



华南理工大学
South China University of Technology

大数据分析视角下广府美食旅游感知价值研究



作者：李钰潇

目录

CONTENTS

01

选题背景与内容

02

研究思路

03

实证分析与成果

04

研究技术难点

05

论文总结



华南理工大学
South China University of Technology



1.1 研究背景

研究背景概述

时政背景

2017年中共的十九大报告指出要“不断满足人民日益增长的美好生活需要”，2018年国务院办公厅提出“推动美食、购物、演艺、游乐等产业与旅游业融合，**打造高品质的文化体验旅游**”

美食旅游良好的发展态势

2019年全国餐饮收入增速高于社会消费品零售总额增幅1.4个百分点；餐饮规模占社会消费品零售总额的比重为11.3%；对**社会消费品零售总额增长贡献率为13.1%**^[1]

千年广府、粤菜名城-广州

2020年广东省餐饮收入排行全国第二，是名副其实的“吃货大省”。2019年广州接待过夜旅行者人数达到6773.15万人次，旅游业创收4454.59亿元，同比增长11.1%^[2]

[1].数据来源：《中国饭店业协会2019年餐饮收入统计》。受新冠疫情影响，2020年数据具有一定特殊性，故本文主要参考2019年度数据报告。

[2].数据来源：《国家统计局2019年餐饮业消费情况回顾》

1.2 研究意义

行业意义

对于打造高品质的文化体验旅游来说，探索游客美食感知价值构成要素对回答“**怎样的美食旅游是人民大众需要的**”、“**如何打造高品质美食文化旅游**”具有重要意义。

学术意义

表2-1 学者感知价值维度划分总结

第一作者	研究角度	感知价值维度划分
Zelthaml	经济效用	收益、成本
Perval	消费者行为学	情感、社会、质量、经济
Petrick	当地一家餐厅	质量、情绪反应、声誉.....
王永贵	服务业	情感价值、感知利失.....
王朝晖	2010年世博会	服务价值、美感价值.....
保继刚	西双版纳傣族泼水节	文化认知价值、情感价值.....
李东伟	中华老字号餐厅	功能价值、认知价值.....

进步性

学界对感知价值的定义从经济效用价值的一维视角逐渐向**多维视角**转变

局限性

对美食这种旅游产品的**特征关照不足**，缺乏区域美食与地方美食差异性对感知价值影响实证研究

研究方法**以问卷调查、现场访谈**为主，采问题的设计在很大程度上会左右最终答案；对文本的**挖掘深度不足**，通常只停留在关键词提取和词频排序，



1.3 研究内容与创新之处



研究内容

- (1) 广府美食旅游感知价值的构成要素，并基于此构建新的美食旅游感知价值量表；
- (2) 游客对广府美食的**整体印象**和集中反映的问题。
- (3) 各价值感知要素对**游客美食体验满意度**的感知结果和原因。

创新之处

- 1 本文利用**词向量工具**，将文本映射到低维向量空间，通过向量距离度量词语语义联系，**为低层次语义信息的深度挖掘提供可能**
- 2 word2vec工具结合k-means聚类的大数据研究方法在探究旅游地意象、股票市场预测等方面已经得到大量运用和实践，但在探究**美食旅游价值感知研究方面尚未有先例**

点都德(聚福楼店) Dian Dou De (Ju FuLoudian)



4.5分
2092条评价

店面中式装修，一层等位区，二楼人。一行六人点了10几份点心
来自 彼岸之澄 的点评

携程、去哪儿网的美食版块
粤菜餐厅游客评论

网络爬虫、数据预处理

数值类数据: 广州热门粤菜和粤菜
品牌、美食旅游目的地等

统计不同餐厅的评论
总数，以评论数量作为
热门程度标准进行
排序

地理坐标类数据: 热门美
粤菜品牌的空间集聚情况

基于餐厅地理位置和评论
数量进行Arcgis核密度分
析

短文本类数据

广府美食旅游感知价值整
体框架与构成要素

游客对广府美食整体印象
和感知价值各维度感知结
果，以及影响游客美食体
验满意度的对象

广府美



口味: 5 环境: 4 服务: 4

店面中式装修，一层等位区，二楼老式的粤式酒楼风格。大圆桌一桌桌大都是当地人。一行六人点了10几份点心。服务员在旁边说，你们吃不掉吧，少点一点。但其实，你低估我们了。最后几乎全部吃光了啊。滚菜单最适合我了。虾饺大颗Q弹虾仁在里面，虽然虾饺皮有点软，口感不足，但虾仁也算加分。红米肠 这个蛮喜欢，分量足，肠粉皮很Q，外软里脆。青芥三文鱼挞 虽然用了cheese，但一点都不觉得腻。芥末和三文鱼组合加上酥松蛋挞皮，口感和谐又美妙。沙爹酱蒸金钱肚 金钱肚不够软烂，也不够入味。流沙包 说是金奖。然而金沙馅料差了那么点意思。豉汁蒸排骨 也是

数据收集时间截止至2021年4月2日，共69880条，每条数据包括**餐厅名称、餐厅位置、用户ID、餐厅评分、评论时间、评论内容**。剔除区域外、手工识别非粤菜类餐厅和不完整数据后剩下**59936条**



注：地图位置标注仅供参考，具体情况以实际道路标示信息为准。

1. 热门美食品牌、美食旅游目的地识别

表3-6 评论数量排名前20的餐厅

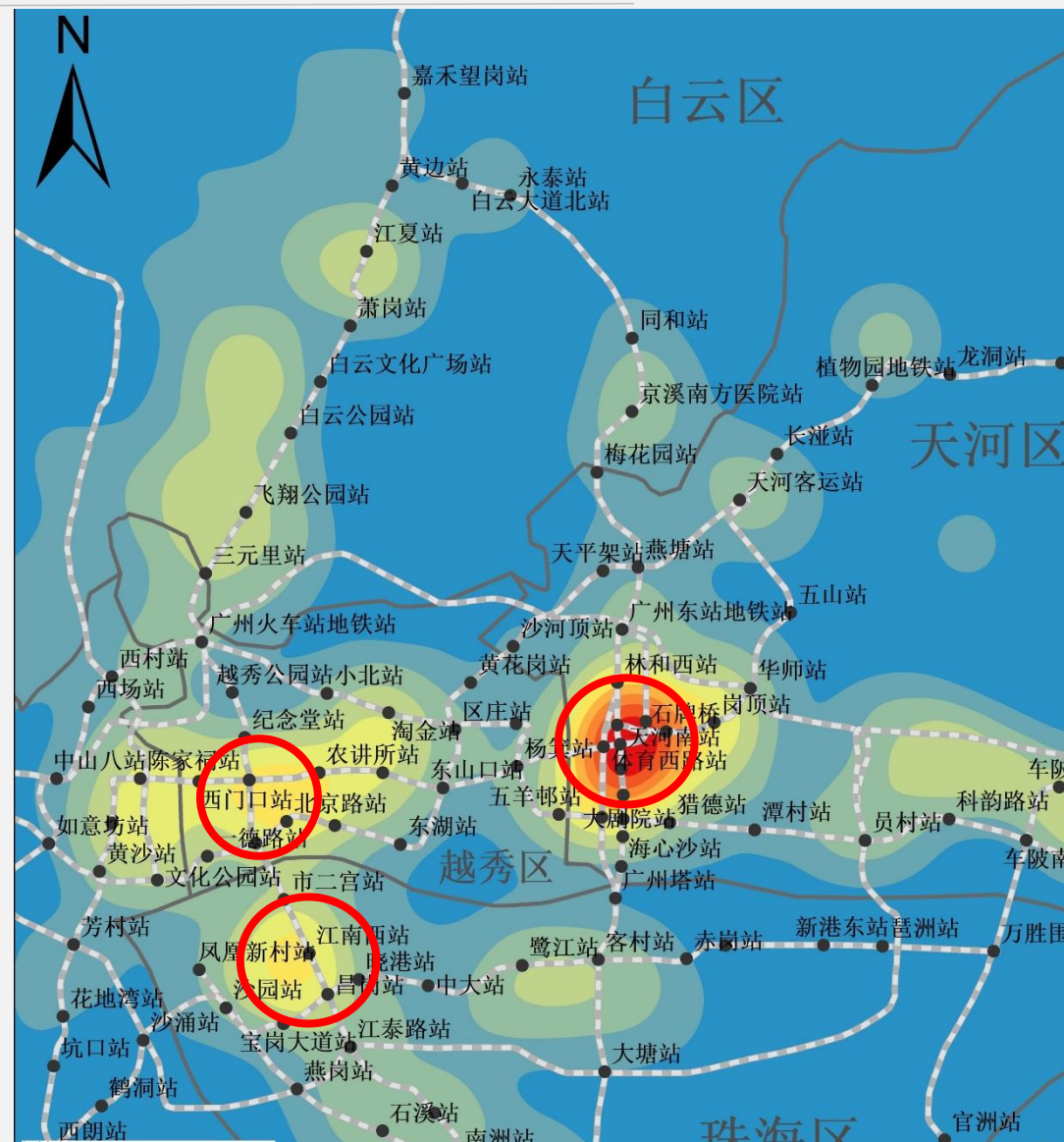
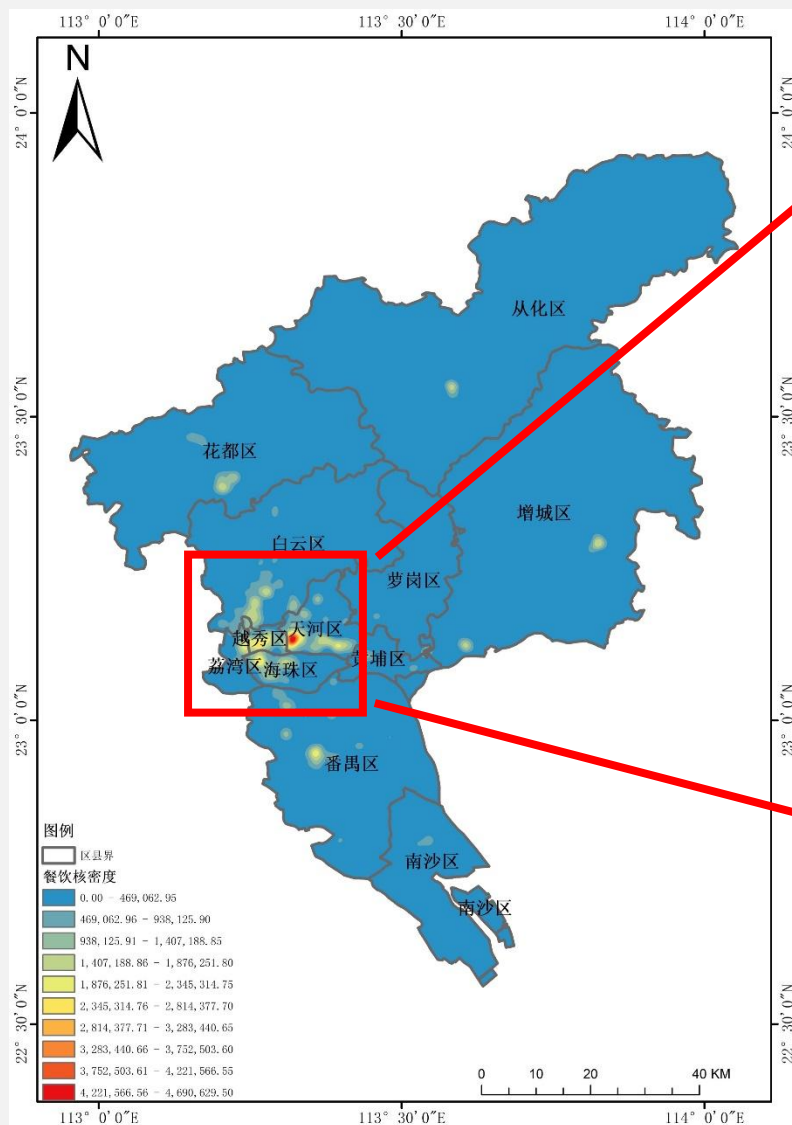
	餐厅名称	携程评论数量	去哪儿网评论数量	评论总量
1	点都德(聚福楼店)	2080	1922	4002
2	广州酒家(文昌店)	1597	1678	3275
3	点都德(喜粤楼店)	1320	1267	2587
4	点都德(北京路壹店)	982	950	1932
5	点都德(花城店)	900	848	1748
6	点都德(大茶楼店)	795	772	1567
7	陳添記(十五甫三巷店)	0	1566	1566
8	炳胜品味(海印总店)	853	703	1556
9	广州酒家(体育东路店)	808	744	1552
10	南信牛奶甜品专家	0	1533	1533
11	点都德(汇胜楼店)	1266	0	1266
12	银记肠粉(上九路店)	0	1228	1228
13	点都德(汇点楼店)	1216	0	1216
14	翠华餐厅(正佳广场店)	995	0	995
15	纯再餐厅(西城都荟店)	988	0	988
16	点都德(德粤楼)	921	0	921
17	宝华面店	0	849	849
18	荔银肠粉(宝华路店)	0	847	847
19	卡朋西餐(较场西路店)	0	825	825
20	广九餐室(中山三路店)	761	0	761

分析与结论

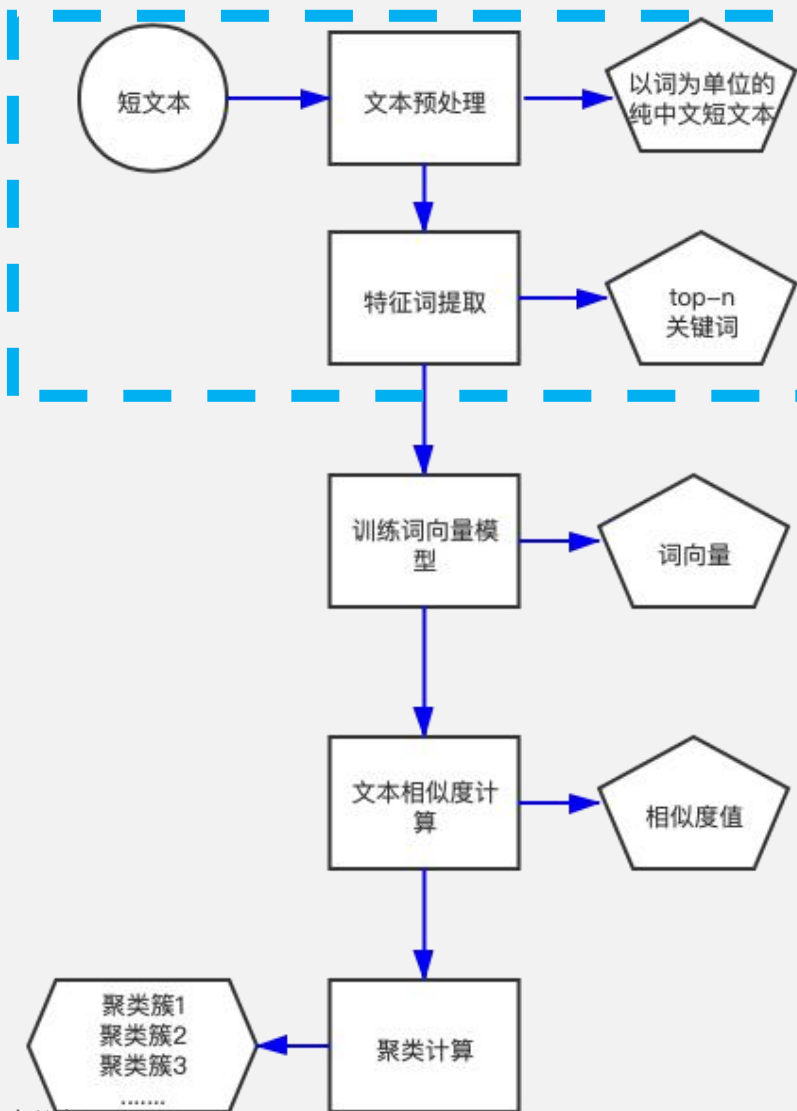
- 从品类上来看，热门餐厅前20名主要以茶餐厅、广式早茶、海鲜、甜品铺、粥粉铺等品类为主。其中，广式早茶和茶餐厅相比其他粤菜品类具有绝对优势。
- 从具体排名来看，广州百年大茶楼点都德具有“霸榜”趋势，占到广州粤菜餐厅热门前20名的8席，说明极具规模和省外知名度的全国连锁型餐厅更容易得到外来游客的追捧。

2. 美食区域的空间布局

热点广府美食旅游目的地多围绕地铁站点和旅游景点分布



3. 结合TF-IDF和词向量的文本聚类分析



TF-IDF 算法

$$w_{i,j} = tf_{i,j} \times idf_i$$

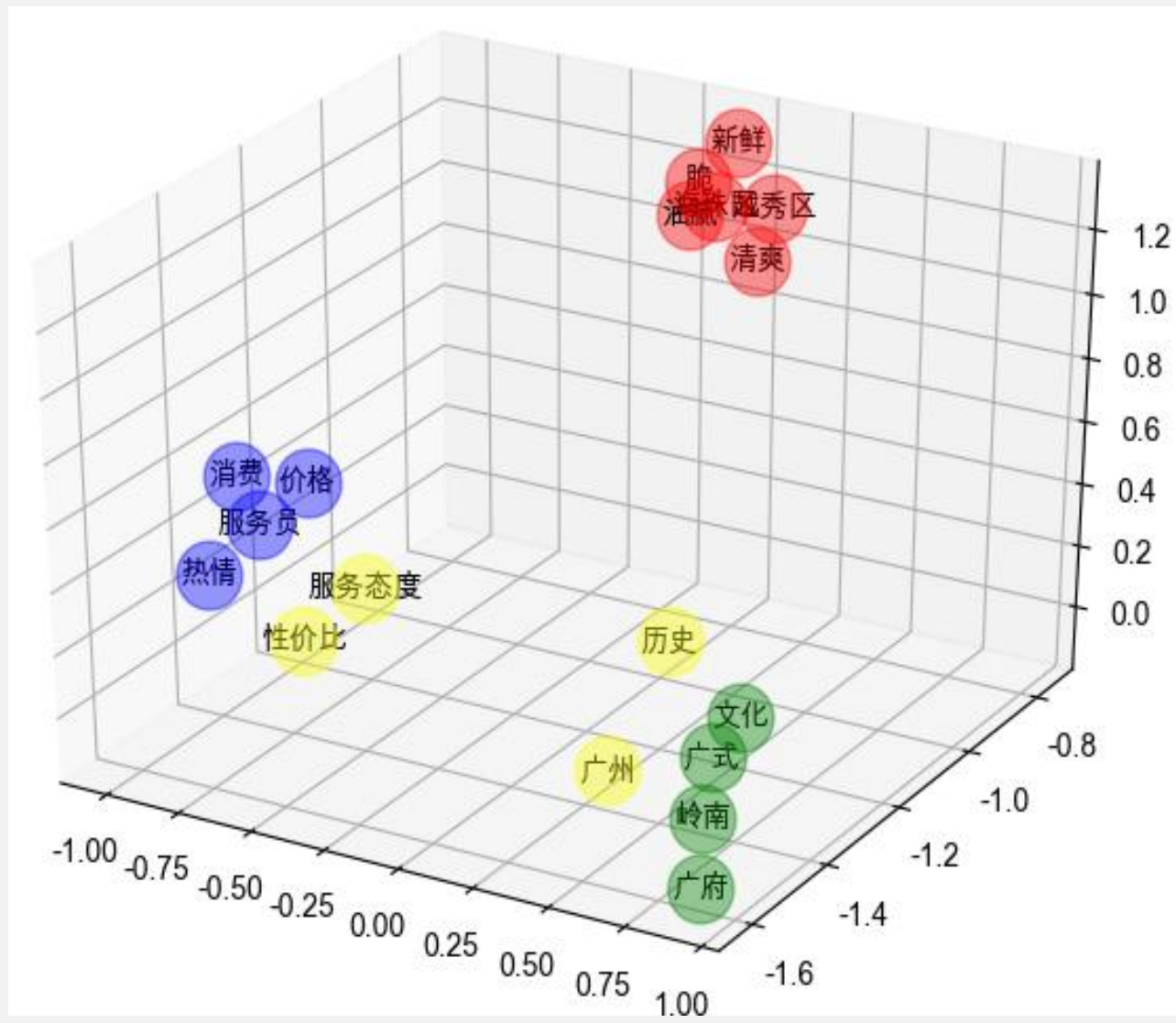
词的重要性随着词语在文件中出现的频率成正比增加，随着词语在语料库中出现的频率成反比下降

表3-7 TF-IDF结果示例

	原文本	根据TF-IDF值提取特征词结果
1	广州好热，今天吃的这家粤菜口味，档次高，价格贵，不敢再去。(共29字)	档次 高 粤菜 贵
2	店里装修别致，用餐很舒适，小朋友很喜欢，菜式不错，力荐烧鹅和巷子里的萝卜牛腩还有十三幺，还有野生菌煲也不错呢，豆腐煲也挺好的，炒芥蓝很喜欢，口感嫩嫩脆脆的，味道挺赞（共82字）	嫩嫩 挺赞 巷子 力荐 别致 脆脆 装修 小朋友 喜欢 牛腩 舒适 菜式
3	出品精致，即点即制，材料新鲜，味道好极了。烧乳鸽只能2人点一个，18块一只，这次的有点小，卖相也不如之前的.....蛋挞热辣辣，就是皮超薄，软软的拿不起.....唯一不喜的是，没有停车场叫楼下付费，12元一小时。（共275字）	茶水 红米肠 超薄 付费 蛋挞 罗汉果 停车场 弹牙 喜欢 鲜 材料 热 卖相 不腻 浓 爽口 皮 一口 楼下 不喜 汁

图3-7 文本聚类分析过程

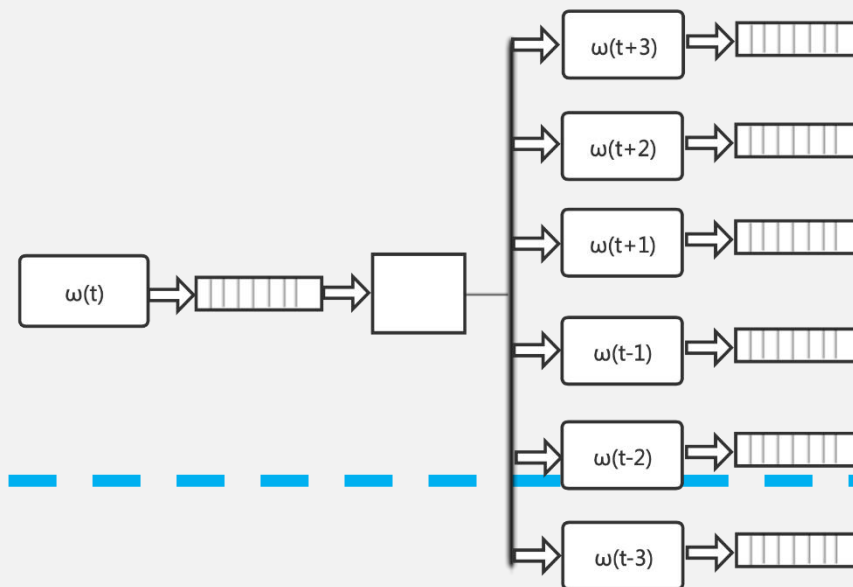
3.结合TF-IDF和词向量的文本聚类分析



输入层

投影层

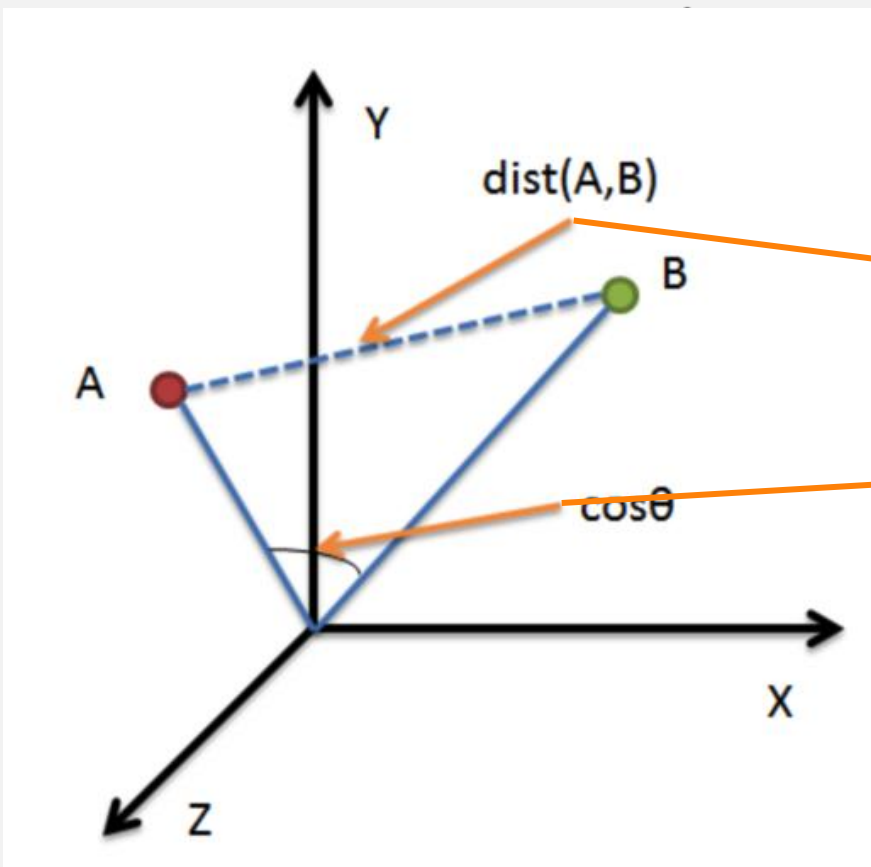
输出层



Word2vec作为词嵌入 (word embedding) 的一种, 本质就是用一层神经网络把one-hot形式的稀疏词向量映射称为一个n维的稠密向量的过程。其中, Skip-Gram 模型根据词语计算出它前后出现某几个词的各个概率。

图3-8 词向量效果示意图

3.结合TF-IDF和词向量的文本聚类分析



缺点

kmeans聚类采用的欧氏距离作为目标函数聚类的度量方式并不适合文本聚类。欧氏距离衡量的是空间各点的绝对距离；而余弦距离衡量的是空间向量的夹角

现有开源机器学习库的k-means没有提供自定义聚类的接口，能直接传如一个距离矩阵，需要用python自主实现，难度较大

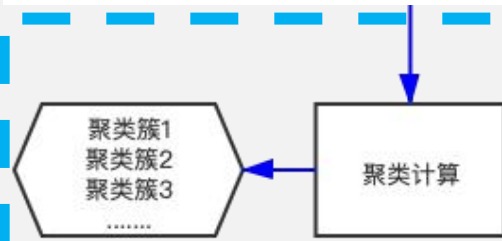


图3-7 文本聚类分析过程

经典kmeans聚类在给定K个初始类簇中心点的情况下，把每个点分到离其距离最近的簇中，然后根据簇内所有点重新计算该类簇的中心点，然后再迭代直至类簇中心点不在变化

3.结合TF-IDF和词向量的文本聚类分析

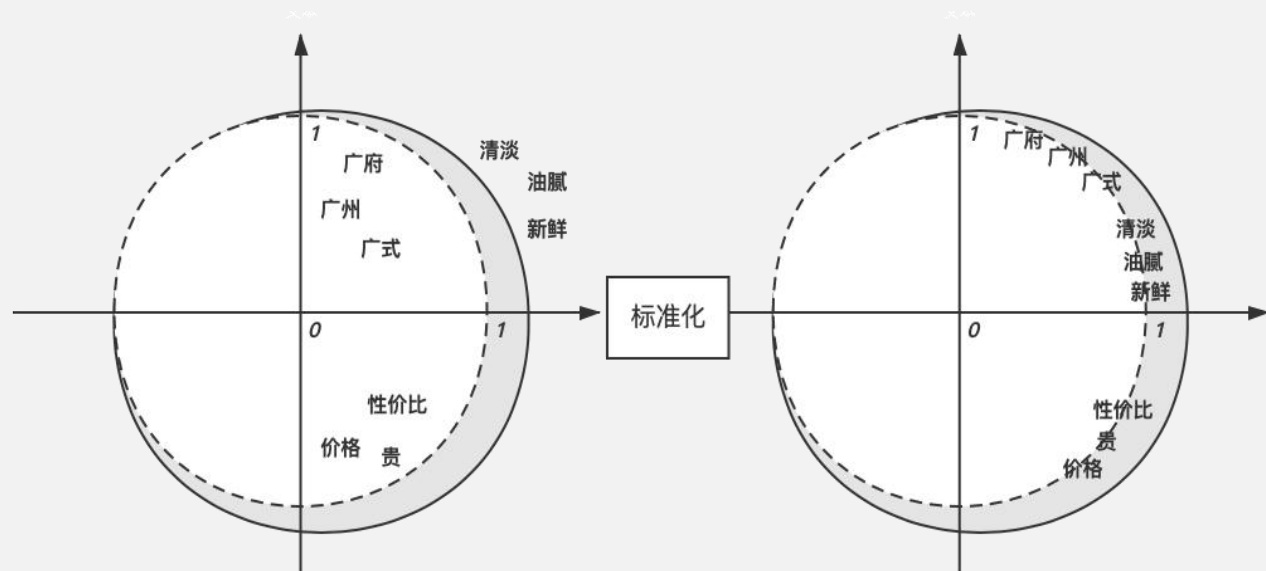


图3-7 图3-4标准化效果示意图

解决

欧氏空间中的任意两个向量，余弦相似度在标准化前后保持不变

$$\begin{aligned}\cos(\tilde{x}, \tilde{y}) &= \sum_{i=1}^n \tilde{x}_i \tilde{y}_i = \sum_{i=1}^n \frac{x_i}{\|x\|} \frac{y_i}{\|y\|} \\ &= \sum_{i=1}^n \frac{x_i y_i}{\|x\| \cdot \|y\|} \\ &= \cos(x, y)\end{aligned}$$

超球聚类算法

经典kmeans聚类在给定K个初始类簇中心点的情况下，把每个点分到离其距离最近的簇中，然后根据簇内所有点重新计算该类簇的中心点，然后再迭代直至类簇中心点不在变化

3. 结合TF-IDF和词向量的文本聚类分析

超球聚类算法：k 值的选择

Inertia 簇内平方和

$$CSS = \sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2)$$

最佳值：拐点

Silhouette Index 轮廓系数

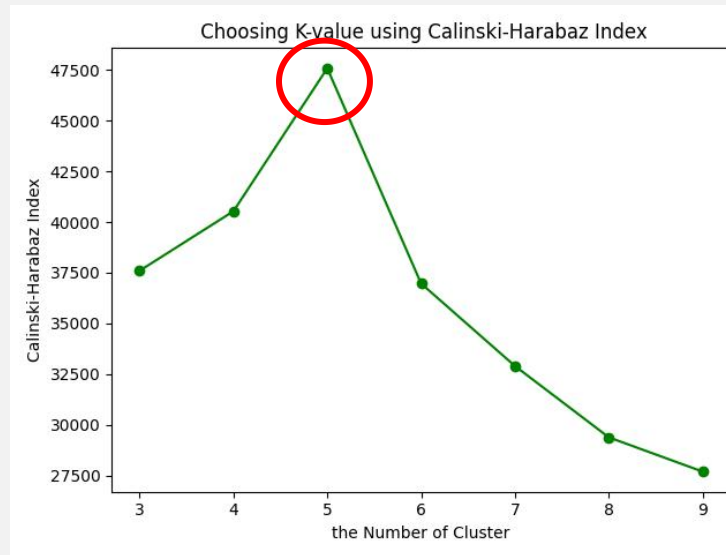
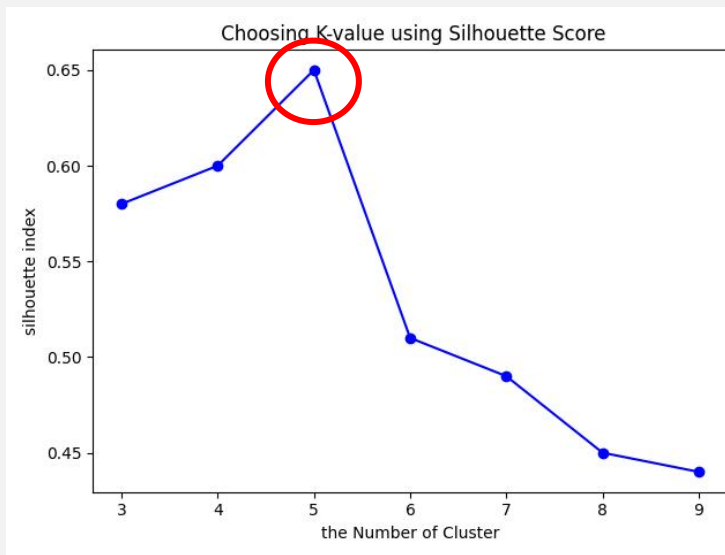
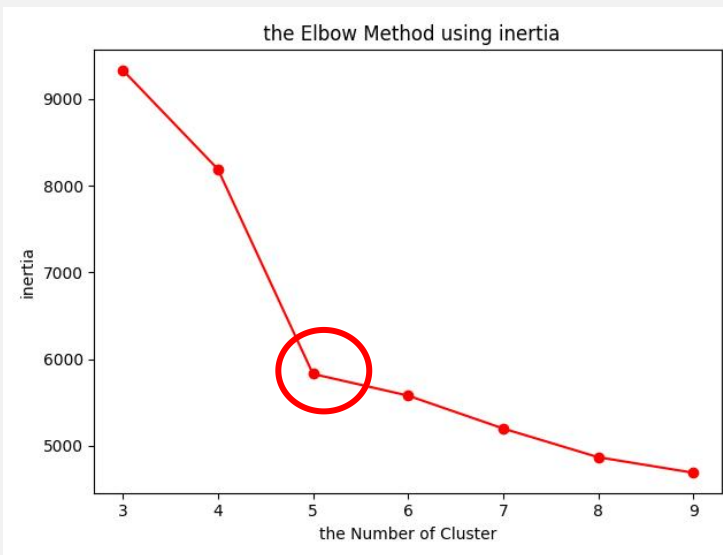
$$s = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

最佳值：接近1

Calfskin-Harabaz CH系数

$$s(k) = \frac{\text{tr}(B_k)m - k}{\text{tr}(W_k)k - 1}$$

最佳值：最大值



3.结合TF-IDF和词向量的文本聚类分析

聚类结果

2. 效用价值

味道、美味、新鲜、清淡、爽、油腻、脆、嫩...

原材料、汤、粥、肠粉、茶水...

营养、健康、正宗、分量、出品、招牌...

4. 价格价值

价格、实惠、菜单、便宜、贵、性价比、消费...

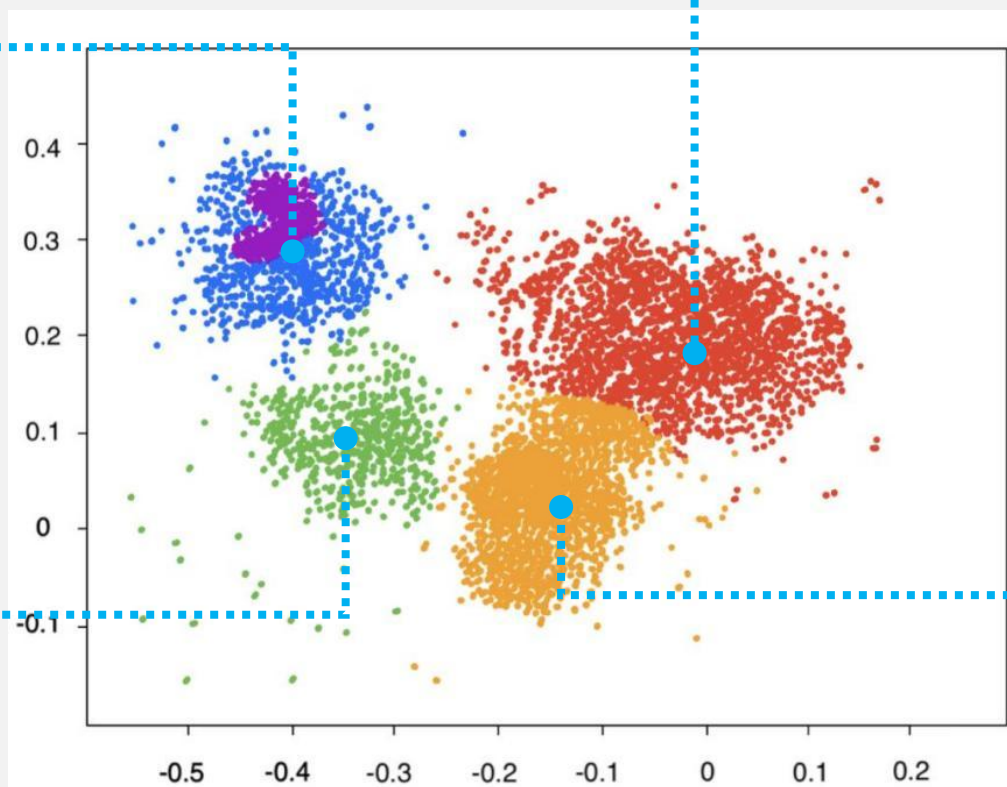


图3-8 k为5时聚类效果

1. 文化认知价值

广州、广式、文化、历史、港式、粤语...
岭南、越秀区、海珠区、西关、顺德...

地道、经典、传统、情怀、怀旧、老字号、老店、多年...

特色、出名、著名、《舌尖上的中国》...

3. 服务价值

服务态度、服务员、热情、不耐烦、细致、速度慢...

环境、餐具、装修、茶杯、氛围、卫生...
交通、位置、停车位、排队、人多...

3.结合TF-IDF和词向量的文本聚类分析

聚类结果

2. 效用价值

该簇面积较大，说明效用价值对美食感知价值贡献较大。游客们对菜品本身和原材料等的关注与粤菜重视烹饪技巧、食材、地方特色的传统相吻合

4. 价格价值

价格价值也是构成广府美食感知价值重要部分。虽然相比其他簇面积较小且内部结构较分散，但货币成本对游客价值感知对影响不容忽视

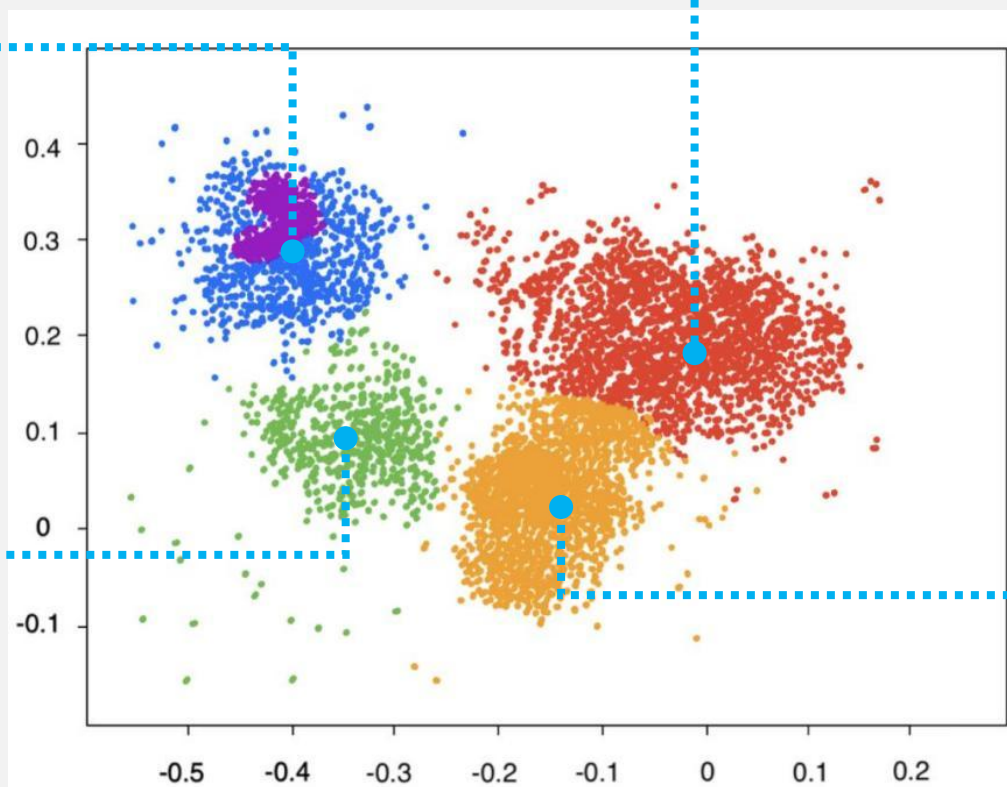


图3-8 k为5时聚类效果

1. 文化认知价值

该簇面积最大，说明文化因素在游客美食评论文本中占比较高，旅客们较为关注当地文化、菜品制作工艺等要素。在因而广府美食体验中，文化认知价值感知维度较其他感知维度更加重要

3. 服务价值

游客们不仅关注于食物本身，也同样关注服务人员的服务态度、餐厅装修、饮食餐具、位置交通等。同时簇1与簇3有重叠部分，说明文化认知价值与服务价值联系紧密、互相影响

4. 文本情感分析

研究设计

本文基于百度AI平台的自然语言处理Python SDK，在python中载入baidu-aip库，利用该库的**情感倾向分析接口sentimentClassify对评论文本进行情感倾向识别**，并返回置信度、分类概率等数据

为了进一步统计各个关键词蕴含的正面情感，即积极情绪比例，本文定义积极系数PI：

$$PI_x = \frac{\sum_{i=1}^n P_{(positive,x,i)}}{n}$$

n代表关键词出现在句子的次数，i代表关键词x存在于句子i，分子代表关键词所在句子的积极类别的概率之和。**积极系数PI实际代表该词出现在一段文本中，该文本为积极情感的概率**

表3-5 情感倾向分析输出参数

参数	是否必须	类型	说明
text	是	string	输入的文本内容
items	是	array	输入的词列表
+sentiment	是	number	表示情感极性分类结果, 0:负向, 1:中性, 2:正向
+confidence	是	number	表示分类的置信度
+positive_prob	是	number	表示属于积极类别的概率
+negative_prob	是	number	表示属于消极类别的概率

资料来源：百度智能云情感倾向分析技术帮助文档

4. 文本情感分析

研究结果

(1) 广府美食游客整体感知结果

表3-9 广府美食情感分析整体结果

游客对广府美食的整体印象为**较好**，属于积极情感的概率达到69%，属于消极情感的概率低至31%，该统计结果的置信度超过88%，数据离散程度较低。

文本数量 (条)	情感倾向分类结果		分类置信度		属于积极类别的概率		属于消极类别的概率	
	均值	标准差	均值	标准差	均值	标准差	均值	标准差
39752	1.397	0.664	0.882	0.174	0.686	0.313	0.314	0.313

*情感极性分类结果 0表示负向，1表示中性，2表示正向

4. 文本情感分析

研究成果

(2) 各感知维度及其关键词积极系数

表3-10 各感知维度及关键词积极系数

感知维度	感知结果		关键词,积极系数PI
	均值	标准差	
文化认知价值	0.726	0.243	(广州,0.688)(广式,0.593) (广府,0.732) (文化,0.825) (历史,0.702) ((地道,0.872) (经典,0.747) (传统,0.708) (老字号,0.624)(老店,0.763)(越秀区,0.638)(海珠区,0.509)(西关,0.661)(顺德,0.495)(特色,0.785)(《舌尖上的中国》,0.816)
效用价值	0.608	0.327	(味道,0.597)(好吃,0.882)(美味,0.764)(原材料,0.582)(原料,0.497)(粥,0.402)(肠粉,0.565)(营养,0.677)(正宗,0.588)(分量,0.633)(出品,0.419)
服务价值	0.385	0.227	(服务态度,0.554)(服务员,0.385)(环境,0.569)(装修,0.428)(交通方便,0.713)(位置,0.316)(卫生,0.637)(排队,0.496)
价格价值	0.419	0.394	(价格,0.576)(实惠,0.783)(太高,0.125)(性价比,0.452).

文化认知价值关键词的积极系数普遍高于平均积极系数，说明游客们对广州美食的文化认知价值认可度和满意度较高，即**广府文化、正宗程度、知名度等要素在游客感知广府美食价值的过程中起到了正向作用**

效用价值维度关键词的积极系数普遍较高，说明**游客对广州美食的功能价值认可度和满意度较高**

4. 文本情感分析

研究结果

表3-10 各感知维度及关键词积极系数

感知维度	感知结果		关键词,积极系数PI
	均值	标准差	
效用价值	0.608	0.327	(味道,0.597)(好吃,0.882)(美味,0.764)(原材料,0.582)(原料,0.497)(粥,0.402)(肠粉,0.565)(营养,0.677)(正宗,0.588)(分量,0.633)(出品,0.419)

表3-11 特色菜品与非特色菜品积极系数对比

特色菜品（部分）及积极系数PI	非特色菜品（部分）及积极系数PI
(汤,0.628)(粥,0.402)(肠粉,0.565)(烧鸡,0.625)(乳鸽,0.757)(虾饺,0.698)(叉烧,0.652)(杨枝甘露,0.592)(云吞面,0.601)(啫啫煲,0.724)...	(米饭,0.428)(酸菜鱼,0.335)(粉丝,0.402)(牛肉,0.535)(茶水,0.334)(白菜,0.405)

广州粤菜餐馆制作烧鸡、乳鸽、虾饺、叉烧等美食的积极普遍较高，说明**越是具有代表性的特色美食，越能激发游客对效用价值的正面感受**，而缺乏地方特色的菜品给游客带来的效用价值感知往往一般，对游客广府美食满意度的贡献度较小

4. 文本情感分析

研究成果

(2) 各感知维度及其关键词积极系数

表3-10 各感知维度及关键词积极系数

感知维度	感知结果		关键词,积极系数PI
	均值	标准差	
服务价值	0.385	0.227	(服务态度,0.554)(服务员,0.385)(环境,0.569)(装修,0.428)(交通方便,0.713)(位置,0.316)(卫生,0.637)(排队,0.496)

● 服务价值关键词的积极系数普遍较低,说明**游客们对广州美食的服务价值认可度和满意度较低**。利用词向量模型提取出与“服务员”语义最相近的10个词进一步分析原因

广州餐馆的整体服务水平还有待加强,针对**服务员个人素质、语言素质**较低问题需要加强培训,针对出菜速度慢的问题需要加强**服务流程规范和优化**

表3-12 “服务员” 相关词

关联词及关联度

('比较慢', 0.898) ('很差', 0.896) ('改进', 0.882) ('抽烟', 0.880) ('提高', 0.880)

('不理', 0.878) ('超好', 0.876) ('听不懂', 0.871) ('不耐烦', 0.870) ('耐心', 0.870)

4. 文本情感分析

研究结果

(2) 各感知维度及其关键词积极系数

表3-10 各感知维度及关键词积极系数

感知维度	感知结果		关键词,积极系数PI
	均值	标准差	
价格价值	0.419	0.394	(价格,0.576)(实惠,0.783)(太高,0.125)性价比,0.452).

表3-13 “性价比”相关词

关联词及关联度

(‘贵’,0.841) (‘蛮高’,0.826) (‘位置’, 0.8241)(‘高档次’, 0.821)(‘服务费’, 0.821)
(‘纸巾’, 0.820) (‘学校’, 0.817) (‘茶水费’, 0.816)(‘优惠券’, 0.815) (‘菜品’, 0.815)

价格价值关键词的积极系数普遍较低，低于平均积极系数，说明**游客们对广州美食的价格价值的满意度较低**

针对“**茶水费**”、“**服务费**”、“**纸巾**”等相关词，可以推测可能是由于除广州外许多省市的餐馆不存在茶水费和会主动提高免费餐巾纸。“优惠券”积极系数较高，说明餐厅适当推出如优惠券、优惠套餐等**让利活动能让游客对性价比的感知提升**



研究启示:

1. 引入感知价值理念，赢取市场竞争优势
2. 开发饮食文化产品，促进饮食产业升级
3. 完善食品安全体系，提升粤菜出品质量
4. 完善餐厅服务规范，构建合理价格体系

研究展望

01

由于社交网站爬虫限制和个人信息保护机制，导致本文结论缺乏人口统计特征对美食旅游感知价值中的影响显著性检验分析

02

本研究属于个案研究，得出的游客感知价值量表尚缺乏普适性。本文仅对广州市粤菜餐厅的网络评论文进行了游客感知价值维度的初测，属于探索性研究

03

文本采用的文本聚类方法仍有一定局限性。不同簇之间余弦相似度存在较大差异，因此可能导致聚成一簇的文本之间的非相似性，最终影响聚类效果。



华南理工大学

South China University of Technology

感谢



作者：李钰潇