

基于主题模型的法院文本典型案例推荐

吕 宾,侯伟亮

(北京华宇信息技术有限公司, 北京 100084)

摘要: 考虑到案件不能量化、业务法官经验局限以及相似案件的认定缺乏统一的标准和尺度等因素,提出一种基于主题模型的法院文本典型案例推荐模型,基于本模型从相似度最大的文本中利用正则和加权的方法根据法院文本中判决年限和对应刑法给出推荐.同时对案件进行深度挖掘和分析,将相似度最高的案件定为最优推荐案件推荐给用户.实验结果表明,提出的基于主题模型的文本相似度计算方法跟传统的方法相比能够取得更好的 F 值.

关键词: 主题模型 LDA; 多粒度文本特征提取; 文本相似度; 法院判决文本

中图分类号:

文献标识码: A

文章编号: 1000-7180(2018)02-0128-05

Typical Case Recommendation of Court Texts Based on Topic Model

LV Bin,HOU Wei-liang

(Beijing Huayu Information Technologys Co., Ltd, Beijing 100084, China)

Abstract: Taking into account that the case cannot be quantified, the limited experience of business judge, the lack of unified standards for the identification of similar cases and other factors, a typical model of court texts based on topic model is proposed in this paper. The recommendation is given by regular and weighted method according to the number of years of judgment and corresponding criminal law from the maximum similarity text in court texts based in this model. While the highest similarity of the case as the best recommendation case to the user. Experimental results show that the method based on topic model proposed in this paper is superior to the traditional text similarity measure in F-metrics.

Key words: topic model LDA; multi-granularity text feature extraction; text similarity; court texts

作者简介:

吕 斌 男, (1973-), 硕士, 副高级工程师, 研究方向为信息法律知识认知技术和法律知识图谱构建.

E-mail: lb@thunisoft.com.

侯伟亮 男, (1984-), 学士, 研究方向为以大数据为中心的多元诉讼服务技术和产品开发.