

数据集

使用的数据来自头条与微信公众号，关于健康的新闻与谣言。只提取数据的标题与内容，整理为五个数据集，保存在 `cleaned_data` 文件夹：

- 头条新闻 (`data_articles_health.csv`) ：4453条健康类新闻
- 头条谣言 (`rumors_positive_health_cleaned.csv`) ：145条健康类谣言
- 微信新闻 (`weChat_facts.csv`) ：1083条健康类新闻
- 微信谣言(全部) (`weChat_rumours_all.csv`) ：5071条健康类谣言(去除了两条空信息)
- 微信谣言(词数大于50) (`weChat_rumours(gt50).csv`) ：3779条健康类谣言

LIWC结果

对五个数据集分别进行liwc统计，结果保存在 `liwc_res` 文件夹。

liwc结果有71个字段：

--	--	--	--	--
pronoun	ppron	i	we	you
they	ipron	TenseM	PastM	PresentM
ProgM	affect	posemo	negemo	anx
sad	death	preps	conj	cogmech
cause	discrep	tentat	certain	inhib
excl	bio	body	health	sexual
social	family	friend	humans	percept
hear	feel	achieve	adverb	assent
filler	funct	home	Interjunction	leisure
motion	MultiFun	negate	nonfl	number
quant	QuanUnit	relativ	relig	space
swear	time	verb	work	youpl

独立样本t检验结果

使用的数据来自头条与微信公众号，关于健康的新闻与谣言。对数据的liwc结果进行独立样本t检验，保存在 `ttest_res` 文件夹。

结果包括两列数值：

- `t_value`：t检验的结果，绝对值越大，相关性越大。

- p_value: p值, 大于0.05可认为两个数值具有相关性。

新闻与谣言 (同一数据集)

1. ttest_res-data_articles_health&rumors_positive_health_cleaned.csv
头条新闻与头条谣言进行独立样本t检验的结果。
2. ttest_res-weChat_facts&weChat_rumours(gt50).csv
微信新闻与微信谣言(词数大于50)进行独立样本t检验的结果。
3. ttest_res-weChat_facts&weChat_rumours_all.csv
微信新闻与微信谣言(全部)进行独立样本t检验的结果。

新闻与谣言 (跨数据集)

4. ttest_res-data_articles_health&weChat_rumours(gt50).csv
头条新闻与微信谣言(词数大于50)进行独立样本t检验的结果。
5. ttest_res-data_articles_health&weChat_rumours_all.csv
头条新闻与微信谣言(全部)进行独立样本t检验的结果。
6. ttest_res-weChat_facts&rumors_positive_health_cleaned.csv
微信新闻与头条谣言进行独立样本t检验的结果。

新闻与新闻

7. ttest_res-data_articles_health&weChat_facts.csv
头条新闻与微信新闻进行独立样本t检验的结果。

谣言与谣言

8. ttest_res-rumors_positive_health_cleaned&weChat_rumours(gt50).csv
头条谣言与微信谣言(词数大于50)进行独立样本t检验的结果。
9. ttest_res-rumors_positive_health_cleaned&weChat_rumours_all.csv
头条谣言与微信谣言(全部)进行独立样本t检验的结果。
10. ttest_res-weChat_rumours(gt50)&weChat_rumours_all.csv
微信谣言(词数大于50)与微信谣言(全部)进行独立样本t检验的结果。

LIWC字段说明

字段名	中文名	
pronoun	代名词	你、她们、在下
ppron	特定人称	他、大家、你们
i	第一人称单数	本人、自己、我
we	第一人称复数	我们、我俩、咱们
you	第二人称	你、您
shehe	第三人称单数	他、她、这位
they	第三人称复数	他们、她们
ipron	非特定人称	一切、这些、其他
TenseM	时态标定词	已经、之前、日后
PastM	过去时态	去年、刚才、昔日
PresentM	现在时态	目前、现今、平时
FutureM	未来时态	之后、即将、将来
ProgM	延续时态	了、至今、近期
affect	情感历程词	气愤、感恩、失望
posemo	正向情感词	信心、满足、祝福
negemo	负向情感词	担忧、猜疑、报复
anx	焦虑词	不安、挣扎、紧绷
anger	生气词	可恶、抱怨、破坏
sad	悲伤词	心痛、沮丧、无力
death	死亡词	亡故、自杀、遗嘱
preps	介词	从、依照、把
conj	连接词	和、一旦、不仅
cogmech	认知历程词	理解、选择、质疑
insight	洞察词	了解、恍然大悟、体会
cause	因果词	引起、使得、变成
discrep	差距词	不足、纳闷、期待
tentat	暂订词	大约、未定、差不多
certain	确切词	不容置疑、必然、保证

字段名	中文名	
inhib	限制词	废止、不准、规则
incl	包含词	包括、附近、添加
excl	排除词	取消、但是、除外
bio	生理历程词	头晕、流汗、拥抱
body	身体词	脖子、皮肤、肠胃
health	健康词	失眠、医生、运动
sexual	性词	上床、性欲、裸体
ingest	摄食词	消化、吃、煮菜
social	社会历程词	家人、接待、打招呼
family	家庭词	公婆、兄弟、孙女
friend	朋友词	同伴、朋友、麻吉
humans	人类词	人民、成员、群众
percept	感知历程词	温暖、经验、注视
see	视觉词	外貌、闪亮、绿色
hear	听觉词	大叫、听见、谈话
feel	感觉词	平滑、弹性、触摸
achieve	成就词	擅长、尽职、高手
adverb	副词	曾经、渐渐、那么
assent	应和词	了解、真的、好
auxverb	助动词	不必、可能、应该
filler	填充赘词	就、像是、话说
funct	功能词	或许、许多、那些
home	家庭词	房子、家人、宠物
Interjunction	语助词	呢、吗、吧
leisure	休闲词	唱歌、轻松、假期
money	金钱词	富有、年薪、折扣
motion	移动词	通过、靠近、参加
MultiFun	多用途词	的、有、是

字段名	中文名	
negate	否定词	不要、不必、没有
nonfl	停顿赘词	呃、然后、那
number	数字	一、八百、两万
PrepEnd	后置词	之中、以上、为止
quant	概数词	一些、所有、众多
QuanUnit	量词	条、头、支
relativ	相对词	以前、相比、达到
relig	宗教词	上帝、慈悲、信仰
space	空间词	里面、街道、台上
SpecArt	特指定词	本、该、每
swear	脏话	去你的、智障、机车
time	时间词	期间、过去、秋天
verb	动词	分享、开车、听
work	工作词	工厂、面试、薪水
youpl	第二人称复数	你们、汝等