



语言数据科学微专业人才培养方案

一、培养目标

在教育部新文科发展战略背景下，上海交通大学外国语学院积极响应国家人工智能，尤其是语言智能重大发展战略，顺应外国语言文学学科交叉与融合的趋势，精选符合国家社会需求与学科前沿的优势课程，强力打造“语言数据科学”微专业。

该项目致力于培养具备语言科学、认知科学、计算机科学等多学科知识，具有广阔国际视野、社会责任感和创新实践能力的语言数据科学拔尖创新人才。

二、规范与要求

按照《上海交通大学章程》提出“本校以精英教育的理念统领人才培养工作，培养具有创新精神和能力的高层次人才”之要求，外国语学院“语言数据科学”微专业具有课程设计对接国际前沿，优势学科交叉培养的特点，培养具有社会使命感和责任感的高端语言数据科学专业人才。微专业课程结课后，学生将具备以下核心能力：

1. 专业知识和理论基础：掌握语言数据科学的基本知识和概念；精通自然语言处理所涉语言数据科学的相关技术；熟悉各层次语言模型及其价值；了解统计学知识与语言数据科学的关联性。

2. 实践能力：利用自然语言处理技术评价文本（语言数据）的能力；为特定任务训练（微调）语言模型的能力；针对不同应用场景运用相关技术的融合能力。

3. 科研能力：掌握语言数据科学研究的基本方法，具有初步的科学研究和实际工作能力；能够有效检索和阅读相关领域的科研文献，撰写综述和研究报告；能够提出新问题和探索新方法的创新思维能力。

4. 职业素养：具备本领域持续学习和自我提升的能力和意识，能够跟踪学科前沿动态，不断更新专业知识和技能。

三、课程体系构成

本微专业课程设计采用模块化设置，分为三个模块，逐层递进，达到学生从技能基础知识到实践应用能力的提升培养。三个模块分别是：基础模块、进阶模块、应用模块，每个模块有相应的学分修读要求如下：

基础模块：旨在为学生打下坚实的语言数据学习基础，为后期的进阶学习和应用做准备。该模块包含4门必修课程，共6学分，分别是：编程与语言数据分析、语言数据统计与分析、语言数据挖掘与机器学习。

进阶模块：要求修满2学分。在基础模块课程学习基础上，进一步提升数据结构和计算能力，本阶段安排一门必修课：自然语言处理。

应用模块：本模块旨在强化学生语言数据知识的应用和提升，本模块是选修课模块，共开设 7 门课程，学生可以根据自身兴趣选修其中 4 门课程，需要修满 2 学分。

以下是模块内课程安排和课程简介：

1. 基础模块（6 学分）

①编程与语言数据分析

学分：2 学时：理论+实践 32 学时

②语言数据统计与分析

学分：2 学时：理论+实践 32 学时

③语言数据挖掘与机器学习

学分：2 学时：理论+实践 32 学时

2. 进阶模块（2 学分）

①自然语言处理

学分：2 学时：理论+实践 32 学时

3. 应用模块（2 学分）

①课程名称：语料库与翻译

学分：0.5 学时：理论+实践 8 学时

②课程名称：机器学习与翻译质量评估

学分：0.5 学时：理论+实践 8 学时

③课程名称：认知科学数据挖掘

学分：0.5 学时：理论+实践 8 学时

④课程名称：大语言模型与应用

学分：0.5 学时：理论+实践 8 学时

⑤课程名称：数字人文

学分：0.5 学时：理论+实践 8 学时

⑥课程名称：计算传播学

学分：0.5 学时：理论+实践 8 学时

⑦课程名称：计算法学

学分：0.5 学时：理论+实践 8 学时

四、学制、毕业条件与学位

学生完成以上模块学分要求，修满 10 学分后，予以结业，获得上海交通大学微专业结业证书。

五、语言数据科学微专业课程设置一览表（最低学分要求：10 学分）

序号	课程名称	学分	学时	学时分项		开课学期 (春/秋)	课程简介 (课程目标和核心 知识点等内容)	开课方式 (线上/线 下/混合式)
				理论	实践			
基础模块必修（6 学分）								
1	编程与语言数据分析	2	32	16	16	秋	讲授基础编程知识和技巧以及语言数据分析中的编程技能；编程基础知识、编程工具与技巧、数据采集与清洗、数据处理案例应用。	线上/线下/ 混合式
2	语言数据统计与分析	2	32	16	16	秋	讲授统计学的基本原理和方法，并将相关方法应用于语言数据处理过程之中；统计学方法及其适用性、数据检验。	线上/线下/ 混合式
3	语言数据挖掘与机器学习	2	32	16	16	春	讲授语言数据挖掘和机器学习的基本原理和方法以及机器学习的常见应用领域；数据预处理与特征工程、数据可视化、监督学习、无监督学习、模型评估。	线上/线下/ 混合式
进阶模块必修（2 学分）								
1	自然语言处理	2	32	16	16	春	本课程介绍自然语言处理的基本概念和任务，如分词、词性标注、句法标注、语义标注等，以及常见应用，如情感分析、阅读理解、自然语言推理等。同时介绍自然语言处理中的常见算法和模型。	线上/线下/ 混合式
应用模块选修（7 门课程中选 4 门，获得 2 个学分即可）								
1	语料库与翻译	0.5	8	4	4	夏	讲授语料库的相关知识、方法、技术和工具	线上/线下/ 混合式

							以及语料库与翻译的关系；语料库技术、语料库翻译、翻译技术。	
2	机器学习与翻译质量评估	0.5	8	4	4	夏	讲授翻译质量评估的基本原理与方法以及机器学习方法加持下的评估模式；翻译质量评估、人评与机评、机器学习算法。	线上/线下/混合式
3	认知科学数据挖掘	0.5	8	4	4	夏	讲授基于机器学习、统计学、数据库、可视化技术等，在认知科学理论的基础上，自动化地分析语言学和翻译学的语料与实验数据，作出归纳性的推理，从中挖掘出潜在的语言特征和认知行为特征。	线上/线下/混合式
4	大语言模型与应用	0.5	8	4	4	夏	本课程介绍大语言模型的历史及现状。主要讲解语言模型的基本原理及其背后的算法，同时介绍目前大语言模型研究的进展以及应用。	线上/线下/混合式
5	数字人文	0.5	8	4	4	夏	讲授数字人文的科学领域及其数字化资源研究对象和所依托的现代通讯技术、超媒体和数字图书馆、数智技术等基础性研究平台和技术手段，并以此作为方法论的共同基础；核心基础活动、远读+细读、人文概念+技术、舆情监测。	线上/线下/混合式
6	计算传播学	0.5	8	4	4	夏	讲授传播网络分析、传播文本挖掘、数据科学等分析工具，采用一种非介入的方式对人类传播行为数据进行大规模收集和分析，挖掘人类传播行为背后的模式和法则，探究其背后的生成机制与基本原理。	线上/线下/混合式
7	计算法学	0.5	8	4	4	夏	讲授计算法学历史沿革与发展展望、现实意义、实际应用；法律数据特征、计算模型、计算结果与法律裁量。	线下/混合式



上海交通大學
SHANGHAI JIAO TONG UNIVERSITY



上海交通大學外國語學院
School of Foreign Languages