

清华大学

综合论文训练

题目：基于浮动车轨迹数据的路况分析

系 别：计算机科学与技术系

专 业：计算机科学与技术

姓 名：吴铮

指导教师：向勇副教授

2017 年 5 月 28 日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名：_____ 导师签名：_____ 日 期：_____

中文摘要

论文的摘要是对论文研究内容和成果的高度概括。摘要应对论文所研究的问题及其研究目的进行描述，对研究方法和过程进行简单介绍，对研究成果和所得结论进行概括。摘要应具有独立性和自明性，其内容应包含与论文全文同等量的主要信息。使读者即使不阅读全文，通过摘要就能了解论文的总体内容和主要成果。

论文摘要的书写应力求精确、简明。切忌写成对论文书写内容进行提要的形式，尤其要避免“第 1 章……；第 2 章……；……”这种或类似的陈述方式。

本文介绍清华大学论文模板 **ThUTHESIS** 的使用方法。本模板符合学校的本科、硕士、博士论文格式要求。

本文的创新点主要有：

- 用例子来解释模板的使用方法；
- 用废话来填充无关紧要的部分；
- 一边学习摸索一边编写新代码。

关键词是为了文献标引工作、用以表示全文主要内容信息的单词或术语。关键词不超过 5 个，每个关键词中间用分号分隔。（模板作者注：关键词分隔符不用考虑，模板会自动处理。英文关键词同理。）

关键词： $\text{T}_{\text{E}}\text{X}$ ； $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ；CJK；模板；论文

ABSTRACT

An abstract of a dissertation is a summary and extraction of research work and contributions. Included in an abstract should be description of research topic and research objective, brief introduction to methodology and research process, and summarization of conclusion and contributions of the research. An abstract should be characterized by independence and clarity and carry identical information with the dissertation. It should be such that the general idea and major contributions of the dissertation are conveyed without reading the dissertation.

An abstract should be concise and to the point. It is a misunderstanding to make an abstract an outline of the dissertation and words “the first chapter”, “the second chapter” and the like should be avoided in the abstract.

Key words are terms used in a dissertation for indexing, reflecting core information of the dissertation. An abstract may contain a maximum of 5 key words, with semi-colons used in between to separate one another.

Keywords: T_EX; L^AT_EX; CJK; template; thesis

目 录

第 1 章 带 English 的标题	1
1.1 封面相关	1
1.2 字体命令	1
1.3 表格样本	2
1.3.1 基本表格	2
1.3.2 复杂表格	3
1.3.3 其它	7
1.4 定理环境	8
1.5 参考文献	11
1.6 公式	12
第 2 章 中华人民共和国	14
2.1 其它例子	14
2.1.1 绘图	14
2.1.2 插图	14
插图索引	17
表格索引	18
公式索引	19
致 谢	20
声 明	21
附录 A 外文资料书面翻译	22
A.1 引言	22
A.2 地图匹配问题	24
A.2.1 问题定义	24
A.2.2 HMM 方法	24

A.3 OHMM 地图匹配算法	25
A.3.1 算法的基本流程	25
A.3.2 表现概率	25
A.3.3 转移概率	26
A.3.4 在线维特比算法	27
A.4 实验装置	28
A.4.1 现场测试数据	28
A.4.2 训练和参数估计	28
A.4.3 性能评价	29
A.5 结果	30
A.6 结论和未来工作	30
在学期间参加课题的研究成果	32

主要符号对照表

HPC	高性能计算 (High Performance Computing)
cluster	集群
Itanium	安腾
SMP	对称多处理
API	应用程序编程接口
PI	聚酰亚胺
MPI	聚酰亚胺模型化合物, N-苯基邻苯酰亚胺
PBI	聚苯并咪唑
MPBI	聚苯并咪唑模型化合物, N-苯基苯并咪唑
PY	聚吡咙
PMDA-BDA	均苯四酸二酐与联苯四胺合成的聚吡咙薄膜
ΔG	活化自由能 (Activation Free Energy)
χ	传输系数 (Transmission Coefficient)
E	能量
m	质量
c	光速
P	概率
T	时间
v	速度
劝学	君子曰：学不可以已。青，取之于蓝，而青于蓝；冰，水为之，而寒于水。木直中绳。輮以为轮，其曲中规。虽有槁暴，不复挺者，輮使之然也。故木受绳则直，金就砺则利，君子博学而日参省乎己，则知明而行无过矣。吾尝终日而思矣，不如须臾之所学也；吾尝跂而望矣，不如登高之博见也。登高而招，臂非加长也，而见者远；顺风而呼，声非加疾也，而闻者彰。假舆马者，非利足也，而致千里；假舟楫者，非能水也，而绝江河，君子生非异也，善假于物也。积土成山，风雨兴焉；积水成渊，蛟龙生焉；积善成德，而神明自得，圣心备焉。故不积跬步，无以至千里；不积小流，无以成江海。

骐驎一跃，不能十步；弩马十驾，功在不舍。锲而舍之，朽木
不折；锲而不舍，金石可镂。蚓无爪牙之利，筋骨之强，上
食埃土，下饮黄泉，用心一也。蟹六跪而二螯，非蛇鳝之穴
无可寄托者，用心躁也。——荀况

第 1 章 带 English 的标题

这是 `ThUThesis` 的示例文档，基本上覆盖了模板中所有格式的设置。建议大家在使用模板之前，除了阅读《`ThUThesis` 用户手册》，这个示例文档也最好能看一看。

小老鼠偷吃热凉粉；短长虫环绕矮高粱^①。

1.1 封面相关

封面的例子请参看 `cover.tex`。主要符号表参看 `denation.tex`，附录和个人简历分别参看 `appendix01.tex` 和 `resume.tex`。里面的命令都很只管，一看即会^②。

1.2 字体命令

苏轼（1037-1101），北宋文学家、书画家。字子瞻，号东坡居士，眉州眉山（今属四川）人。苏洵子。嘉佑进士。神宗时曾任祠部员外郎，因反对王安石新法而求外职，任杭州通判，知密州、徐州、湖州。后以作诗“谤讟朝廷”罪贬黄州。哲宗时任翰林学士，曾出知杭州、颖州等，官至礼部尚书。后又贬谪惠州、儋州。北还后第二年病死常州。南宋时追谥文忠。与父洵弟辙，合称“三苏”。在政治上属于旧党，但也有改革弊政的要求。其文汪洋恣肆，明白畅达，为“唐宋八大家”之一。其诗清新豪健，善用夸张比喻，在艺术表现方面独具风格。少数诗篇也能反映民间疾苦，指责统治者的奢侈骄纵。词开豪放一派，对后代很有影响。《念奴娇·赤壁怀古》、《水调歌头·丙辰中秋》传诵甚广。

坡仙擅长行书、楷书，取法李邕、徐浩、颜真卿、杨凝式，而能自创新意。用笔丰腴跌宕，有天真烂漫之趣。与蔡襄、黄庭坚、米芾并称“宋四家”。能画竹，学文同，也喜作枯木怪石。论画主张“神似”，认为“论画以形似，见与儿童邻”；高

① 韩愈（768-824），字退之，河南河阳（今河南孟县）人，自称郡望昌黎，世称韩昌黎。幼孤贫刻苦好学，德宗贞元八年进士。曾任监察御史，因上疏请免关中赋役，贬为阳山县令。后随宰相裴度平定淮西迁刑部侍郎，又因上表谏迎佛骨，贬潮州刺史。做过吏部侍郎，死谥文公，故世称韩吏部、韩文公。是唐代古文运动领袖，与柳宗元合称韩柳。诗力求险怪新奇，雄浑重气势。

② 你说还是看不懂？怎么会呢？

度评价“诗中有画，画中有诗”的艺术造诣。诗文有《东坡七集》等。存世书迹有《答谢民师论文帖》、《祭黄几道文》、《前赤壁赋》、《黄州寒食诗帖》等。画迹有《枯木怪石图》、《竹石图》等。

易与天地准，故能弥纶天地之道。仰以观於天文，俯以察於地理，是故知幽明之故。原始反终，故知死生之说。精气为物，游魂为变，是故知鬼神之情状。与天地相似，故不违。知周乎万物，而道济天下，故不过。旁行而不流，乐天知命，故不忧。安土敦乎仁，故能爱。范围天地之化而不过，曲成万物而不遗，通乎昼夜之道而知，故神无方而易无体。

[无 \youyuan 字体。] 有天地，然后万物生焉。盈天地之间者，唯万物，故受之以屯；屯者盈也，屯者物之始生也。物生必蒙，故受之以蒙；蒙者蒙也，物之穉也。物穉不可不养也，故受之以需；需者饮食之道也。饮食必有讼，故受之以讼。讼必有众起，故受之以师；师者众也。众必有所比，故受之以比；比者比也。比必有所畜也，故受之以小畜。物畜然后有礼，故受之以履。

履而泰，然后安，故受之以泰；泰者通也。物不可以终通，故受之以否。物不可以终否，故受之以同人。与人同者，物必归焉，故受之以大有。有大者不可以盈，故受之以谦。有大而能谦，必豫，故受之以豫。豫必有随，故受之以随。以喜随人者，必有事，故受之以蛊；蛊者事也。

[无 \lishu 字体。] 有事而后可大，故受之以临；临者大也。物大然后可观，故受之以观。可观而后有所合，故受之以噬嗑；嗑者合也。物不可以苟合而已，故受之以贲；贲者饰也。致饰然后亨，则尽矣，故受之以剥；剥者剥也。物不可以终尽，剥穷上反下，故受之以复。复则不妄矣，故受之以无妄。

有无妄然后可畜，故受之以大畜。物畜然后可养，故受之以颐；颐者养也。不养则不可动，故受之以大过。物不可以终过，故受之以坎；坎者陷也。陷必有所丽，故受之以离；离者丽也。

1.3 表格样本

1.3.1 基本表格

模板中关于表格的宏包有三个：`booktabs`、`array` 和 `longtabular`，命令有一个 `\hlinewd`。三线表可以用 `booktabs` 提供的 `\toprule`、`\midrule` 和 `\bottomrule`。它们与 `longtable` 能很好的配合使用。如果表格比较简单的话

可以直接用命令 `\hlinewd{<width>}` 控制。

表 1.1 模板文件。如果表格的标题很长，那么在表格索引中就会很不美观，所以要像 `chapter` 那样在前面用中括号写一个简短的标题。这个标题会出现在索引中。

文件名	描述
thuthesis.ins	L ^A T _E X 安装文件，DocStrip ^①
thuthesis.dtx	所有的一切都在这里 ^② 。
thuthesis.cls	模板类文件。
thuthesis.cfg	模板配置文。 <code>cls</code> 和 <code>cfg</code> 由前两个文件生成。
thuthesis.bst	参考文献 BIB _T E _X 样式文件。
thuthesis.sty	常用的包和命令写在这里，减轻主文件的负担。

① 表格中的脚注

② 再来一个

首先来看一个最简单的表格。表 1.1 列举了本模板主要文件及其功能。请大家注意三线表中各条线对应的命令。这个例子还展示了如何在表格中正确使用脚注。由于 L^AT_EX 本身不支持在表格中使用 `\footnote`，所以我们不得不将表格放在小页中，而且最好将表格的宽度设置为小页的宽度，这样脚注看起来才更美观。

1.3.2 复杂表格

我们经常会在表格下方标注数据来源，或者对表格里面的条目进行解释。前面的脚注是一种不错的方法，如果不喜欢脚注，可以在表格后面写注释，比如表 1.2。

表 1.2 复杂表格示例 1

x \ y	First Half		Second Half	
	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East*	20.4	27.4	90	20.4
West**	30.6	38.6	34.6	31.6

注：数据来源 《ThU_{Th}ESIS 使用手册》。

*：东部

**：西部

此外，表 1.2 同时还演示了另外两个功能：1) 通过 `tabularx` 的 `|X|` 扩展实现表格自动放大；2) 通过命令 `\diagbox` 在表头部分插入反斜线。

为了使我们的例子更接近实际情况，我会在必要的时候插入一些“无关”文字，以免太多图表同时出现，导致排版效果不太理想。第一个出场的当然是我的最爱：风流潇洒、骏马绝尘、健笔凌云的**李太白**了。

李白，字太白，陇西成纪人。凉武昭王暠九世孙。或曰山东人，或曰蜀人。白少有逸才，志气宏放，飘然有超世之心。初隐岷山，益州长史苏颋见而异之，曰：“是子天才英特，可比相如。”天宝初，至长安，往见贺知章。知章见其文，叹曰：“子谪仙人也。”言于明皇，召见金銮殿，奏颂一篇。帝赐食，亲为调羹，有诏供奉翰林。白犹与酒徒饮于市，帝坐沉香亭子，意有所感，欲得白为乐章，召入，而白已醉。左右以水颰面，稍解，援笔成文，婉丽精切。帝爱其才，数宴见。白常侍帝，醉，使高力士脱靴。力士素贵，耻之，摘其诗以激杨贵妃。帝欲官白，妃辄沮止。白自知不为亲近所容，恳求还山。帝赐金放还。乃浪迹江湖，终日沉饮。永王璘都督江陵，辟为僚佐。璘谋乱，兵败，白坐长流夜郎，会赦得还。族人阳冰为当涂令，白往依之。代宗立，以左拾遗召，而白已卒。文宗时，诏以白歌诗、裴旻剑舞、张旭草书为三绝云。集三十卷。今编诗二十五卷。——《全唐诗》诗人小传

浮动体的并排放置一般有两种情况：1) 二者没有关系，为两个独立的浮动体；2) 二者隶属于同一个浮动体。对表格来说并排表格既可以像图 1.3、图 1.4 使用小页环境，也可以如图 1.5 使用子表格来做。图的例子参见第 2.1.2.2 节。

表 1.3 第一个并排子表格

111	222
222	333

表 1.4 第二个并排子表格

111	222
222	333

然后就是忧国忧民，诗家楷模杜工部了。杜甫，字子美，其先襄阳人，曾祖依艺为巩令，因居巩。甫天宝初应进士，不第。后献《三大礼赋》，明皇奇之，召试文章，授京兆府兵曹参军。安禄山陷京师，肃宗即位灵武，甫自贼中遁赴行在，拜左拾遗。以论救房琯，出为华州司功参军。关辅饥乱，寓居同州同谷县，身自负薪采椽，餽糒不给。久之，召补京兆府功曹，道阻不赴。严武镇成都，奏为参谋、检校工部员外郎，赐绯。武与甫世旧，待遇甚厚。乃于成都浣花里种竹

植树，枕江结庐，纵酒啸歌其中。武卒，甫无所依，乃之东蜀就高适。既至而适卒。是岁，蜀帅相攻杀，蜀大扰。甫携家避乱荆楚，扁舟下峡，未维舟而江陵亦乱。乃溯沿湘流，游衡山，寓居耒阳。卒年五十九。元和中，归葬偃师首阳山，元稹志其墓。天宝间，甫与李白齐名，时称李杜。然元稹之言曰：“李白壮浪纵恣，摆去拘束，诚亦差肩子美矣。至若铺陈终始，排比声韵，大或千言，次犹数百，词气豪迈，而风调清深，属对律切，而脱弃凡近，则李尚不能历其藩翰，况堂奥乎。”白居易亦云：“杜诗贯穿古今，尽工尽善，殆过于李。”元、白之论如此。盖其出处劳佚，喜乐悲愤，好贤恶恶，一见之于诗。而又以忠君忧国、伤时念乱为本旨。读其诗可以知其世，故当时谓之“诗史”。旧集诗文共六十卷，今编诗十九卷。

表 1.5 并列子表格

(a) 第一个子表格		(b) 第二个子表格	
111	222	111	222
222	333	222	333

不可否认 \LaTeX 的表格功能没有想象中的那么强大，不过只要足够认真，足够细致，同样可以排出来非常复杂非常漂亮的表格。请参看表 1.6。

表 1.6 复杂表格示例 2

Network Topology		# of nodes	# of clients			Server
GT-ITM	Waxman Transit-Stub	600	2%	10%	50%	Max. Connectivity
Inet-2.1		6000				
Xue	Rui	Ni	ThUThESIS			
	ABCDEF					

最后就是清新飘逸、文约意赅、空谷绝响的王大侠了。王维，字摩诘，河东人。工书画，与弟缙俱有俊才。开元九年，进士擢第，调太乐丞。坐累为济州司仓参军，历右拾遗、监察御史、左补阙、库部郎中，拜吏部郎中。天宝末，为给事中。安禄山陷两都，维为贼所得，服药阳喑，拘于菩提寺。禄山宴凝碧池，维潜赋诗悲悼，闻于行在。贼平，陷贼官三等定罪，特原之，责授太子中允，迁中庶子、中书舍人。复拜给事中，转尚书右丞。维以诗名盛于开元、天宝间，宁薛诸王驸马豪贵之门，无不拂席迎之。得宋之问辋川别墅，山水绝胜，与道友裴

迪，浮舟往来，弹琴赋诗，啸咏终日。笃于奉佛，晚年长斋禅诵。一日，忽索笔作书数纸，别弟缙及平生亲故，舍笔而卒。赠秘书监。宝应中，代宗问缙：“朕常于诸王坐闻维乐章，今存几何？”缙集诗六卷，文四卷，表上之。敕答云，卿伯氏位列先朝，名高希代。抗行周雅，长揖楚辞。诗家者流，时论归美。克成编录，叹息良深。殷璠谓维诗词秀调雅，意新理惬。在泉成珠，著壁成绘。苏轼亦云：“维诗中有画，画中有诗也。”今编诗四卷。

要想用好论文模板还是得提前学习一些 \TeX/L\TeX 的相关知识，具备一些基本能力，掌握一些常见技巧，否则一旦遇到问题还真是比较麻烦。我们见过很多这样的同学，一直以来都是使用 Word 等字处理工具，以为 \LaTeX 模板的用法也应该类似，所以就沿袭同样的思路来对待这种所见非所得的排版工具，结果被折腾的焦头烂额，疲惫不堪。

如果您要排版的表格长度超过一页，那么推荐使用 `longtable` 或者 `supertabular` 宏包，模板对 `longtable` 进行了相应的设置，所以用起来可能简单一些。表 1.7 就是 `longtable` 的简单示例。

表 1.7 实验数据

测试程序	正常运行 时间 (s)	同步 时间 (s)	检查点 时间 (s)	卷回恢复 时间 (s)	进程迁移 时间 (s)	检查点 文件 (KB)
CG.A.2	23.05	0.002	0.116	0.035	0.589	32491
CG.A.4	15.06	0.003	0.067	0.021	0.351	18211
CG.A.8	13.38	0.004	0.072	0.023	0.210	9890
CG.B.2	867.45	0.002	0.864	0.232	3.256	228562
CG.B.4	501.61	0.003	0.438	0.136	2.075	123862
CG.B.8	384.65	0.004	0.457	0.108	1.235	63777
MG.A.2	112.27	0.002	0.846	0.237	3.930	236473
MG.A.4	59.84	0.003	0.442	0.128	2.070	123875
MG.A.8	31.38	0.003	0.476	0.114	1.041	60627
MG.B.2	526.28	0.002	0.821	0.238	4.176	236635
MG.B.4	280.11	0.003	0.432	0.130	1.706	123793
MG.B.8	148.29	0.003	0.442	0.116	0.893	60600
LU.A.2	2116.54	0.002	0.110	0.030	0.532	28754
LU.A.4	1102.50	0.002	0.069	0.017	0.255	14915
LU.A.8	574.47	0.003	0.067	0.016	0.192	8655
LU.B.2	9712.87	0.002	0.357	0.104	1.734	101975

续下页

续表 1.7 实验数据

测试程序	正常运行 时间 (s)	同步 时间 (s)	检查点 时间 (s)	卷回恢复 时间 (s)	进程迁移 时间 (s)	检查点 文件 (KB)
LU.B.4	4757.80	0.003	0.190	0.056	0.808	53522
LU.B.8	2444.05	0.004	0.222	0.057	0.548	30134
EP.A.2	123.81	0.002	0.010	0.003	0.074	1834
EP.A.4	61.92	0.003	0.011	0.004	0.073	1743
EP.A.8	31.06	0.004	0.017	0.005	0.073	1661
EP.B.2	495.49	0.001	0.009	0.003	0.196	2011
EP.B.4	247.69	0.002	0.012	0.004	0.122	1663
EP.B.8	126.74	0.003	0.017	0.005	0.083	1656

1.3.3 其它

如果不想让某个表格或者图片出现在索引里面，请使用命令 `\caption*`。这个命令不会给表格编号，也就是出来的只有标题文字而没有“表 XX”，“图 XX”，否则索引里面序号不连续就显得不伦不类，这也是 \LaTeX 里星号命令默认的规则。

有这种需求的多是本科同学的英文资料翻译部分，如果觉得附录中英文原文中的表格和图片显示成“表”和“图”不协调的话，一个很好的办法就是用 `\caption*`，参数随便自己写，比如不守规矩的表 1.111 和图 1.111 能满足这种特殊需要（可以参看附录部分）。

表 1.111 这是一个手动编号，不出现在索引中的表格。

THUThesis

Figure 1.111 这是一个手动编号，不出现在索引中的图。

薛瑞尼

如果的确想让它编号，但又不想让它出现在索引中的话，目前模板上不支持。

最后，虽然大家不一定会独立使用小页，但是关于小页中的脚注还是有必要提一下。请看下面的例子。

柳宗元，字子厚（773-819），河东（今永济县）人^①，是唐代杰出的文学家，哲学家，同时也是一位政治改革家。与韩愈共同倡导唐代古文运动，并称韩柳^②。

① 山西永济水校。

② 唐宋八大家之首二位。

唐朝安史之乱后，宦官专权，藩镇割据，土地兼并日渐严重，社会生产破坏严重，民不聊生。柳宗元对这种社会现实极为不满，他积极参加了王叔文领导的“永济革新”，并成为这一运动的中坚人物。他们革除弊政，打击权奸，触犯了宦官和官僚贵族利益，在他们的联合反扑下，改革失败了，柳宗元被贬为永州司马。

1.4 定理环境

给大家演示一下各种和证明有关的环境：

假设 1.1： 待月西厢下，迎风户半开；隔墙花影动，疑是玉人来。

$$c = a^2 - b^2 \quad (1-1)$$

$$= (a + b)(a - b) \quad (1-2)$$

千辛万苦，历尽艰难，得有今日。然相从数千里，未曾衰戚。今将渡江，方图百年欢笑，如何反起悲伤？（引自《杜十娘怒沉百宝箱》）

定义 1.1： 子曰：「道千乘之国，敬事而信，节用而爱人，使民以时。」

千古第一定义！问世间、情为何物，只教生死相许？天南地北双飞客，老翅几回寒暑。欢乐趣，离别苦，就中更有痴儿女。君应有语，渺万里层云，千山暮雪，只影向谁去？

横汾路，寂寞当年箫鼓，荒烟依旧平楚。招魂楚些何嗟及，山鬼暗啼风雨。天也妒，未信与，莺儿燕子俱黄土。千秋万古，为留待骚人，狂歌痛饮，来访雁丘处。

命题 1.1： 曾子曰：「吾日三省吾身——为人谋而不忠乎？与朋友交而不信乎？传不习乎？」

多么凄美的命题啊！其日牛马嘶，新妇入青庐，奄奄黄昏后，寂寂人定初，我命绝今日，魂去尸长留，揽裙脱丝履，举身赴清池，府吏闻此事，心知长别离，徘徊庭树下，自挂东南枝。

注释 1.1： 天不言自高，水不言自流。

$$\begin{aligned}\varphi(x, z) &= z - \gamma_{10}x - \gamma_{mn}x^m z^n \\ &= z - Mr^{-1}x - Mr^{-(m+n)}x^m z^n \\ \zeta^0 &= (\xi^0)^2, & (1-3) \\ \zeta^1 &= \xi^0 \xi^1, & (1-4) \\ \zeta^2 &= (\xi^1)^2, & (1-5)\end{aligned}$$

天尊地卑，乾坤定矣。卑高以陈，贵贱位矣。动静有常，刚柔断矣。方以类聚，物以群分，吉凶生矣。在天成象，在地成形，变化见矣。鼓之以雷霆，润之以风雨，日月运行，一寒一暑，乾道成男，坤道成女。乾知大始，坤作成物。乾以易知，坤以简能。易则易知，简则易从。易知则有亲，易从则有功。有亲则可久，有功则可大。可久则贤人之德，可大则贤人之业。易简，而天下矣之理矣；天下之理得，而成位乎其中矣。

公理 1.1： 两点间直线段距离最短。

$$\begin{aligned}x &\equiv y + 1 \pmod{m^2} & (1-6) \\ x &\equiv y + 1 \pmod{m^2} & (1-7) \\ x &\equiv y + 1 \pmod{m^2} & (1-8)\end{aligned}$$

《彖曰》：大哉乾元，万物资始，乃统天。云行雨施，品物流形。大明始终，六位时成，时乘六龙以御天。乾道变化，各正性命，保合大和，乃利贞。首出庶物，万国咸宁。

《象曰》：天行健，君子以自强不息。潜龙勿用，阳在下也。见龙再田，德施普也。终日乾乾，反复道也。或跃在渊，进无咎也。飞龙在天，大人造也。亢龙有悔，盈不可久也。用九，天德不可为首也。

引理 1.1: 《猫和老鼠》是我最爱看的动画片。

$$\begin{aligned} \int_a^b \left\{ \int_a^b [f(x)^2 g(y)^2 + f(y)^2 g(x)^2] - 2f(x)g(x)f(y)g(y) dx \right\} dy \\ = \int_a^b \left\{ g(y)^2 \int_a^b f^2 + f(y)^2 \int_a^b g^2 - 2f(y)g(y) \int_a^b fg \right\} dy \end{aligned}$$

行行重行行，与君生别离。相去万余里，各在天一涯。道路阻且长，会面安可知。胡马依北风，越鸟巢南枝。相去日已远，衣带日已缓。浮云蔽白日，游子不顾返。思君令人老，岁月忽已晚。弃捐勿复道，努力加餐饭。

定理 1.1: 犯我强汉者，虽远必诛 ——陈汤（汉）

$$y = 1 \quad (1-9a)$$

$$y = 0 \quad (1-9b)$$

道可道，非常道。名可名，非常名。无名天地之始；有名万物之母。故常无，欲以观其妙；常有，欲以观其徼。此两者，同出而异名，同谓之玄。玄之又玄，众妙之门。上善若水。水善利万物而不争，处众人之所恶，故几于道。曲则全，枉则直，洼则盈，敝则新，少则多，多则惑。人法地，地法天，天法道，道法自然。知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者寿。

证明 燕赵古称多感慨悲歌之士。董生举进士，连不得志于有司，怀抱利器，郁郁适兹土，吾知其必有合也。董生勉乎哉？

夫以子之不遇时，苟慕义强仁者，皆爱惜焉，矧燕、赵之士出乎其性者哉！然吾尝闻风俗与化移易，吾恶知其今不异于古所云邪？聊以吾子之行卜之也。董生勉乎哉？

吾因子有所感矣。为我吊望诸君之墓，而观于其市，复有昔时屠狗者乎？为我谢曰：“明天子在上，可以出而仕矣！” ——韩愈《送董邵南序》 □

推论 1.1: 四川话配音的《猫和老鼠》是世界上最好看最好听最有趣的动画片。

$$V_i = v_i - q_i v_j, \quad X_i = x_i - q_i x_j, \quad U_i = u_i, \quad \text{for } i \neq j; \quad (1-10)$$

$$V_j = v_j, \quad X_j = x_j, \quad U_j u_j + \sum_{i \neq j} q_i u_i. \quad (1-11)$$

迢迢牵牛星，皎皎河汉女。纤纤擢素手，札札弄机杼。终日不成章，泣涕零如雨。河汉清且浅，相去复几许。盈盈一水间，脉脉不得语。

例 1.1: 大家来看这个例子。

$$\begin{cases} \nabla f(\mathbf{x}^*) - \sum_{j=1}^p \lambda_j \nabla g_j(\mathbf{x}^*) = 0 \\ \lambda_j g_j(\mathbf{x}^*) = 0, \quad j = 1, 2, \dots, p \\ \lambda_j \geq 0, \quad j = 1, 2, \dots, p. \end{cases} \quad (1-12)$$

练习 1.1: 清列出 Andrew S. Tanenbaum 和 W. Richard Stevens 的所有著作。

猜想 1.1: *Poincare Conjecture* If in a closed three-dimensional space, any closed curves can shrink to a point continuously, this space can be deformed to a sphere.

问题 1.1: 回答还是不回答，是个问题。

如何引用定理 1.1 呢? 加上 \label 使用 \ref 即可。妾发初覆额，折花门前剧。郎骑竹马来，绕床弄青梅。同居长干里，两小无嫌猜。十四为君妇，羞颜未尝开。低头向暗壁，千唤不一回。十五始展眉，愿同尘与灰。常存抱柱信，岂上望夫台。十六君远行，瞿塘滟滪堆。五月不可触，猿声天上哀。门前迟行迹，一一生绿苔。苔深不能扫，落叶秋风早。八月蝴蝶来，双飞西园草。感此伤妾心，坐愁红颜老。

1.5 参考文献

当然参考文献可以直接写 \bibitem, 虽然费点功夫, 但是好控制, 各种格式可以自己随意改写。

本模板推荐使用 **BIB_TE_X**, 样式文件为 thuthesis.bst, 基本符合学校的参考文献格式 (如专利等引用未加详细测试)。看看这个例子, 关于书的^[1], 还有这些^[2], 关于杂志的^[3], 硕士论文^[4], 博士论文^[5], 标准文件^[6], 会

议论文^[? ?]，技术报告^[?]，电子文献^[? ?]。中文参考文献^[?]应增加 `lang="zh"` 字段，以便进行相应处理。另外，本模板对中文文献^[?]的支持并不是十全十美，如果有不如意的地方，请手动修改 `bb1` 文件。

有时候不想要上标，那么可以这样^[?]，这个非常重要。

有时候一些参考文献没有纸质出处，需要标注 **URL**。缺省情况下，**URL** 不会在连字符处断行，这可能使得用连字符代替空格的网址分行很难看。如果需要，可以将模板类文件中

```
\RequirePackage{hyperref}
```

一行改为：

```
\PassOptionsToPackage{hyphens}{url}
\RequirePackage{hyperref}
```

使得连字符处可以断行。更多设置可以参考 `url` 宏包文档。

1.6 公式

贝叶斯公式如式 (1-13)，其中 $p(y|\mathbf{x})$ 为后验； $p(\mathbf{x})$ 为先验；分母 $p(\mathbf{x})$ 为归一化因子。

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} \quad (1-13)$$

论文里面公式越多， \TeX 就越 happy。再看一个 `amsmath` 的例子：

$$\det \mathbf{K}(t = 1, t_1, \dots, t_n) = \sum_{I \in \mathbf{n}} (-1)^{|I|} \prod_{i \in I} t_i \prod_{j \in I} (D_j + \lambda_j t_j) \det \mathbf{A}^{(\lambda)}(\bar{I}|\bar{I}) = 0. \quad (1-14)$$

前面定理示例部分列举了很多公式环境，可以说把常见的情况都覆盖了，大家在写公式的时候一定要好好看 `amsmath` 的文档，并参考模板中的用法：

$$\begin{aligned} & \int_a^b \left\{ \int_a^b [f(x)^2 g(y)^2 + f(y)^2 g(x)^2] - 2f(x)g(x)f(y)g(y) dx \right\} dy \\ &= \int_a^b \left\{ g(y)^2 \int_a^b f^2 + f(y)^2 \int_a^b g^2 - 2f(y)g(y) \int_a^b fg \right\} dy \end{aligned}$$

其实还可以看看这个多级规划：

$$\left\{ \begin{array}{l} \max_x F(x, y_1^*, y_2^*, \dots, y_m^*) \\ \text{subject to:} \\ G(x) \leq 0 \\ (y_1^*, y_2^*, \dots, y_m^*) \text{ solves problems } (i = 1, 2, \dots, m) \\ \left\{ \begin{array}{l} \max_{y_i} f_i(x, y_1, y_2, \dots, y_m) \\ \text{subject to:} \\ g_i(x, y_1, y_2, \dots, y_m) \leq 0. \end{array} \right. \end{array} \right. \quad (1-15)$$

这些跟规划相关的公式都来自于刘宝碁老师《不确定规划》的课件。

第 2 章 中华人民共和国

2.1 其它例子

在第 1 章中我们学习了贝叶斯公式 (1-13)，这里我们复习一下：

$$p(y|\mathbf{x}) = \frac{p(\mathbf{x}, y)}{p(\mathbf{x})} = \frac{p(\mathbf{x}|y)p(y)}{p(\mathbf{x})} \quad (2-1)$$

2.1.1 绘图

本模板不再预先装载任何绘图包（如 `pstricks`，`pgf` 等），完全由用户来决定。个人觉得 `pgf` 不错，不依赖于 `Postscript`。此外还有很多针对 \LaTeX 的 GUI 作图工具，如 `XFig(jFig)`, `WinFig`, `Tpx`, `Ipe`, `Dia`, `Inkscape`, `LaTeXPiX`, `jPicEdt`, `jaxdraw` 等等。

2.1.2 插图

强烈推荐《 $\text{\LaTeX} 2_{\epsilon}$ 插图指南》！关于子图形的使用细节请参看 `subcaption` 宏包的说明文档。

2.1.2.1 一个图形

一般图形都是处在浮动环境中。之所以称为浮动是指最终排版效果图形的位置不一定与源文件中的位置对应^①，这也是刚使用 \LaTeX 同学可能遇到的问题。如果要强制固定浮动图形的位置，请使用 `float` 宏包，它提供了 `[H]` 参数，比如图 2.1。



图 2.1 利用 `Xfig` 制图

① This is not a bug, but a feature of \LaTeX !

大学之道，在明明德，在亲民，在止于至善。知止而后有定；定而后能静；静而后能安；安而后能虑；虑而后能得。物有本末，事有终始。知所先后，则近道矣。古之欲明明德于天下者，先治其国；欲治其国者，先齐其家；欲齐其家者，先修其身；欲修其身者，先正其心；欲正其心者，先诚其意；欲诚其意者，先致其知；致知在格物。物格而后知至；知至而后意诚；意诚而后心正；心正而后身修；身修而后家齐；家齐而后国治；国治而后天下平。自天子以至于庶人，壹是皆以修身为本。其本乱而未治者否矣。其所厚者薄，而其所薄者厚，未之有也！

——《大学》

2.1.2.2 多个图形

如果多个图形相互独立，并不共用一个图形计数器，那么用 `minipage` 或者 `parbox` 就可以。否则，请参看图 2.2，它包含两个小图，分别是图 2.2(a)和图 2.2(b)。推荐使用 `\subcaptionbox`，因为可以像图 2.2 那样对齐子图的标题，也可以使用 `subcaption` 宏包的 `\subcaption`（放在 `minipage` 中，用法同 `\caption`）或是 `subfigure`、`subtable` 环境，像图 2.3，不要再用 `\subfloat`、`\subfigure` 和 `\subtable`。



(a) 第一个小图形



(b) 第二个小图形，注意这个图略矮些。如果标题很长的话，它会自动换行

图 2.2 包含子图形的大图形 (`subcaptionbox` 示例)

古之学者必有师。师者，所以传道受业解惑也。人非生而知之者，孰能无惑？惑而不从师，其为惑也，终不解矣。生乎吾前，其闻道也固先乎吾，吾从而师之；生乎吾后，其闻道也亦先乎吾，吾从而师之。吾师道也，夫庸知其年之先后生于吾乎！是故无贵无贱无长无少，道之所存，师之所存也。

嗟乎！师道之不传也久矣，欲人之无惑也难矣。古之圣人，其出人也远矣，犹且从师而问焉；今之众人，其下圣人也亦远矣，而耻学于师。是故圣益圣，愚



(a) 第一个小图形

清华大学

(b) 第二个小图形，注意这个图略矮些。sub-figure 中同一行的子图在顶端对齐。

图 2.3 包含子图形的大图形 (subfigure 示例)



清华大学
Tsinghua University

图 2.4 并排第一个图



清华大学
Tsinghua University

图 2.5 并排第二个图

益愚。圣人之所以为圣，愚人之所以为愚，其皆出於此乎？爱其子，择师而教之，於其身也，则耻师焉，惑焉。彼童子之师，授之书而习其句读者，非吾所谓传其道、解其惑者也。句读之不知，惑之不解，或师焉，或不焉，小学而大遗，吾未见其明也。巫医、乐师、百工之人不耻相师，士大夫之族曰“师”曰“弟子”之云者，则群聚而笑之。问之，则曰：彼与彼年相若也，道相似也，位卑则足羞，官盛则近谄。呜呼！师道之不复，可知矣。巫医、乐师、百工之人。吾子不齿，今其智乃反不能及，其可怪也欤！圣人无常师。孔子师郯子、苌子、师襄、老聃。郯子之徒，其贤不及孔子。孔子曰：“三人行，必有我师。”是故弟子不必不如师，师不必贤於弟子。闻道有先後，术业有专攻，如是而已。

如果要把编号的两个图形并排，那么小页就非常有用了：

李氏子蟠，年十七，好古文、六艺，经传皆通习之，不拘於时，学於余。余嘉其能行古道，作师说以贻之。

——韩愈（唐）

插图索引

图 2.1	利用 Xfig 制图	14
图 2.2	包含子图形的大图形 (subcaptionbox 示例).....	15
图 2.3	包含子图形的大图形 (subfigure 示例)	16
图 2.4	并排第一个图	16
图 2.5	并排第二个图	16

表格索引

表 1.1	模板文件	3
表 1.2	复杂表格示例 1.....	3
表 1.3	第一个并排子表格	4
表 1.4	第二个并排子表格	4
表 1.5	并排子表格.....	5
表 1.6	复杂表格示例 2.....	5
表 1.7	实验数据	6

公式索引

公式 1-1	8
公式 1-2	8
公式 1-3	9
公式 1-4	9
公式 1-5	9
公式 1-6	9
公式 1-7	9
公式 1-8	9
公式 1-9a	10
公式 1-9b	10
公式 1-10	10
公式 1-11	11
公式 1-12	11
公式 1-13	12
公式 1-14	12
公式 1-15	13
公式 2-1	14

致 谢

衷心感谢导师 xxx 教授和物理系 xxx 副教授对本人的精心指导。他们的言传身教将使我终生受益。

在美国麻省理工学院化学系进行九个月的合作研究期间，承蒙 xxx 教授热心指导与帮助，不胜感激。感谢 xx 实验室主任 xx 教授，以及实验室全体老师和同学们的热情帮助和支持！本课题承蒙国家自然科学基金资助，特此致谢。

感谢 THUTHESIS，它的存在让我的论文写作轻松自在了许多，让我的论文格式规整漂亮了许多。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

附录 A 外文资料书面翻译

应用于实时交通传感应用基于隐马尔科夫模型的在线地图匹配

摘要: 在许多智能交通系统 (ITS) 应用程序中, 探测车辆的数据来源, 一个关键的步骤是准确地将 GPS 轨迹实时地映射到道路网络。这个过程被称为地图匹配, 通常需要考虑数据的噪声和稀疏性, 因为 (1) 高度精确的 GPS 轨迹信息非常稀有, 以及 (2) 密集轨迹对于实况传输和存储是昂贵的。

我们提出了一种基于对噪声和稀疏性鲁棒的隐马尔可夫模型 (HMM) 的在线地图匹配算法。我们集中在现有的基于 HMM 的算法的两个改进: (1) 使用最佳局部化策略, 可变滑动窗口 (VSW) 方法, 保证在不确定的未来输入下的在线解决方案质量, 和 (2) 使用机器学习的空间, 时间和拓扑信息。我们使用在城市和农村地区的公交路线上收集的现场测试数据来评估我们的算法的准确性。此外, 我们还研究了在处理实时输入流时精度和输出延迟之间的关系。

在我们对现场测试数据的测试中, VSW 在精度和输出延迟方面都优于传统的定位方法。我们的研究表明, 它可以应用于对延迟要求较为苛刻的应用程序, 如交通流量传感。

A.1 引言

由大量车辆收集的实时传感器数据, 例如城市地区的出租车和公共汽车, 为交通感测, 交通事故检测, 旅行时间预测 [3], 车辆管理 [4] 和路线建议 [5], [6]。这些系统的可用性取决于通过可用的地图匹配算法提取的数据的可靠性, 地图匹配算法做的就是将 GPS 轨迹投影到数字地图上的相应路段。

诸如时间戳, 位置和速度的信息通常由探测车辆记录。然而, 处理这些大量数据所需的高额到难以负担的存储和带宽导致了现在收集稀疏样本的做法, 间隔范围从几十秒到几分钟 [7]。此外, 已知安装在这些探头上的传感器容易出现各种错误 [7], 例如不精确的位置和速度测量, 重复传输和时间戳不匹配。在诸如交通感测的实时应用中, 还要求在未来数据点未给出的时候就执行地上的地图匹配, 即算法必须在线。可靠的在线地图匹配算法因此必须考虑这些问题并且保证精度和及时性高的输出。

地图匹配算法被定性为全局或增量/在线。全局算法在生成解之前对整个输入轨迹进行批处理。增量/在线算法采用将输入轨迹划分为更小段并按顺序处理的局部化策略，有时导致次优解。在地图匹配中应用的技术包括几何分析 [12]，信念函数理论 [13]，扩展卡尔曼滤波器 [14] 和隐马尔可夫模型 [11] [15] [16]。这些方法的优势和局限性已在 [9] 中综述。特别地，已经采用基于 **HMM** 的算法及其变体 [10]，[17] 用于同时评估实际映射的多个假设的能力，以便找到最终的最大似然解。这些方法已经被证明可以抵抗高度噪声测量，例如来自 **GSM** 塔的位置指纹 [16]，并且它们的精度随着轨迹的时间稀疏度的增加而退化 [10]，[15]，[17]。

我们提出的在线 **HMM** (**OHMM**) 地图匹配算法受到最近基于 **HMM** 的方法 [10]，[15] - [17] 的启发。我们解决了在以前的相关工作中没有集中的两个具体问题：(i) 需要一个在线算法来管理精度和输出延迟之间的权衡，以及 (ii) 多个评分函数的融合以估计转换可能性。

大多数现有的增量/在线算法使用简单的定位策略，例如固定滑动窗口和固定深度递归预测。滑动窗口方法简单地将轨迹划分为固定大小的输入序列并且独立地处理它们。较大的窗口尺寸导致更好的精度 [10]，[17] 但更长的输出延迟，反之亦然。递归预测方法将每个点的决定延迟固定数量的步长，以评估未来路径替代方案 [18]。这两种方法虽然易于实现，但可能导致次优解决方案和长输出延迟。当地图匹配算法在实时输入流上操作并且需要在短时间窗口内产生输出时，这显然是不可行的。受实时应用需求的驱使，我们提出了一个最优的局部化策略，使用可变滑动窗口 (**VSW**) 将输入分成更小的子问题，并证明找到全局最优解。我们的方法在概念上类似于在线维特比算法 [19]，[20]。此外，我们还开发了 **VSW** 的次优变体，其提供对延迟性能的最坏情况保证。在第五节中，我们将使用定长滑动窗口 (**FSW**) 方法作为基准来比较这两种方法的性能。

其次，我们推导了一种概率评分机制，其在 **HMM** 中的表现和转移概率的建模中结合了各种传感器数据和拓扑信息：(i) **GPS** 坐标 (ii) 车辆速度 (iii) 速度限制 (iv) 推断的车辆前进方向，以及 (vi) 拓扑约束。我们使用支持向量机 (**SVM**) 来学习转移概率函数，而不是选择简单的先验模型 [10]，[15]，[17]，然后估计其参数。这种方法的主要优点是它提供了一个数据驱动的框架，用于集成多个过渡评分函数。

我们使用在新加坡包括农村和城市地区的 4 条公交线路收集的现场测试

数据来评估我们的方法的性能。性能是根据精度和输出延迟标准来测量的。我们的研究表明，当操作在实时输入流时，所提出的算法实现了优化或接近最优解，实际上具有低输出延迟。

本文结构如下。第二节根据 **HMM** 来制定地图匹配问题。方法将在第三节中描述。我们的实验设置在第四部分解释。结果在第五节中介绍。最后，第六部分总结了我们的贡献，并讨论了未来工作的空间。

A.2 地图匹配问题

A.2.1 问题定义

定义 1: 轨迹, $T = (t_n | n = 1, \dots, N)$, 是由车辆收集的 N 个数据点组成的序列。每个轨迹点由其经度 ($t_n.lon$), 纬度 ($t_n.lat$), 速度 ($t_n.v$) 和时间戳 ($t_n.t$) 指定。

定义 2: 段落, $r = (p_m | m = 1, \dots, M)$, 是一条 M 点折线表示的道路段曲线。它由连接顶点的一系列线段 p_1, \dots, p_m 组成。按顺序, 其中每个顶点由其经度和纬度指定。段落也由其道路宽度 ($r.w$), 速度限制 ($r.v$) 和双向行驶 ($r.d$) 的允许性定义 ($r.d = \{\text{true}, \text{false}\}$)。

定义 3: 数字地图, $G = \{r_k | k = 1, \dots, K\}$, 是一组代表 K 个段落的集合。

给定轨迹 T , 地图匹配的目标是找到 T 每个轨迹点到 G 中段落的对应关系。

A.2.2 HMM 方法

在基于 **HMM** 的地图匹配算法中, 候选路径被顺序地生成并基于它们的似然性来评估。当遇到新的轨迹点时, 解决方案的过去的假设被扩展以考虑新的观测值。在最后阶段的所有候选中, 具有最高联合概率的幸存路径然后被选择为最终解。

对于每个轨迹点, 我们首先从该数据最有可能被采集的地点识别一组候选道路段落。这些候选中的每一个表示为马尔可夫链中的隐藏状态, 并且具有表现概率, 其是在候选段是真实匹配的情况下观察 **GPS** 点的可能性。直观地, 如果点被发现更接近它, 我们将给段更高的概率。然后, 我们计算链中每对相邻隐藏状态的转移概率, 使得后者的概率仅取决于前者, 因此服从马尔可夫假设。我们的目标是找到具有最高联合表现和传输概率的马尔科夫链上的最大似然路径。

该过程如图 1 所示。

正式地，我们将表现概率表示为 $p(t | h)$ ，其是给定隐藏状态（段）的观察
到轨迹点 t 的概率。从隐藏状态到隐藏状态的转移概率是 $f(s, h)$ 。给定一个 N
个点轨迹，隐藏状态的似然序列， $T = (t_n | n = 1, \dots, N)$ ，隐藏状态的似然序列，
 $S^* = (S_n \in A_n | n = 1, \dots, N)$ 满足以下递推关系。

$$V_{n,h} = p(t_n | h) \max_{s \in A_{n-1}} \{f(s, h) V_{n-1,s}\}.$$

这里 $V_{0,h} = p(t_0 | h)$ 和 A_n 表示第 n 阶段的隐藏状态集合。然后，我们可以从
最后一个元素中找到 $S^*, s_N = \arg \max_{h \in A_n} \{V_{N,h}\}$ ，向后工作找到最大联合概率序
列 S_{N-1}, \dots, S_1 。我们在第 3-E 节将提供一个在线算法找到 S^* 。

A.3 OHMM 地图匹配算法

A.3.1 算法的基本流程

- 对于每个轨迹点，找到围绕其半径为 50m 的所有候选分段。施加该阈值的原因
是双重的：（1）丢弃具有非常低的表现概率（低于 10^{-4} 的范围）的所有
候选，以及（2）避免由于过多候选而导致执行速度的惩罚。
- 针对每个候选分段（隐藏状态）计算表现概率，而将转移概率分配给在隐
藏状态上发生的每个边缘。
- VSW 算法在更新的 Markov 链上执行回溯，并给出部分求解（如果可用）。
否则，输出将延迟一个阶段。
- 对下一个轨迹点重复上述过程。当到达最后一点时，算法终止。

A.3.2 表现概率

对于在轨迹点附近找到的每个候选分段，我们用 1D 高斯函数对其观测概率
进行建模，如下所示。

$$p(\text{observation}) = \frac{1}{2w} \int_{-w}^w \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{(l-d)^2}{2\pi\sigma_g^2}} dl.$$

这里 w 是指道路 r 的半宽 ($w=0.5*r.w$), d 是指 t 和 r 间点到曲线之间的大圆距离, σ_g 是 GPS 误差的估计标准偏差。虽然 GPS 误差已知具有非高斯分布, 我们采用这种模型是因为它易于实现和在之前的工作被证明有效 [10], [15], [17]。我们的方法是不同的, 因为我们也考虑道路宽度。这将允许我们更好地在路段之间区分, 特别是在路口。

此外, 基于假设驾驶员不大可能大大超过速度限制, 我们引入超速的惩罚机制。目的是帮助区分从相同交叉点分支出的可能具有不同速度限制的紧密间隔的平行道路。我们注意到在这些情况下, 单独的位置测量不足以区分段落, 因为记录的轨迹点可能落在它们之间。我们定义惩罚函数 S 如下,

$$S(v_t, v_r) = \frac{v_r}{\max(0, v_t - v_r) + v_r}.$$

这里 v_t 指传感器记录的时间, 而 v_r 指该路段的限制速度。如果遵守了速度限制, 则 $\max(0, v_t - v_r) = 0$ 。(即惩罚机制不启动)。结合 (2) 和 (3), 我们定义表现概率, $p(t|r)$ 如下

$$p(t|r) = S(v_t, v_r)p(observations).$$

A.3.3 转移概率

令 i, j 分别表示归属于两个连续轨迹点 t_n 和 t_{n+1} 的一对候选段, 其中 $i \in A_n$, $j \in A_{n+1}$ 。我们将内插路径 $P_{i \rightarrow j}$ 定义为当从 i 行进到 j 时车辆最可能采取的段的序列。假设驾驶员选择最短路程的路 (路径距离短的被选择可能性高), 我们可以使用 A* 路径寻找算法找到插值路径 [21]。对于 K 个片段的给定序列, $P_{i \rightarrow j}$ 中的 (r_1, r_k) , 我们设计如下的两个评分函数,

方法 1: 距离差异函数 T 测量传感器推测的行进距离和插值路径长度之间的差异,

$$T(d_{i \rightarrow j}, D_{i \rightarrow j}) = \frac{|d_{i \rightarrow j} - D_{i \rightarrow j}|}{D_{i \rightarrow j}},$$

这里 $d_{i \rightarrow j} = \bar{v}_{i \rightarrow j} \Delta t$ 是车辆在时间间隔 Δt 以平均速度 $\bar{v}_{i \rightarrow j}$ 行驶的距离, 而 $D_{i \rightarrow j}$ 是 $P_{i \rightarrow j}$ 的路径长度。上面的方法 1 通过比较其长度与推测 (DR) 估计评估了假设路径 $P_{i \rightarrow j}$ 的可行性。如果 $P_{i \rightarrow j}$ 是真实路径, 差异会被假定为接近于零。

方法 2: 选用动量变化函数 M 测量车辆对于在 $P_{i \rightarrow j}$ 中采取的每个路段所引起的平均动量变化,

$$M(v_0, v_1, l_1, \dots, v_k, l_k) = \frac{\sum_i^K l_i \|v_i - v_{i-1}\|}{\bar{v}_{i \rightarrow j} \sum_i^K l_i},$$

其中 (v_1, \dots, v_k) 是 $P_{i \rightarrow j}$ 中每个段的车辆的速度矢量, 并且 $(l_1 \dots l_k)$ 是相应的段长度。我们假设矢量幅度随时间从 $|v_1|$ 线性变化到 $|v_n|$, 而它们的方向与段曲线平行。注意, 附加参数 v_0 是从先前转换的终端速度继承的初始速度矢量。通过类似的逻辑, 我们将使用 v_k , 当前终端速度作为下一转换中的初始速度, 等等。图 2 示出了该概念。上述措施 2 可以被描述为“平滑因子”, 其对由许多突然转弯组成的不可行转移做惩罚。

上面介绍的评分函数 T 和 M 为转变 $i \rightarrow j$ 提供两种不同的“适合度量”。这表明, 可以通过将这两个度量融合在一起来导出转移概率。我们使用标记为正确或不正确转换的实例训练支持向量机 (SVM) 分类器, 其中特征向量由测量 1 和测量 2 给出的分量分数组成。利用该分类方法, $P_{i \rightarrow j}$ 是输入得分组合属于“正确转换”类别的概率。我们将在第 4-B 节更详细地描述训练过程。

A.3.4 在线维特比算法

我们的目标是使用增量法找到全局地图匹配解决方案。这意味着该算法需要在不知道未来输入的情况下沿着马尔可夫链进行不可逆的在线决策, 同时确保部分解在组合时产生全局最优解。为了实现这一点, 我们应用在线动态规划来解决 (1) 中的递推关系。关键的见解是, 当当前幸存路径在马尔科夫链中的某点 (收敛点) 收敛时, 所有未来的幸存路径将包含直到收敛点的相同子路径。相关证据可在 [19] 和 [20] 中找到。

我们制定我们的 OHMM 算法的伪码如下。算法 1(OHMM 地图匹配) 递增地处理轨迹点, 并且在每个阶段, 它输出算法 2 返回的部分解, 如果可用的话。否则, 它给出一个空输出, 并引发一个延迟阶段。算法 2 (在线维特比算法) 检查

在解链中是否存在任何收敛点，并返回到该点的最大似然子序列 (如果有的话)。

方便地描述算法 1 和 2 在滑动窗口方面的工作原理。当在由窗口覆盖的马尔科夫链中的任何地方找到收敛点时，窗口随着新的轨迹点被处理并从后面收缩而向前扩展。注意，滑动窗口的大小可以根据状态空间的结构而变化，因此名称可变滑动窗口。图 3 示出了 **VSW** 的工作原理。

然而，**VSW** 的一个缺点是不能保证最坏情况的窗口大小；因此在极端情况下输出延迟可以是任意大的。我们通过设置窗口大小的上限来修改算法 1，使得当达到阈值时，算法将输出直到当前阶段的最大似然解。我们将标记这个修改的方法有界滑动窗 (**BVSW**) 方法。但是与 **VSW** 不同，这种方法可能导致次优解决方案。

A.4 实验装置

A.4.1 现场测试数据

使用支持 **GPS** 的智能手机，我们收集了在新加坡的 4 条预定的公共汽车路线的地面实况数据，如图 4 所示。由于我们关心地图匹配结果的路径精度（将在第 6-C 节中定义），因此对实际测试路径（地面实况路径）的了解足以使我们验证算法。为了比较不同环境条件下的表现，我们选择了涵盖新加坡农村和城市地区的 4 条路线。农村路线 (**R1** 和 **R2**) 涉及较少的转弯，并且主要包括通过开放区域的直线路线，例如高速公路。城市路线 (**U1** 和 **U2**) 除了更加高度分支外，还包括密集地包含高层建筑的城市街区。**R1**，**R2**，**U1** 和 **U2** 的长度分别为 36.3km，11.3km，27.3km 和 32.5km。此外，为了模拟变化的采样频率的轨迹，我们以 10 秒到 5 分钟的采样间隔对我们的原始数据（每 1-3 秒记录一次）进行二次采样。

A.4.2 训练和参数估计

参数 σ_g 需要在 (2) 中估计体现概率，并且转换概率需要 **SVM** 训练。

我们通过分析我们的地面实际数据的扰动来估计 σ_g 。对于每个轨迹点，我们计算该点离其最近路段中心的大圆距离。然后，基于距离的中值绝对偏差 (**MAD**) 计算标准偏差，

$$\sigma_g = 1.4826 \text{median}_i(|d_i - \text{median}_j(d_j)|).$$

这里 d_i 表示单个轨迹点与其匹配段之间的垂直距离。距离的分布允许我们估计地面真实路径周围的轨迹点的一维扰动。注意在 (7) 中, **MAD** 由常数因子 1.4826 缩放, 因为我们假设 **GPS** 测量误差是正态分布的。我们采用 **MAD** 方法因为它对抗能够更好地对抗数据集中的异常值。在 [7], [15] 中采用了相同的估计标准差的方法。基于整个数据集, 我们获得 $\sigma_g = 6.86\text{m}$ 。

为了推断转移概率, 我们使用 3,000 个标记的实例训练 **SVM** 分类器, 其中每个实例对应于正确的转换 (类别标记为 '1') 或不正确的转换 (类别标记为 '2')。每个实例是 **2D** 特征向量 (5) 和 (6) 计算得到的得分值, 并且两个分量被缩放为 [0,1]。缩放函数是 $(1+x)^{-1}$, 其中 x 是特征的任一分量。使用网格搜索参数空间和 5 重交叉验证, 我们发现参数的最佳组合为 $C = 0.25$ 和 $g = 0.5$, 其中 C 是软边际参数, g 是径向基函数 (**RBF**) 内核参数。训练结果如图 5 所示。

A.4.3 性能评价

我们将使用两个性能度量来评估我们的算法: 精度和输出延迟。

精度被定义为地面实际路径中正确匹配的轨迹点的分数。当轨迹点被映射到包含在地面实况路径中的任何道路段时, 记录正确的匹配。这种准确性的测量避免了惩罚“边界类”, 其中点位于路口的正中间。我们注意到, 精确地确定收集每个轨迹点的路段是不切实际的, 原因在于两个原因: (i) “边界类”可以归因于交叉点的任一出口上的路段, 以及 (ii) 可能的误校准的数字地图。因此, 路径精度测量是更合适的评估标准。

输出延迟是算法对每个轨迹点引起的平均输出延迟。它通过在获得匹配结果之前花费的时间来量化。

测试如下进行:

- 我们对农村和城市测试路线的不同采样间隔的测试轨迹执行地图匹配, 范围从 3 秒到 5 分钟。对于每个路由类别, 我们聚合两个测试路由获得的结果。
- 我们比较了三种定位策略在精度和输出延迟方面: **VSW**, **BVSW** 和 **FSW**。

对于 **BVSW** 和 **FWS**，测试了不同的窗口尺寸。对于每个窗口大小，我们聚集整组测试数据的结果（4 个测试路由，采样间隔在 10 秒到 5 分钟之间）。

A.5 结果

图 6 展示了城乡测试路线的地图匹配精度的比较。结果表明，除了大于 4 分钟的抽样间隔外，农村路线的准确性比城市路线的准确度大约为 5%。以小于 1 分钟的间隔，两条路线的精度都高于 0.9。在这两种情况下，精度随着采样间隔的增加而恶化。

图 7 和图 8 中，虚线表示使用 **VSW** 定位策略实现的最优结果。**BVSW** 方法在 $w = 8$ 及以上时收敛到 0.921 的最佳精度。在所有情况下，**FSW** 方法给出一致的较低精度，并且没有收敛到最优解，即使在 $w = 20$ 。

在图 8 中，**VSW** 的平均输出延迟为 82s。与 **FSW** 相比，它实现了显着更低的延迟，而没有折衷解决方案的最优性。使用 **BVSW**，在 4 和以上的窗口大小的延迟性能中没有显着的优点。这表明马尔可夫链中的大多数决策点发生在达到窗口界限之前。在 **FSW** 的情况下，延迟与窗口大小成比例地增加，但是在某个阈值点之后精度增益减小到几乎为零。

A.6 结论和未来工作

在本文中，我们描述了一种用于地图匹配的在线算法，并分析其在地面实况数据上的性能。我们设计了 **VSW** 和 **BVSW** 方法来寻找在线解决方案。两者在精度和输出延迟方面都优于先前基于 **HMM** 的算法中使用的传统 **FSW** 定位策略。我们还开发了一种数据驱动的方法，用于推断在地图匹配过程中融合传感器测量和拓扑信息的转换概率。总而言之，这些方法提供了用于设计基于在线 **HMM** 的地图匹配算法的一般框架，其适合于使用浮动车辆数据的实时应用。算法的其它变体可以在估计体现和转移概率时结合附加的传感器数据，例如加速度和高度测量。

对于未来的工作，我们可以探索地图匹配算法的设计与动态参数检测和适应不同的环境设置，如在城市或农村地区，其中 **GPS** 准确性可能会变化。传感器信息，例如用于 **GPS** 测量的精度（**DOP**）值的稀释可能被证明对于实现这个

目标是有用的。此外，我们建议更好的方法 [22] 内插轨迹点，而不是假设它们之间的最短路径。对实际行进路径的更好近似可以提高地图匹配精度。

在学期间参加课题的研究成果

个人简历

xxxx 年 xx 月 xx 日出生于 xx 省 xx 县。

xxxx 年 9 月考入 xx 大学 xx 系 xx 专业, xxxx 年 7 月本科毕业并获得 xx 学士学位。

xxxx 年 9 月免试进入 xx 大学 xx 系攻读 xx 学位至今。

发表的学术论文

- [1] Yang Y, Ren T L, Zhang L T, et al. Miniature microphone with silicon- based ferroelectric thin films. Integrated Ferroelectrics, 2003, 52:229-235. (SCI 收录, 检索号:758FZ.)
- [2] 杨轶, 张宁欣, 任天令, 等. 硅基铁电微声学器件中薄膜残余应力的研究. 中国机械工程, 2005, 16(14):1289-1291. (EI 收录, 检索号:0534931 2907.)
- [3] 杨轶, 张宁欣, 任天令, 等. 集成铁电器件中的关键工艺研究. 仪器仪表学报, 2003, 24(S4):192-193. (EI 源刊.)
- [4] Yang Y, Ren T L, Zhu Y P, et al. PMUTs for handwriting recognition. In press. (已被 Integrated Ferroelectrics 录用. SCI 源刊.)
- [5] Wu X M, Yang Y, Cai J, et al. Measurements of ferroelectric MEMS microphones. Integrated Ferroelectrics, 2005, 69:417-429. (SCI 收录, 检索号:896KM)
- [6] 贾泽, 杨轶, 陈兢, 等. 用于压电和电容微麦克风的体硅腐蚀相关研究. 压电与声光, 2006, 28(1):117-119. (EI 收录, 检索号:06129773469)
- [7] 伍晓明, 杨轶, 张宁欣, 等. 基于 MEMS 技术的集成铁电硅微麦克风. 中国集成电路, 2003, 53:59-61.

研究成果

- [1] 任天令, 杨轶, 朱一平, 等. 硅基铁电微声学传感器畴极化区域控制和电极连接的方法: 中国, CN1602118A. (中国专利公开号)
- [2] Ren T L, Yang Y, Zhu Y P, et al. Piezoelectric micro acoustic sensor based on ferroelectric materials: USA, No.11/215, 102. (美国发明专利申请号)

综合论文训练记录表

学生姓名		学号		班级	
论文题目					
主要内容以及进度安排	<div>指导教师签字：_____</div> <div>考核组组长签字：_____</div> <div>年 月 日</div>				
中期考核意见	<div>考核组组长签字：_____</div> <div>年 月 日</div>				

指导教师评语	<div>指导教师签字：_____</div> <div>年 月 日</div>
评阅教师评语	<div>评阅教师签字：_____</div> <div>年 月 日</div>
答辩小组评语	<div>答辩小组组长签字：_____</div> <div>年 月 日</div>

总成绩：_____

教学负责人签字：_____

年 月 日