算法所在论文对应任务标注规范

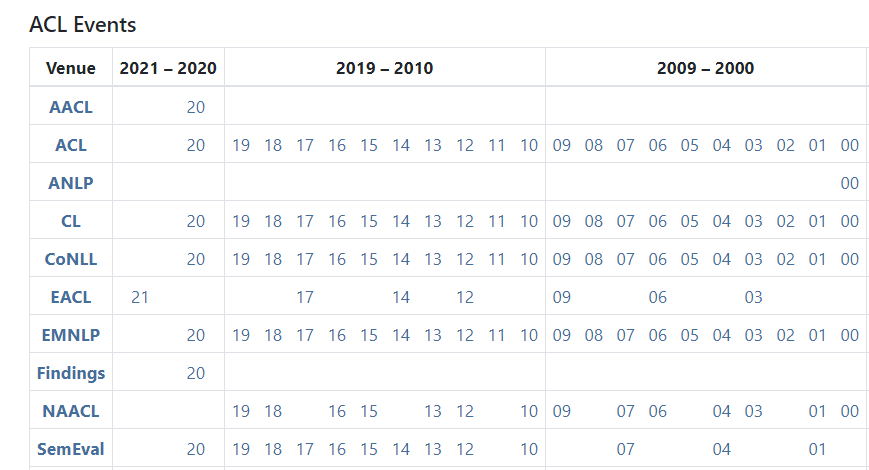
**1标注任务说明**

本标注任务主要针对ACL年会论文，对论文中解决的任务进行识别。任务的标注主要分为三步，首先从会议程序册中获取原始的任务标签，而后对高频单任务标签进行合并得到初始的自然语言处理领域任务体系。随后，基于此对低频的单任务标签进行了修正。最后，针对较为宽泛的标签（如机器学习、NLP应用等）进行人工标注。

**2任务类别获取**

**2.1 收集会议手册**

打开ACL Anthology官网（https://www.aclweb.org/anthology/），在首页上能够看到各个会议每年的相关信息，如图1所示。



**图 1 ACL官网主页**

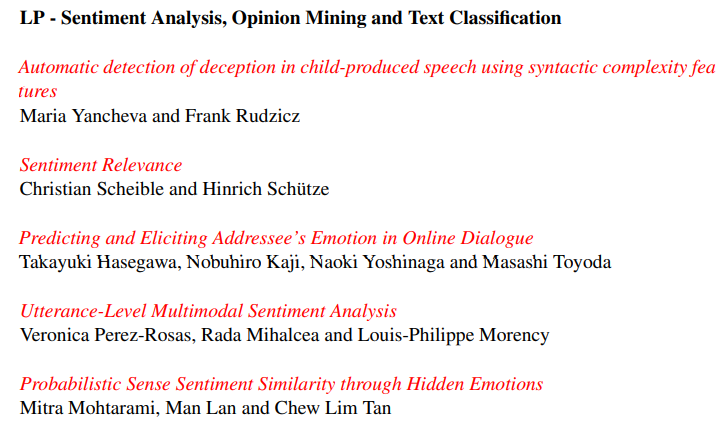
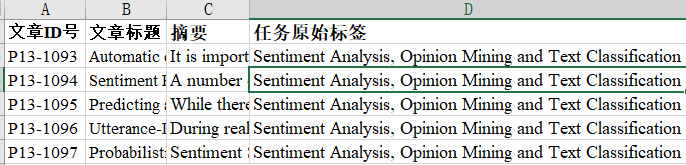
打开需要标注的ACL会议每一年网页后，最上方展示了该会议该年的所有部分。需要下载的会议手册为以上每一部分的第一篇文章（文档编号为X000），将会议所有年份的会议手册下载下来，会议手册文章样例如图2所示。



**图 2 会议手册示例**

**2.2 原始任务标签标注**

将所有会议手册下载下来之后，打开会议手册，找到手册中各论文的日程安排部分，各session的名称为该节所有文章的原始任务标签，本轮标注仅需复制所有文章的原始标签，无需对其进行任何更改。需要注意的是，不是所有年份会议手册的顺序与论文的顺序都相同，标注任务时需要看标题来一一确认。有的年份缺少Poster部分的任务标签，在表格中将单元格标黄即可。标注样例如图3所示。



**图 3 会议原始任务标注样例**

**2.3单任务标签的合并原则**

将原始标签标注完成后，将单任务标签提取出来，采用专家评审的方式利用频率较高的任务制定了一些单任务标签合并的规则，如表1所示。

**表1 专家提供的单任务标签合并规则**

|  |  |
| --- | --- |
| 原始标签 | 最终标签 |
| Machine Translation（机器翻译） | Machine Translation |
| Semantics（语义学） | Semantics |
| Information Extraction（信息抽取） | Information Extraction |
| Machine Learning（机器学习） | Machine Learning |
| Question Answering（自动问答） | Question Answering |
| Generation（生成） | 全称为Natural Language Generation，归入Natural Language Generation |
| Summarization（摘要） | Summarization |
| Parsing（语法分析） | Parsing |
| NLP Applications（NLP应用） | 过于宽泛，需要根据具体论文做具体的任务分类 |
| Sentiment Analysis（情感分析） | 与Sentiment Analysis and Opinion Mining合并，合并后为Sentiment Analysis and Opinion Mining |
| Dialogue and Interactive Systems（对话与交互式系统） | 与Dialogue合并，合并后为Dialogue |
| Discourse（话语分析） | Discourse |
| Resources and Evaluation（资源与评估） | Resources and Evaluation |
| Machine Translation and Multilinguality（机器翻译与多语言） | 与Machine Translation合并，合并后为Machine Translation |
| Dialogue（对话） | Dialogue |
| Social Media（社交媒体） | 与NLP for the Web and Social Media合并，合并后为NLP for the Web and Social Media |
| Information Retrieval（信息检索） | Information Retrieval |
| Language Resources and Evaluation（语言资源与评估） | 与Resources and Evaluation合并，合并后为Resources and Evaluation |
| Spoken Language + IR（口语+信息检索） | 与Information Retrieval合并，合并后为Information Retrieval |
| Multilinguality（多语言） | 与Machine Translation合并，合并后为Machine Translation |
| Sentiment Analysis and Argument Mining（情感分析与论元挖掘） | 与Sentiment Analysis and Opinion Mining合并，合并后为Sentiment Analysis and Opinion Mining |
| Phonology, Morphology and Word Segmentation（音韵学、形态学与分词） | 细分为Phonology、 Morphology 、 Word Segmentation三个任务，根据具体论文做判定 |
| Syntax（句法分析） | Syntax |
| Spoken Language Processing（口语处理） | Spoken Language Processing |
| Sentiment（情感分析） | 与Sentiment Analysis and Opinion Mining合并，合并后为Sentiment Analysis and Opinion Mining |
| Lexical Semantics（词汇语义学） | 归入Semantics |
| Vision, Robotics, Multimodal and Grounding（视觉、机器人、多模态等） | Vision, Robotics, Multimodal and Grounding等归入Natural Language in Multimodal and Multimedia Systems |
| Sentiment Analysis and Opinion Mining（情感分析与观点挖掘） | Sentiment Analysis and Opinion Mining |
| Document Analysis（文件分析） | 需要根据具体论文做具体的任务分类 |
| Speech（讲话） | Speech相关的合并到Spoken Language Processing（注意不要与Part-of-Speech Tagging混淆，这个特指词性标注） |
| Text Mining（文本挖掘） | 过于宽泛，需要根据具体论文做具体的任务分类 |
| Dependency Parsing（依存句法分析） | 合并到Parsing，合并后为Parsing |
| Interpretability and Analysis of Models（模型可解释性与分析） | 合并到Machine Learning |
| Word Sense Disambiguation（词义消歧） | Word Sense Disambiguation |
| Interpretability and Analysis of Models for NLP（NLP模型可解释性与分析） | 归入Machine Learning |
| Language Generation（语言生成） | 归入Natural Language Generation |
| Social Media and Computational Social Science（计算机社会科学与社交媒体） | 与NLP for the Web and Social Media合并，合并后为NLP for the Web and Social Media |

利用专家给出的合并建议，对剩下单任务标签进行合并与消歧，将结果保存到新的一列单元格中。剩余标签主要有以下两种标注类别：

（1）仅写法不同

不同会议、不同年份对同一个任务的描述方法不同，这些任务可以合并，如“Resources & Evaluation”标注为“Resources and Evaluation”

（2）表达方式不同

根据专家标注意见进行消歧，有不确定的地方两位标注者通过讨论确定所属标签。如“Evaluation Methods”标注为“Resources and Evaluation”。

通过人工标注与讨论后，所有单任务标签都被标注完成，根据这些标签整理出任务体系，应用于下一步宽泛标签的标注中。

**3任务类别标注**

由于部分标签涵盖范围较为广泛、标签与标签之间划分的边界较为模糊，如机器学习、文本挖掘、自然语言处理应用等。因此在将单标签任务合并之后，需要将这类宽泛标签中的论文再做细分标注，标注原则和标注Tips如下所示：

**3.1 宽泛标签标注原则**

（1）对于标签为“Machine Learning”、“Machine Learning Models and Methods”、“Statistical and Machine Learning Methods”、“Statistical and Machine Learning Methods in NLP”等机器学习方法相关的论文，需要根据论文的题目与摘要，确定论文的中心是提出的机器学习方法还是应用机器学习技术解决自然语言处理领域具体问题，若为前者则将论文任务标注为“Machine Learning”，若为后者则需要根据任务体系与论文解决的具体任务对论文任务进行标注。

（2）对于标签为“Machine Learning-Classification”、“Machine Learning: Embeddings”等机器学习具体应用的论文，需要根据任务体系与标签中论文应用的具体任务对论文任务进行标注。

（3）对于标签为“NLP applications”、“Natural Language Applications”、“Natural Language Processing Applications”、“NLP Apps”等自然语言处理应用相关的论文，需要根据论文的题目与摘要判断论文应用的具体任务，并根据任务体系进行标注。若具体任务数量极少且并没有在任务体系中，如自动评分，则保留“NLP applications”作为标注的结果。

（4）对于标签为“Text Mining”的论文，需要根据论文的题目与摘要判断论文中的文本挖掘具体任务，若为文本分类任务，则标注为“Text Categorization”；若为其他任务，则根据任务标注体系进行标注，标注结果不保留为“Text Mining”。

**3.2 宽泛标签标注Tips**

（1）论文任务为语法纠错等与语法相关的论文应标注为“Grammar”。

（2）论文标签为“Text Representation”的论文与论文主题为“Representation Learning”统一标注为“Representation Learning and Text Representation”。

（3）遇到论文任务与多个任务都比较相关的情况，标注结果为论文最主要的任务。

（4）论文任务为主题模型，需判断重点是模型还是应用。重点为模型应标注为“Machine Learning”；重点为应用应标注为“Topic Segmentation and Recognition”。