



HITSCIR_Run: COAE2015 微博观点句识别任务分析系统

作者：李泽魁，赵妍妍，秦兵，刘挺

演讲者：唐都钰

单位：社会计算与信息检索研究中心

院校：哈尔滨工业大学

日期：2015-08-25



COAE 2015 Task2任务介绍

- 微博观点句识别任务

- 输入：微博
- 输出：微博的情感倾向(三类)

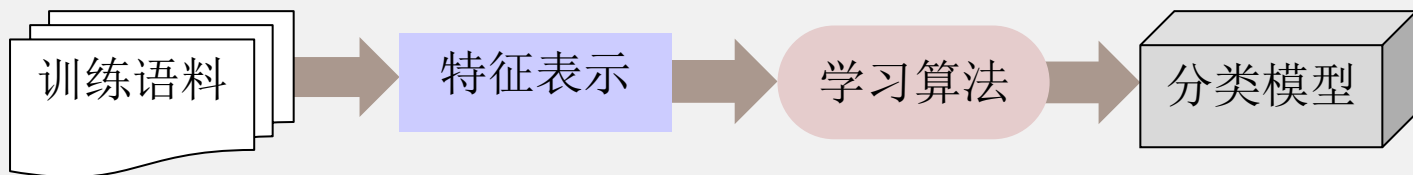
- 微博情感倾向标注样例

- 褒义：“节目足够吸引您的眼球！”
- 贬义：“米国的80%的宾馆房间都不能让人抽烟，真是扯蛋。”
- 褒贬混合：“外形虽不那么cute,但也算体面中略带个性。”

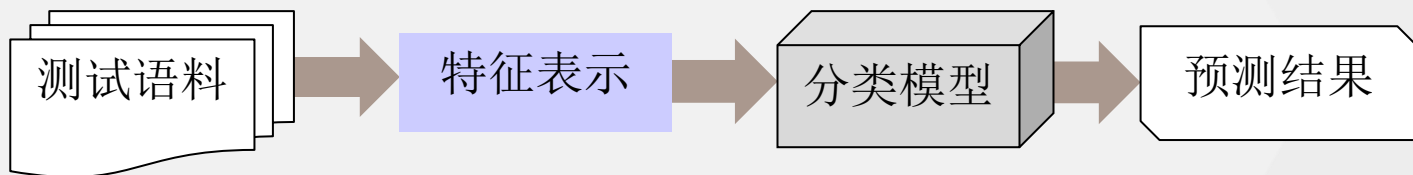
情感分类——有监督的机器学习任务

- 有监督的机器学习算法

- 训练过程



- 预测过程





基于特征的文本情感分类

- Webis (Top System in SemEval 2015)
 - 特征工程 (Feature Engineering)
 - N-Gram特征
 - 词典特征
 - 词性特征
 -
- Coooolll (2nd System in SemEval 2014)
 - 特征学习 (Feature Learning)
 - 词向量的训练学习
 - 文本中的词向量加入到特征模板

特征工程(1)

- N元语法特征 (N-gram特征)
 - 一元语法特征 (Unigram特征) 和二元语法特征 (Bigram特征)
- 词典特征
 - 6个中文情感分类褒贬词典
 - 统计不同词典在一句文本的一些数值特征, 例如褒义词数目、贬义词数目、褒贬义词数目差
- 多词典规则评分特征

微博文本示例	文本现象	评分策略
这道菜好吃。	情感词	+情感词分值
这道菜真好吃。	情感词+强程度副词	+情感词分值*2
这道菜还算好吃。	情感词+弱程度副词	+情感词分值*0.5
这道菜不怎么好吃。	情感词+短距离否定	-情感词分值
我不认为这道菜好吃。	情感词+长距离否定	-情感词分值
这道菜哪儿好吃了?!	情感词+短距离疑问	忽略情感词



特征工程(2)

- 情感专属词向量特征 (SSWE)
 - 统计句中每个词语的词向量的每个维度，追加每一维的池化 (Pooling) 信息，这里池化信息指简单的最大值、最小值和平均值信息
- 词性特征
 - 词性特征本文特指和句子中情感倾向相关的一些词性的出现频率，例如形容词、副词、标点等
- 表情符特征
 - 表情符数目、情感倾向
- 否定词特征
 - 文本进行否定标记处理 (“不是很喜欢” -> “不是很_NEG 喜欢_NEG”)
 - 否定词数目、否定区间情感词命中数目



特征工程(3)

- 感叹词特征
 - 统计感叹词数目、最后一个词是否是感叹词或者感叹型标点
- 疑问词特征
 - 统计疑问词数目、最后一个词是否是疑问词或者疑问型标点
- 话题Hashtag特征
 - 是否有Hashtag
 - Hashtag内文本的分词结果

特征工程--词典资源

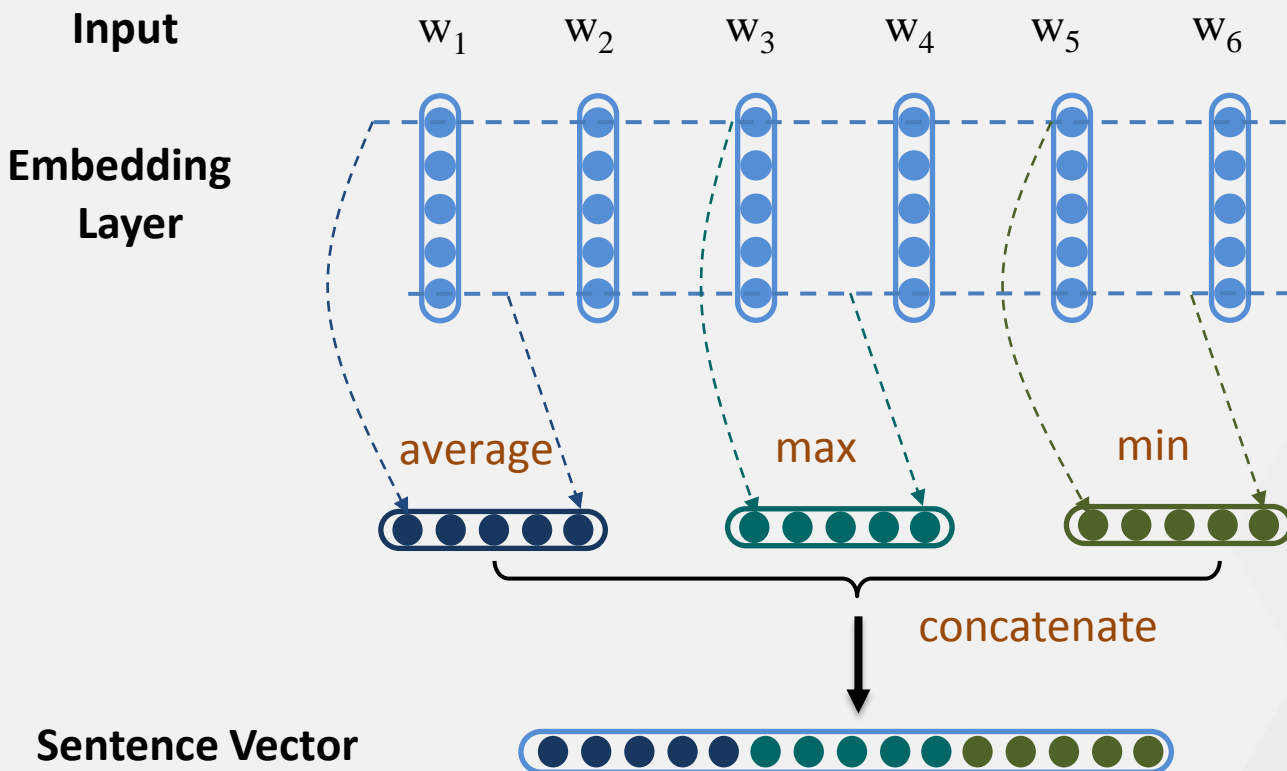
- 情感词典资源

ID	词典名称	创作版权	情感分值	词典规模
1	哈工大褒贬词典	哈尔滨工业大学 社会计算与信息检索研究中心	无	21690
2	大连理工褒贬词典	大连理工大学信息检索研究室	有	22012
3	台湾大学NTUSD	台湾大学	无	11086
4	清华大学褒贬词典	清华大学李军等人	无	10035
5	清华大学情感词典	清华大学原博等人	有	23419
6	哈工大大规模情感词词典	哈尔滨工业大学 社会计算与信息检索研究中心	无	150000

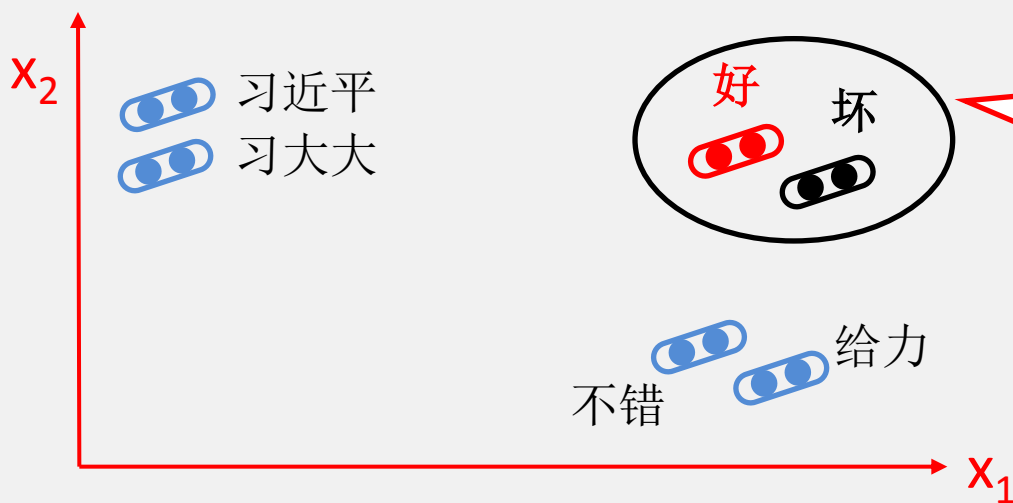
- 非情感词典资源

词典资源类型	词典名称	词典规模	词汇样例
修饰词词典	否定词	107	不、别
	强程度副词	193	真的、最
	弱程度副词	45	稍微、有点
	疑问副词	15	哪里、有啥
连词词典	总结或转折连词	40	总之、但是

特征学习



- 词向量表示

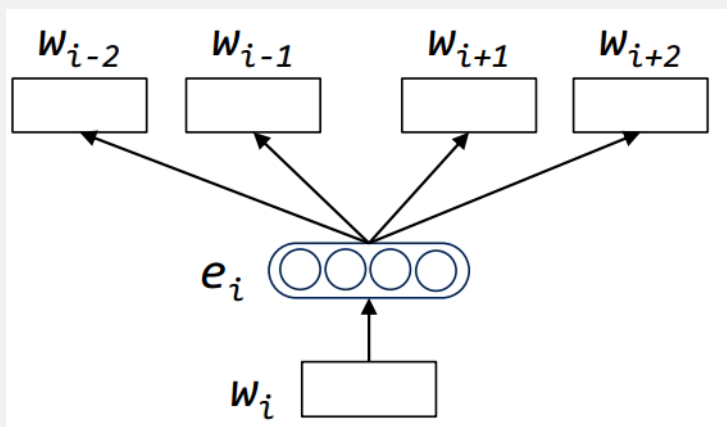


这两个词有相似的上下文表示，但是极性却相反。直接使用深度学习，会给出错误结果。

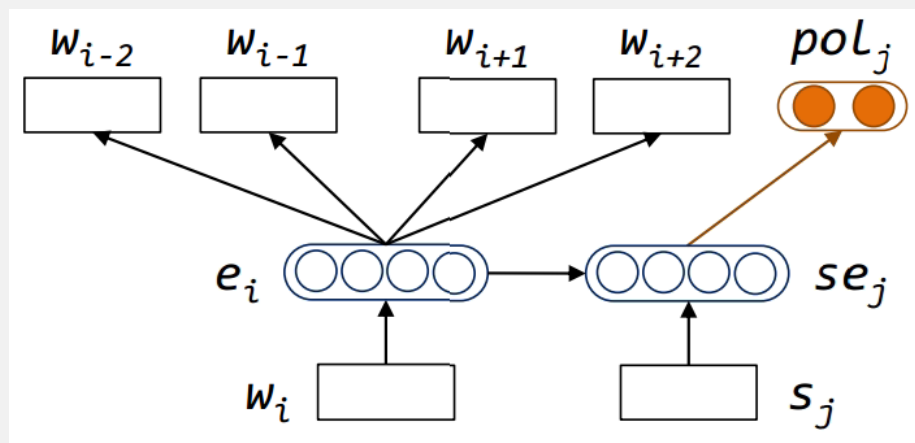
这是一个好消息
这是一个坏消息

情感专属词向量学习

- 情感专属词向量 (Tang et al. ACL2014, COLING2014)
 - (SSWE, Sentiment Specified Word Embedding)



word2vec



sswe-hard

情感专属词向量学习

- 训练输入

- 约170万句根据表情符自动弱标注的褒贬中三元情感的中文微博
- 褒/贬/中：414K/1038K/293K

情感倾向	表情符数目	举例
褒义	60	
贬义	43	

- 训练输出

- Embedding vocabulary size = 274,789
- Embedding dimension = 50



实验设计——训练语料的构建

- 评测语料分布分析（简单统计）
 - 微博比例较小
 - 商品评论（手机、汽车、相机等领域）占很大一部分
- 训练语料构成
 - 共计30,049句褒贬中训练语料
 - 褒/贬/中：12,622/11,553/5,874

ID	语料领域	极性分布	语料规模	标注单位
1	新浪微博	褒贬中	7019	哈工大社会计算与信息检索研究中心
2	“蒙牛”产品评论	褒贬	2278	COAE 2014情感分类训练语料
3	新闻中性文本	中性	1000	哈工大社会计算与信息检索研究中心
4	相机领域产品评论	褒贬中	19752	哈工大社会计算与信息检索研究中心

实验设计——分类模型的调优与集成

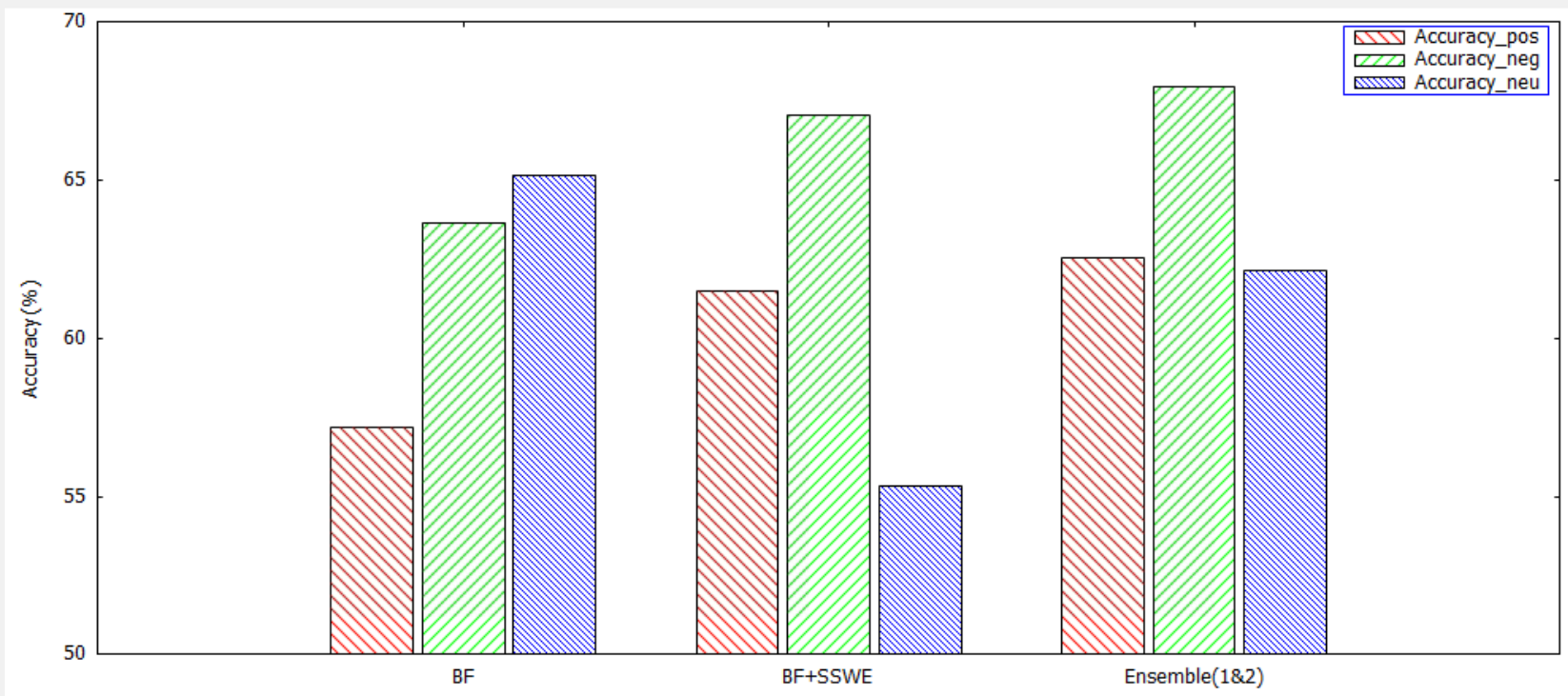
- 模型设计

ID	模型代号	特征抽取模板
1	Basic Features(BF)	除去SSWE特征外的其他基础特征
2	BF +SSWE	Basic Features +SSWE特征
3	Ensemble(1&2)	模型1和模型2的集成模型

- 实验结果(交叉验证)

ID	模型代号	褒义精确率 (%)	贬义精确率 (%)	中性精确率 (%)
1	Basic Features(BF)	57.20	63.65	65.14
2	BF +SSWE	61.47	67.04	55.35
3	Ensemble(1&2)	62.56	67.93	62.14

实验设计——分类模型的调优与集成



- BF 模型与 BF+SSWE 模型集成后
 - 褒贬分类精确率提升
 - 中性分类精确率下降



实验设计——分类模型的调优与集成

- BF 模型与 BF+SSWE 模型集成后
 - 褒贬分类精确率提升
 - 中性分类精确率下降
- 原因分析
 - Embedding的中性训练数据质量低于褒贬训练数据，影响中性分类效果
 - 如何获得高质量大规模的中性数据是一个有待解决的难题

评测结果及分析

- 评测结果(系统排名5/27, 机构排名3/15)
 - HITSCIR_Run2: Ensemble (1 & 2) Model
 - HITSCIR_Run1: BF Model

System	pos_F1	neg_F1	mix_F1	mirco_F1	marco_F1
HITSCIR_Run2	0.7562	0.7114	0.2236	0.6411	0.6106
HITSCIR_Run1	0.6596	0.6840	0.1993	0.5573	0.5821
Best	0.8617	0.7868	0.4233	0.8113	0.6945
Medium	0.7233	0.5800	0.1617	0.6057	0.4947

- 结果分析
 - 褒义贬义混合判别准确率不高
 - 训练语料偏执为中性, 而非褒义贬义混合
 - 训练数据中没有mix



结论

- 本文提出了一个基于特征的有指导情感倾向性分析系统，在参评15家机构中排名第3位
- 使用了特征工程和特征学习技术，实验结果表明了SSWE可以提升褒贬分类的精确率



哈爾濱工業大學
社会计算与信息检索研究中心



Thanks Q&A