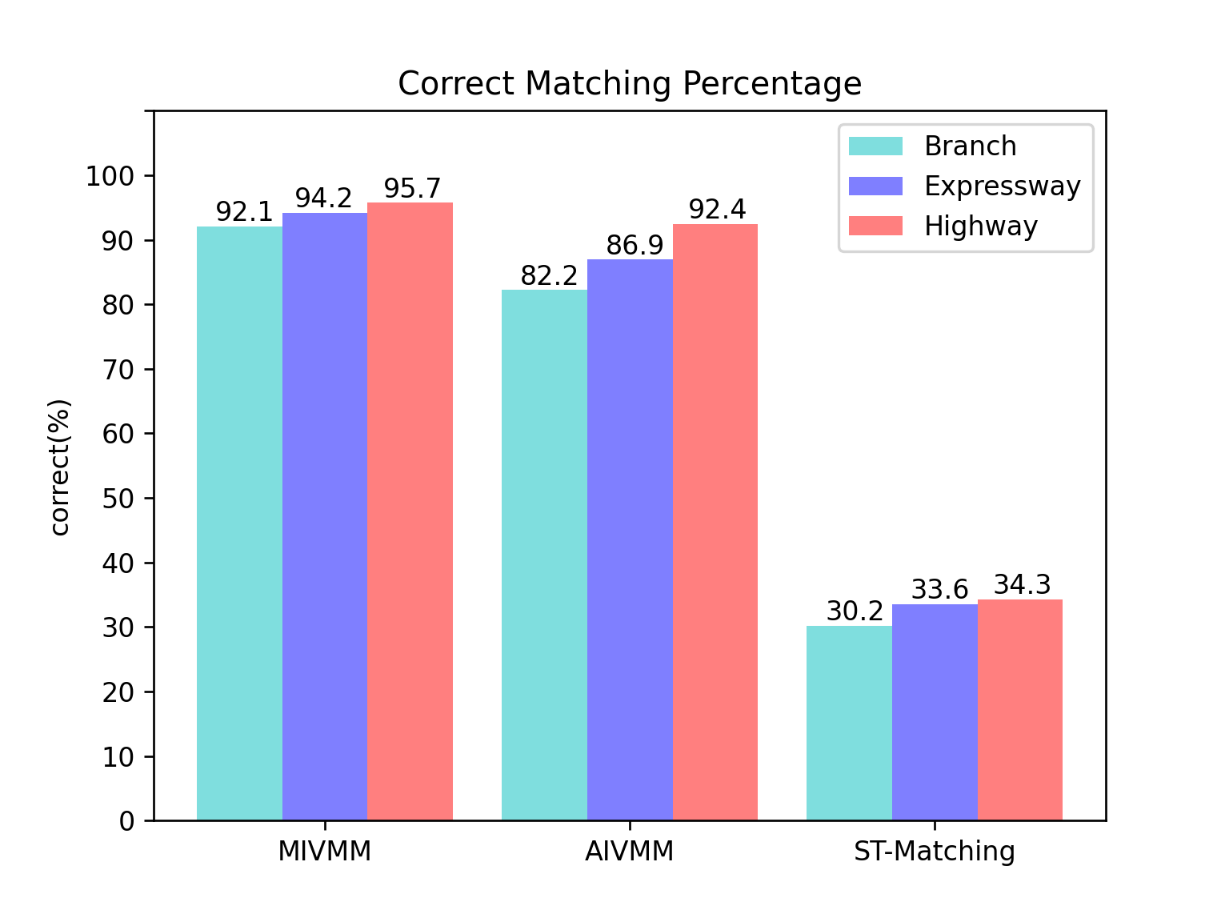
## 实验设置

为了得到更加严谨与全面的实验结果，在实验中我们分别取评估了候选点个数为2、3、4、5、6、7时MIVMM算法在各个性能指标中的数据，KNN算法不划定查询范围而是查询整个路网。在[]中提出了GPS设备采集的数据的偏移符合正态分布，其偏移量在范围内的概率为95.44%。正态分布函数的参数为：，，距离权重参数。本算法使用Python编程语言实现，运行平台配置为：英特尔i7-6700HQ处理器，16GB运行内存，Windows11操作系统。

## 准确率

准确率是指算法匹配道路的正确性，为比较直观的对比出算法的准确率，我们对比了AIVMM算法以及ST-Matching算法在三种不同类型道路上（分别为支路、主干道以及高速公路）的准确率表现，使用CMP方法来评估三种算法的准确率：

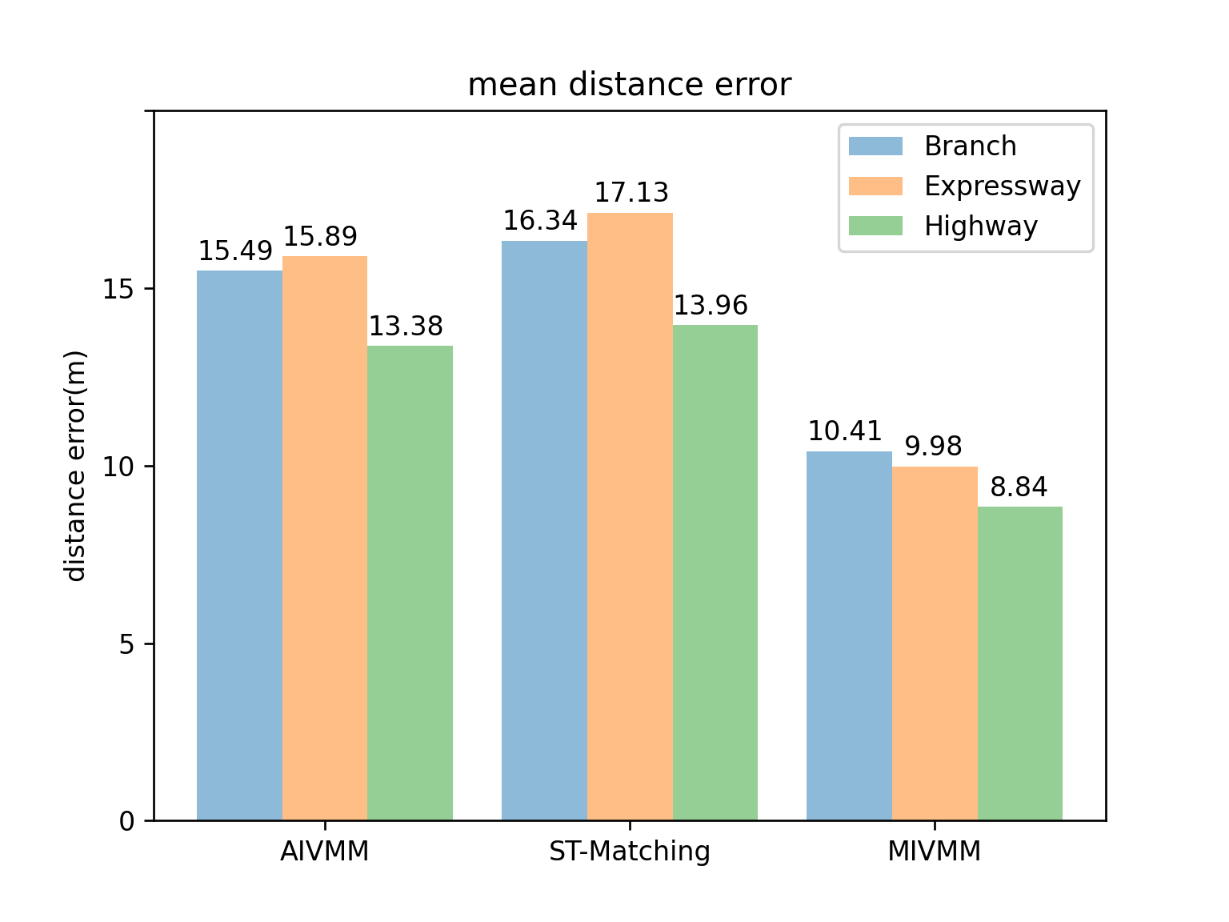
各个算法的准确率如下图所示：



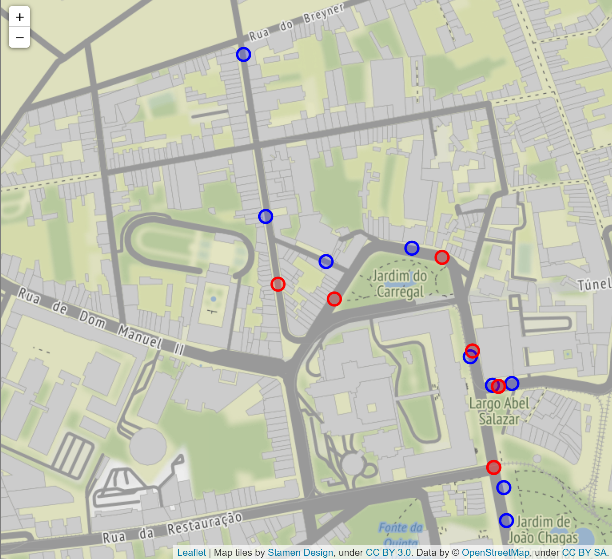
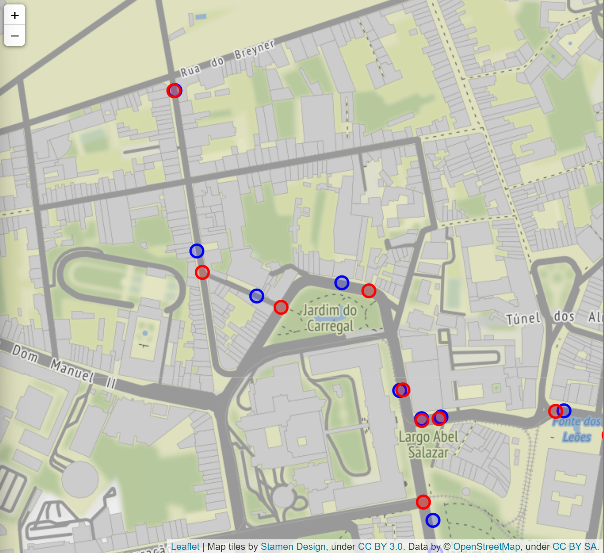
图中我们可以得出以下结论：三种算法的匹配准确率与路网的复杂度呈反比关系，即路网结构越复杂（路网复杂度：支路或次干道>主干道>高速公路），算法匹配的准确率就越低。ST-Matching的整体准确率较低，AIVMM算法较高，而本文提出的MIVMM算法无论是哪种道路类型都高于其他算法，都在90%以上。

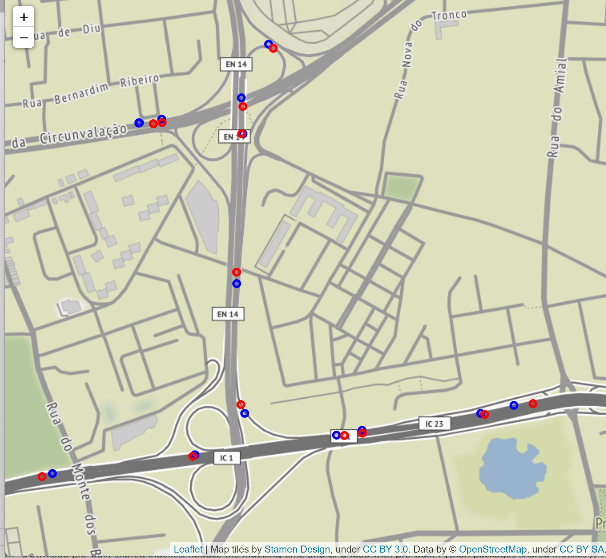
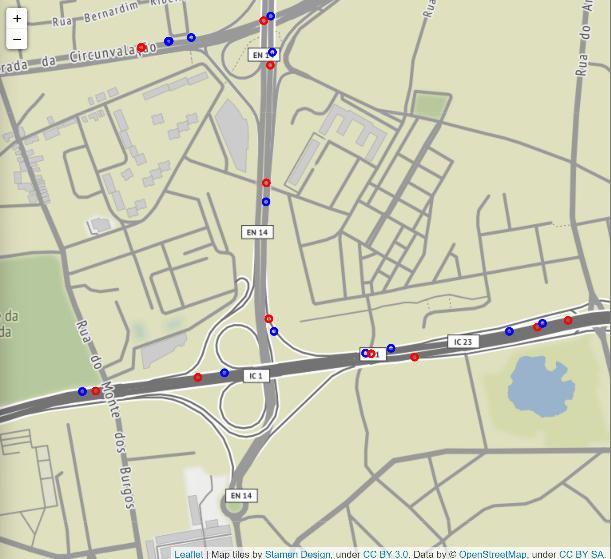
## 平均距离误差MDE

平均距离误差MDE是指算法在匹配正确的点到采样点的平均距离，平均距离越小，说明算法的精度越高，越能反应道路真实情况。MDE只考虑匹配正确的点且使用球面坐标系计算距离而不是常规的欧式距离，这样做是为了提高实验的准确度。

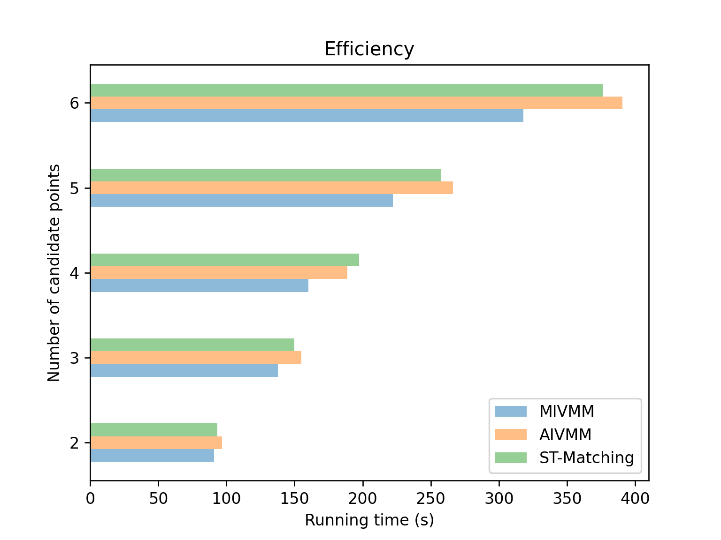


在平均距离误差的衡量标准下，AIVMM算法与ST-Matching算法在同一水平，前者MDE值略低，本文提出的MIVMM算法由于使用了全局KNN搜索以及墨卡托投影的加入降低了

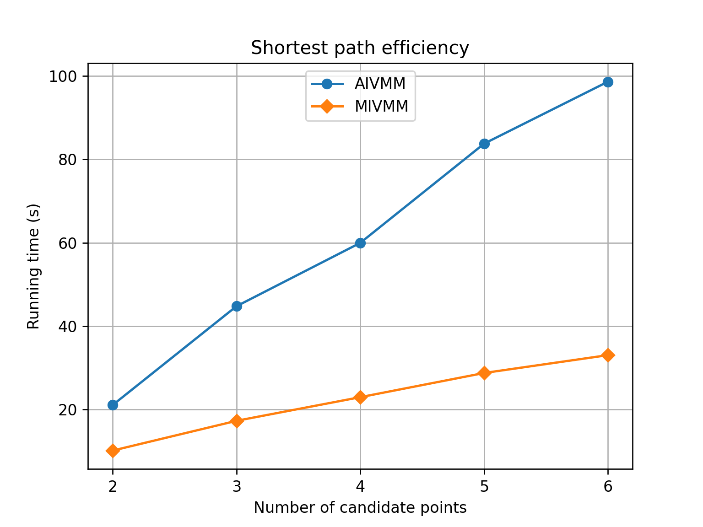
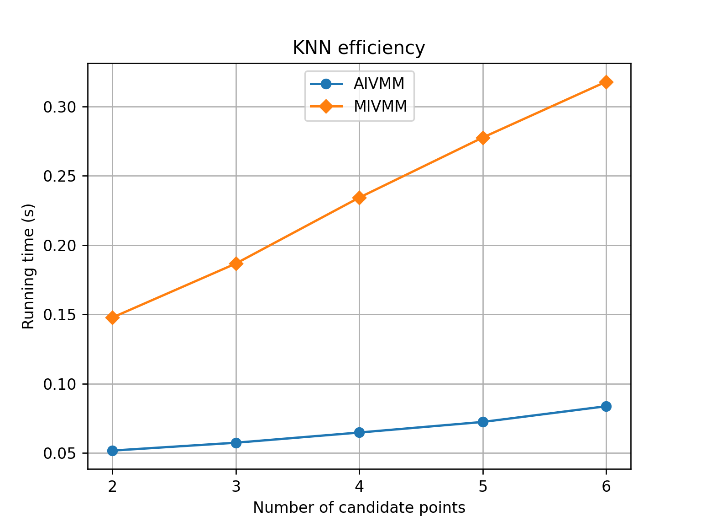
 

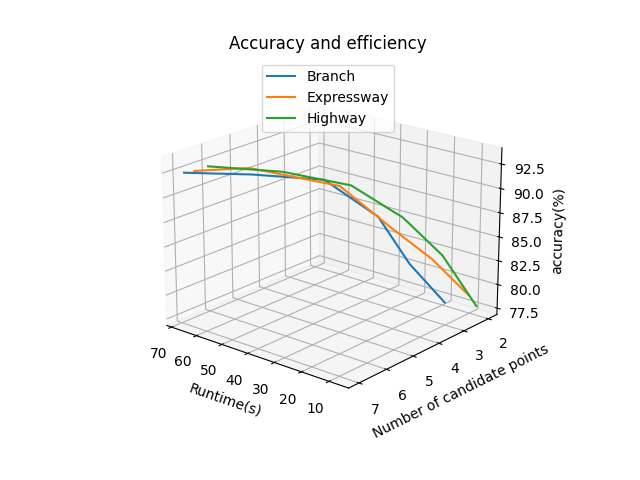
## 算法效率

图表, 折线图

描述已自动生成



## 准确率与效率



# 总结

## 

## 

参考文献

致谢