

# 基于多特征和分类器的用户意图领域分类系统

李大宇<sup>1</sup>, 张鹏<sup>1</sup>, 张文跃<sup>1</sup>, 王素格<sup>1,2</sup>

(1. 山西大学 计算机与信息技术学院, 太原 030006;

2. 山西大学 计算智能与中文信息处理教育部重点实验室, 太原 030006)

wsg@sxu.edu.cn

**摘要:** 随着互联网的飞速发展, 人机对话技术成为学术界和工业界的研究热点。用户意图领域分类是其中的一项重要研究任务, 其性能直接影响到特定领域的人机对话质量。本文针对用户意图领域分类的任务, 从特征对多类别间的区别能力的角度, 讨论了多种特征以及多种分类器对其分类效果的影响。在此基础上, 设计了一种多特征的用户意图领域分类系统, 并在 SMP2017 的中文人机对话技术评测 (ECDT) 任务一中取得了较好的成绩。

**关键词:** 人机对话, 用户意图领域分类, 多特征, 分类器, SMP2017

**Abstract:** With the rapid development of Internet, Chinese human-computer dialogue technology has become the focus of academic and industrial research. Domain classification of user intention is one of the most important research tasks, and its performance has a direct impact on the quality of human-computer dialogue in specific domains. Aiming at the task of domain classification of user intention, this paper discusses the influence of multiple features and multiple classifiers on classification performance. On this basis, a multi-feature domain classification of user intention system is designed, and achieved good results in The Evaluation of Chinese Human-Computer Dialogue Technology (ECDT) of SMP2017.

**Key words:** Human-Computer Dialogue Technology, Domain classification of user intention, multi-feature, classifier, SMP2017

## 1 引言

随着互联网的飞速发展, 人机对话技术成为学术界和工业界的研究热点。对于学术界而言, 人机对话技术对语音识别、自然语言处理、信息检索以及情感分析等领域有着重要的推动作用; 对于工业界而言, 人机对话技术是个人事务助理、虚拟情感陪护机器人以及娱乐性聊天机器人等产品的核心技术, 其性能直接影响了产品的市场竞争力。

第六届全国社会媒体处理大会 (The Sixth China National Conference on Social Media Processing) 举办了中文人机对话技术评测 (ECDT)。本次中文人机对话技术评测分为两个任务, 任务一为用户意图领域分类, 任务二为特定域任务型人机对话在线评测。其中任务一旨在正确识别用户所感兴趣的领域, 在用户与机器的交互中, 存在多种意图, 不同的意图会触发人机对话系统中的多个领域 (domain), 包括任务型垂直领域 (如查询机票、酒店、公交车等)、知识型问答以及闲聊等。因而, 人机对话系统的一个关键任务就是正确地将用户的输入分类到相应的领域 (domain) 中, 从而返回正确的回复结果。

我们山西大学文本挖掘团队参加了本次 ECDT 评测的任务一: 用户意图领域分类。用户意图领域分类的主要任务为, 正确判断用户所描述的问题属于哪个领域。例如, “你好啊, 很高兴见到你!” 为 “闲聊类”, “我想订一张去北京的机票。” 为 “任务型垂类 (订机票)”, “我想找一家五道口附近便宜干净的快捷酒店” 为 “任务型垂类 (订酒店)”。其中 “任务型垂类” 细分为了 30 个垂直领域。任务一评测又细分为封闭性测试和开放性测试, 封闭性测

试是指，仅使用主办方提供的训练集和开发集，开放性测试是指可以使用除主办方提供的数据外的任何数据。

针对 ECDT 任务一，以往的研究大都将其形式化为文本多分类任务，但是由于特征以及分类器的多样性，采用不同的特征以及不同的分类器效果各不相同，如何将不同的特征及分类器进行有效的组合使 ECDT 任务一效果达到最佳成为亟待研究的课题。本文提出了一种融合多语言粒度特征的用户意图领域分类系统，并且比较了不同特征组合以及不同分类器对结果的影响，并在 ECDT 任务一的测试集中取得了较好的成绩。

## 2 用户意图领域分类方法

本节主要介绍中文文本分类的一些关键技术对用户意图领域分类的影响，主要包括预处理、特征选择以及分类器等。

### 2.1 预处理

中文文本语料与英文语料不同，词与词之间没有用空格进行分割，因此中文文本在进行分类之前，一般都要经过一些预处理，包括分字或分词、去停用词等，但这些预处理是否会使用户意图领域分类效果提升尚不明确，在实验部分我们对其进行验证。

### 2.2 文本向量表示

为了使分类器能够识别文本数据，我们需要将文本数据表示成向量。文本向量表示主要包括特征选择和特征权重两部分。

我们采用的特征选择方法为传统的 N-gram 特征，分别测试了 1-gram、2-gram、3-gram 以及 4-gram 特征对分类效果的影响。

特征权重的计算方法主要有：布尔权重（boolean）、词频权重（Term Frequency, TF）、词频-逆文档频率（Term Frequency – Inverse Document Frequency, TF-IDF）。布尔权重向量的每一维上的值为 0 或 1，若这篇文档中含有这个词，则在这个维度上的值为 1，反之为 0。词频权重向量的每一维上的值为文档中这个词的出现次数。词频-逆文档频率权重向量的计算公式如下：

$$\text{TF-IDF} = \frac{\text{TF}}{\text{IDF}} = \frac{\text{TF}}{\log\left(\frac{|\text{D}|}{1 + |\{d \in \text{D} : t \in d\}|}\right)} \quad (1)$$

其中 TF 表示词频，|D|表示文档的总个数， $|\{d \in \text{D} : t \in d\}|$ 表示词在多少个文档中出现过。

### 2.3 分类器

本系统采用了常见的三种文本分类器：朴素贝叶斯分类器、决策树分类器、SVM 分类器。

## 3 实验结果与分析

本次评测提供的数据共包含 31 个领域，具体包括聊天类（chat）和垂类（30 个垂直领域），垂直领域有 app、bus、calc、cinemas、contacts 等。数据分为训练集和开发集，训练集包含 2299 个文档，开发集包含 770 个文档，总共 3069 个文档。

为了验证不同特征对用户意图领域分类结果的影响，本次实验采用控制变量法，通过改变单个因素并且固定其他因素来观察这个单一因素对结果的影响。实验前我们假设这些因素可能对实验结果产生影响：分词/分字、去停用词、分类器、特征、特征权重。评价指标采用准确率（Accuracy），宏平均精确率（Precision），宏平均召回率（Recall），宏平均 F 值（F-measure）。

### 3.1 分词/分字及去停用词

分词以及去停用词是中文文本预处理的常用方法，本节考察分词以及去停用词对结果的影响。分类器采用 Lib-SVM 分类器，特征采用 1-gram，特征权重采用 Boolean 权重。实验结果如表 1 所示。

表 1 分词/分字及去停用词结果表

Model		Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
分字	去停用词	88.96	90.76	<b>88.12</b>	88.89
	未去停用词	<b>89.35</b>	<b>91.72</b>	88.10	<b>89.34</b>
分词	未去停用词	82.46	87.07	78.38	81.30

由表 1 可知：

(1) 在分字的情况下，去停用词与否的实验结果可以看出，去停用词后，结果略低于不去停用词的结果。主要原因是停用词中的标点符号等对类别的判断有指导作用，比如问号是询问类的主要特征之一，去掉停用词降低了分类器对类别的区分能力。

(2) 分词采用 python 常用分词工具包结巴分词，对比分词和分字的实验结果可以看出，虽然分词为中文文本预处理中常用的步骤，但是分词的文本分类结果明显低于分字的分类结果。这是由于分词的过程中产生了较多的分词错误造成的，这对人机对话类的短文本影响较大，若短文本中的类别关键词出现分词错误，则直接会导致短文本的分类错误。

后续的实验部分均采用分字以及未去停用词。

3.2 分类器

这一部分我们讨论各分类器对分类结果的影响，特征仅选用了 1-gram，特征权重采用 Boolean 权重进行对比实验。实验结果如表 2 所示。

表 2 各分类器结果表

Model	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
朴素贝叶斯	85.84	87.94	81.32	83.45
决策树	83.39	79.84	80.92	80.90
Lib-SVM	89.35	91.72	88.10	89.34

从表 2 中可以看出，Lib-SVM 分类器的文本分类结果优于朴素贝叶斯分类器和决策树分类器。由于先验知识的不足，使得朴素贝叶斯分类器降低。而在决策树构造主要依赖选取的特征，当特征较多时，产生的树的分支也较多，导致分类的性能下降。后续的实验部分均采用 Lib-SVM 分类器。

3.3 特征选择

这一部分我们讨论特征选择对结果的影响，权重采用 TF-IDF。实验结果如图 1 所示。

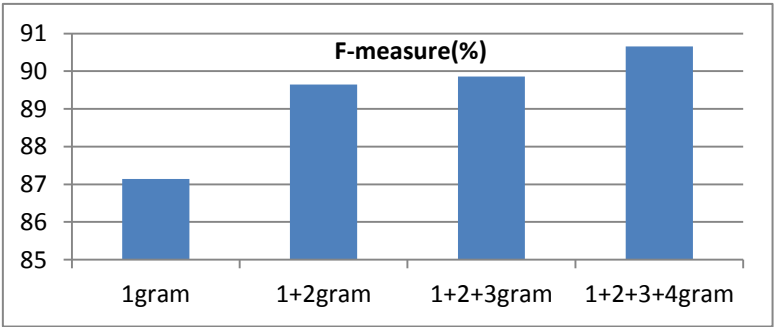


图 1 不同特征结果对比图

从图中可以看出，随着特征的增加，F 值也逐渐增加。这是由于特征的增加使得文本的向量表示更加全面，1+2+3+4gram 特征的效果最佳，其向量维度也最大，为 2642。在后续的实验部分均采用 1+2+3+4gram 特征。

3.4 特征权重

这一部分我们讨论特征权重对结果的影响实验结果如表 3 所示。

表 3 各特征权重结果表

Model	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
Boolean	89.35	91.86	88.50	89.61
TF	89.87	91.98	89.41	90.30
TF-IDF	90.51	93.14	89.07	90.66

从表 3 中可以看出，采用 TF-IDF 权重的效果最佳，这是由于 TF-IDF 权重综合考虑了词频以及文档频率，相较于 Boolean 权重和 TF 权重更加全面的刻画了一个词的重要程度。

### 3.5 一对多 SVM 分类器

在上述实验中，我们将分类任务当作文本多分类来直接放入 SVM 分类器中进行分类器，而 SVM 分类器擅长二分类任务，因此我们将多分类任务转化为一对多分类任务，将某一类当成正类，而将其他类当成负类进行训练。实验结果如表 4 所示。

表 4 多分类 SVM 与一对多 SVM 对比结果表

Model	Accuracy(%)	Precision(%)	Recall(%)	F-measure(%)
多分类 SVM	90.51	93.14	89.07	90.66
一对多 SVM	91.16	93.64	89.64	91.23

从表 4 中可以看出，一对多 SVM 分类效果优于多分类 SVM。这是由于 SVM 分类器更加擅长二分类任务，尽管将多分类任务转化为二分类任务会增加二分类的不平衡性，但是其效果仍然由于多分类 SVM。

### 3.6 最终评测结果

最终的用户意图领域分类评测系统采用的特征为：分字并且未去停用词，特征采用 1+2+3+4gram，权重采用 TF-IDF，分类器采用一对多 Lib-SVM。

在最终的封闭剂测试中，我们取得了第 3 名，最终结果为 F 值 0.9089。

## 4 总结与展望

本文针对 SMP2017 人机对话评测任务一，提出了一个基于多特征和分类器的用户意图领域分类系统。系统通过结合不同特征、特征权重以及分类器等，不断提升系统的分类精度，最终在封闭测试阶段取得了第三名的好成绩。

本系统中融入的特征以及分类器等都是基础传统的文本多分类框架，在未来的工作中，我们希望加入深度学习等技术来解决用户意图领域分类任务。

## 参考文献

- [1] Chang C C, Lin C J. LIBSVM: A library for support vector machines[M]. ACM, 2011.
- [2] Mohammad S M, Kiritchenko S, Zhu X. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets[J]. Computer Science, 2013.
- [3] Aida-Zade K, Rustamov S, Mustafayev E, et al. Human-computer dialogue understanding hybrid system[C]// International Symposium on Innovations in Intelligent Systems and Applications. IEEE, 2012:1-5.
- [4] 魏松. 人机对话系统中若干关键问题研究[D]. 北京邮电大学, 2007.
- [5] 侯晓兵. 基于自然语言的人机对话系统的研究[D]. 北方交通大学 北京交通大学, 1998.