文本挖掘技术被广泛运用在各个学科中，用于处理大批量数据。现在，对于文本挖掘，通用的定义为：文本挖掘的过程即是从文本数据中挖掘内在的，未知的，且有用的模型的过程（Tan，1999）。因此，文本挖掘技术可以在大批量数据中高效提取出有用的信息。

文本挖掘分析技术包括文本结构分析、文本摘要、文本分类、文本聚类、文本关联分析、分布分析和趋势预测等（袁军鹏等，2006）。郭金龙等（2012）学者认为在人文社科领域中较为常用的文本挖掘技术是文本分类技术和文本聚类技术。文本分类技术属于有监督的机器学习应用，采用的常见技术有：决策树、朴 素贝叶斯( NB) 、支持向量机( SVM) 、K - 近邻等，可以用于对主题的分类、对风格的分类、对情感的分类以及文章的题材等分类。在文本分类技术中。然而文本分类技术的最终结果与所采取的停用词等有很强的关系，而且自动化程度不高。而文本聚类技术属于无监督的机器学习应用。文本聚类技术虽然也需要分词，但不需要重复复杂的训练，自动化程度较为文本分类技术更高。

文本聚类技术中，用于挖掘潜在语义知识的模型有：LSA，PLSA以及LDA等模型。董婧灵（2011）等通过比较研究，认为LDA模型有着较为突出的优点：首先，LDA是完全概率生成模型，具有丰富的结构，成熟的算法和训练模型；其次，LDA模型更适合在大规模语料库中构建模型。LDA模型是由Blei等（2003）提出的,是一个集合概率模型。近十几年来，LDA模型在实际运用上都有广泛的讨论。例如：唐晓波（2014）等将LDA模型运用在微博热点的搜集。然而，这些文献较少涉及到微信平台数据的使用，以及将其所得模型运用到实际问题中。因此本文将运用LDA模型去研究微信娱乐圈公众号数据，并将其运用至广告营销中。

参考文献：

[1]Tan A H. Text mining: The state of the art and the challenges[C]//Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases. 1999, 8: 65-70.

[2]Blei, D., Ng, A., and Jordan,M. (2003). Latent Dirichlet allocation. Journal of Machine Learning Research, 3:993–1022.

[3]袁军鹏, 朱东华, 李毅, 等. 文本挖掘技术研究进展[J]. 计算机应用研究, 2006, 2: 1-4.

[4]郭金龙, 许鑫, 陆宇杰. 人文社会科学研究中文本挖掘技术应用进展[J]. 图书情报工作, 2012, 56(8): 10-17.

[5]董婧灵, 李芳, 何婷婷. 基于 LDA 模型的文本聚类研究[J]. 孙茂松, 陈群秀. 中国计算语言学研究前沿进展 (2009r 2011). 北京: 清华大学出版社, 2011.

[6]唐晓波, 向坤. 基于 LDA 模型和微博热度的热点挖掘[J]. 图书情报工作, 2014, 58(5): 58-63.