

NLPIR深度机器学习的文本分类

产品名称	NLPIR深度机器学习的文本分类
公司名称	灵玖中科软件（北京）有限公司
价格	.00/个
规格参数	
公司地址	北京市海淀区苏州街49-3号5层11号
联系电话	010-62648216

产品详情

近年来，随着Internet的迅猛发展，网络信息和数据信息不断扩张，如何有效利用这一丰富的数据信息，已成为广大信息技术工作者所关注的焦点之一。为了快速、准确的从大量的数据信息中找出用户所需要的信息，文本信息的自动分析也成为了当前的迫切需要。对文本信息的分析中的一个主要技术就是文本分类。

文本分类问题是自然语言处理的一个基本问题，很多相关的研究都可以归结为分类问题。文本分类是将文本按一定的规则归于一个或多个类别中的技术。近年来，许多统计的方法和机器学习的方法都应用到文本分类方面。

文本分类是指在给定的分类体系下,根据文语义元是统计语义方法中的原子,是不可分本的内容自动确定文本类别的过程.当前的文本割的最小单位,在文本分类中语义元是词;

文本分类一般包括了文本的表达、分类器的选择与训练、分类结果的评价与反馈等过程，其中文本的表达又可细分为文本预处理、索引和统计、特征抽取等步骤。文本分类系统的总体功能模块为:

- (1) 预处理：将原始语料格式化为同一格式，便于后续的统一处理;
- (2) 索引：将文档分解为基本处理单元，同时降低后续处理的开销;
- (3) 统计：词频统计，项(单词、概念)与分类的相关概率;
- (4) 特征抽取：从文档中抽取反映出文档主题的特征;
- (5) 分类器：分类器的训练;
- (6) 评价：分类器的测试结果分析。

NLPIR采用基于内容的文本自动分类过滤和基于规则的文本分类过滤分类，并采用深度神经网络对分类体系进行了综合训练。能够进行多级分类，分类速度每秒100篇以上，平均准确率90%以上，能够进行中英文分类和中英文的混合分类。用户可以灵活、方便的更换模板，来实现对不同的主题的分类过滤。

文本过滤功能能够从大量文本中快速识别和过滤出符合特殊要求的信息，可应用于品牌报道监测、垃圾信息屏蔽、敏感信息审查等领域。

NLPIR深度文本分类，可以用于新闻分类、简历分类、邮件分类、办公文档分类、区域分类等诸多方面。此外还可以实现文本过滤，能够从大量文本中快速识别和过滤出符合特殊要求的信息，可应用于品牌报道监测、垃圾信息屏蔽、敏感信息审查等领域。

目前，对大量的文本进行自动分类处理已然成为信息检索、自然语言处理、数据库、人工智能等领域的热点;文本分类技术已成为一项具有较大实用价值的关键技术，主要体现在以下领域：信息检索、Web文档自动分类、数字图书馆、自动文摘、分类新闻组、文本过滤等等。