

情感分析

Sentiment Analysis

徐冰 副教授
哈尔滨工业大学



课程介绍

- 课程名称：情感分析与计算
- 课程学时：**32学时，讲授+讨论**
- 上课时间：第**1-8周**（周一 **9-10节**；周三 **9-10节**）
- 先修课程：机器学习、自然语言处理
- 参考教材：
 - **1、情感分析-挖掘观点、情感和情绪，刘兵著，机械工业出版社，2017**
 - **2、Sentiment Analysis and Opinion Mining, Bing Liu, 2012**
- **考核方式：研讨成绩20% + 课程报告40% + 系统实现40%**



群名称：2019春情感分析课程群
群 号：764376340

内容安排

- 情感分析的基本概念
- 情感分类问题：文档级，句子级，属性级
- 情感要素抽取：属性，实体
- 情感词典
- 比较型观点分析
- 观点摘要和检索
- 意图挖掘

人工智能的重要发展方向

- 按照人工智能最为通行的定义——“人工智能是对计算机系统如何能够履行那些只有依靠人类智慧才能完成的任务的理论研究。”
 - 这里的人类智慧当然不是单指智商或者智力，还有人类情感，情感是人类神经系统对外界价值关系产生的主观反映。
- 人工智能在我们眼中一直不是普通的智力工具，所以我们一方面需要它具备超越人类的感知计算能力，另一方面也需要它能够与人进行情感交流，在重视人工智能完成任务和功能强化的同时更要建立和满足人的情感和心理需求，这才是人工智能的最终定义。

- 从实际应用来看，也出现了越来越多具备情感计算能力、能与人类进行感性交互的技术和产品。
 - **MIT教授、TED讲者雪莉·特克**在《一起孤独》一书中提出，**社交本能使人类很容易将社交关系投射到各种各样的人和事物上，当一件物品在我们的关怀下茁壮成长时，我们会觉得它是智能的。**但更重要的是，我们会觉得自己与它形成了某种关系。我所说的这种关系并不是来源于计算机真实的情感或智能，因为它们根本没有。这种关系，来自于我们自身被它们所唤起的部分。
 - 能激发孩子情感依恋的玩偶，比如菲比小精灵和真娃娃机器人玩偶。
 - 一个叫帕罗的机器海豹，用来充当老年人的伴侣动物。

- 微软认为，在人工智能领域内，新一代人工智能系统的首要任务就是需要具备“感性”的情感连接能力，这样才能以更像真实人类的方式满足人们普遍心理和情感需求，从而逐步**建立信任和依赖感**。
 - 如微软小冰，小冰通过动态决策具有了情感记忆功能，当前一天某个用户提到了自己一些心情，小冰会注意到这一点，后面的很长时间还会反复了解用户的情况，询问用户的身体有没有好一点。
 - 因此，比起直接解决问题的感知计算型人工智能，情感计算是先与用户建立一种**信任关系**，然后在此基础上形成一种**情感交流和需求满足**的良性循环。

• 如何进行情感计算？

- 人类情感的东西，实际上**隐性的包含在数据里面**，小冰之所以能够对懂人类情感，是因为今天可以借助于运算能力和算法去收集和分析用户的数据，然后产生这样的情感分析能力，小冰用这样的能力反过来再去跟用户进行交流，而其他具有“情感”的机器人也是基于这种原因。
- 我们现在可能无法弄懂人类情感的本质，但我们可以找到对应**各种情感的表现信号**，比如**寂寞对应的文字**，**开心对应的表情**，**愤怒对应的语调**，**兴奋对应的脑电波**，等等，机器通过对这些信号的分析就能找出所对应的情感，从而做到了读懂人类的情感和表达自己的情感，这样机器就具有了与人类进行情感交互的能力。

• 大数据在情感计算中的重要性

- 互联网、社交媒体、移动设备和廉价的传感器，这个世界产生的数据量急剧增加。
- 随着对这些数据的价值的不断认识，用来管理和分析数据的新技术也得到了发展。大数据是人工智能发展的助推剂，这是因为有些人工智能技术使用统计模型来进行数据的概率推算，比如图像、文本或者语音，通过把这些模型暴露在数据的海洋中，使它们得到不断优化，或者称之为“训练”——现在这样的条件随处可得。
- 深度学习来说，它善于在高维度的数据中摸索出错综复杂的结构，因此能应用在许多不同的领域。除了图像识别和语音识别，它还在许多方面击败了其他机器学习技术，在自然语言理解方面，特别是话题分类、情感分析、问答系统和语言翻译等不同的任务上。

• 情感计算的未来

- 像小冰这类产品的进化过程是人工智能中情感计算的缩影，对于人工智能的诉求不应仅限于比特的流动，还应该在于一种情感的表达，因为工具可以被取代，效率可以逐步被提高，数据可以失去价值，但**机器和用户之间因为情感互动和需求满足所形成的那种循环**却将一直持续。



第1章 引言

• 1、情感分析：Sentiment analysis, Opinion mining

- 又称：观点挖掘、观点分析、观点信息抽取、情感挖掘、主观性分析、倾向性分析、情绪分析、评论挖掘，**统一归为情感分析任务**。
- 情感分析术语最早出现在**2003年**（**Nasukawa and Yi**），观点挖掘也出现在同一年。
- 其他研究开展的更早
 - **Wiebe 1990,1994,1999; Hearst 1992 ; Picard 1997; Pang et al. 2002; Turney, 2002**
 - 隐喻分析、情感形容词抽取、倾向性计算、主观性分析等
 - **Elkan 2001:**
 - 这个专利把情感、得体性、幽默以及许多其他概念看做文本的类别标签。
- 现有情感分析的研究和应用多是**针对文本数据**，因此这一研究逐步成为**NLP领域的重要研究方向**。

• 2、情感和观点的区别：

- 研究者容易混淆，这不是语言学中的概念，而是从计算机科学派生的
- 在**Webster**词典中：
 - 情感是一种态度、想法，或者是感性的判断；
 - 观点是一种论断、判断，或称之为一种在人脑中形成的对于某一事物的评价
- 我们发现：观点更多指的是一个人对于某一事物形成的具体看法；情感更多的指一个人内在的某种情感。
 - *(1) I am concerned about the current state of the economy.*
 - *(2) I think the economy is not doing well.*
 - (1)表达了对经济现状的一种情感，担心；
 - (2)表达了一种对经济的观点和态度
- 在很多情况下，观点信息都包含着正面的或负面的观点。
- 也有例外：*I think he will go to Canada next year.*

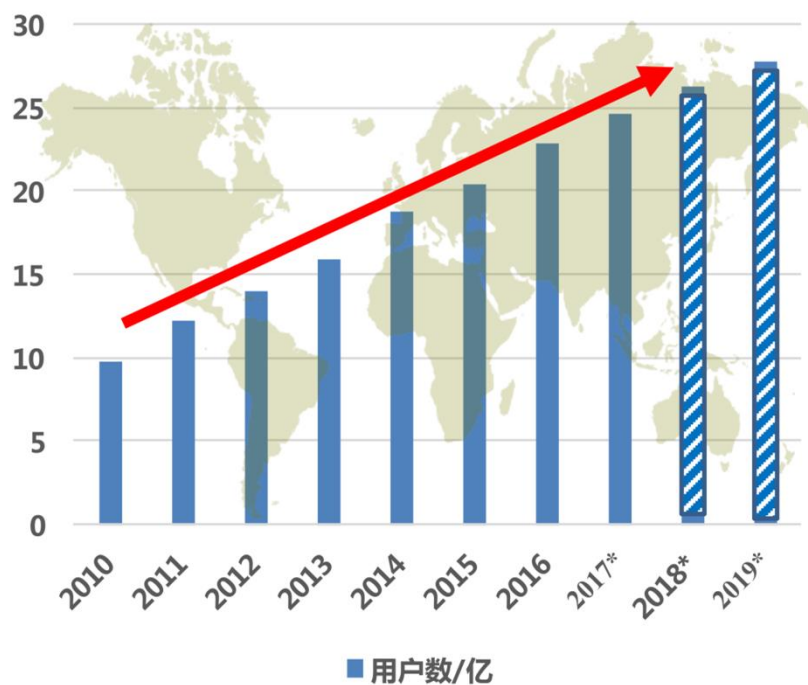
• 3、情感分析或观点挖掘的目标:

- 是从文本中识别出那些**显示或隐式**表达的褒、贬观点情感，同时包括这些观点或情感所描述的对象。
 - 观点类型的划分与社会心理学中的态度类型划分类似：
 - **Eagly and Chaiken(1998)**: 把人们对某种特定事物喜欢或不喜欢时所产生的心理倾向性称为态度。还有一些中性情感，不包含任何倾向性。
 - 还有**感情 (affect)**，**情绪 (emotion)**，**心情 (mood)**，也是意识的心理状态。
- 观点或情感本身传递了作者主观性的信息，这类表达情感或观点的句子称为主观性句子。
 - 有些客观性的句子里也隐含情感倾向，也需要注意。
 - *(1) I bought the car yesterday and it broke today.*
 - *(2) After sleeping on the mattress for a month, a valley has formed in the middle.*


• 4、社交网络分析和情感分析

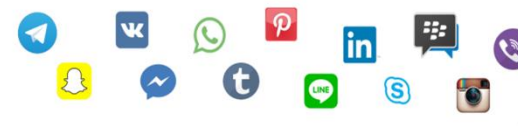
- 近**15**年，随着互联网以及社会媒体的迅猛发展，情感分析技术也迅速发展。
- 从评论、博客、**YouTube**、**Facebook**、**Twitter**、微博这些平台上，有大量的对某一事物的观点、评论和看法，研究者从社会媒体上获得大量观点信息作为分析数据
- 个人或者机构在进行决策时，都会寻求他人的观点意见，情感分析具有十分重要的意义。
- 衍生的许多研究：情感画像，用于产品推荐、服务推荐等；虚假评论检测，用于提升某种产品、服务或传播价值观的目的。

全球社交媒体用户数



数据来源：The Statistics Portal (2017)

社交媒体	月活跃用户数
 Facebook	20.47亿
 微信	9.38亿
 QQ	8.61亿
 QQ空间	6.38亿
 Twitter	3.28亿
 新浪微博	3.13亿
 百度贴吧	3亿



关于《流浪地球》的评论

中国联通 4G 13:48 50%
 商品 详情 评论 相关

全部(10297) 最新(10297) 有图(61) 长评(5)
 好评(10287) 中评(5) 差评(5) 装帧精美(86)

徐柳依 [已购]
 ★★★★★ 10分

很早就知道了这本书，马上过年放假了，要好好静下心来读一读~买书真的也会上瘾的，好的书就是看了这一本，会告诉你下一本更加好看的书~自从用了当当，根本停不下来，并且乐此不疲，去不了的地方感受不到的爱，文字可以做到，身临其境.....



2019-01-27

18 评论

solanen [已购]
 ★★★★★ 10分

大刘的故事真没法连着读.....小时候喜欢估计是因为都是断续地在《科幻世界》上看的。基本主角都是特别酷炫的科学家或老派军人，伟光正程度超越党员宇宙尺度的想象不错。社会细节就差很多。小时候这么

店铺 收藏 购物车

立即购买

加入购物车

中国联通 4G 13:48 49%
 评论详情

今年春节档的第一大片非《流浪地球》莫属，这部大年初一刚刚上映的影片，不到24小时就在豆瓣上创造了8.4分、五万多人评价的佳绩，其中近一半的观影者给出了5星好评；在专业影评人那里，这部片也几乎是零差评。

流浪地球 (2019)



豆瓣评分 8.4
 5星 45.2%
 4星 38.8%
 3星 14.2%
 2星 1.8%
 1星 1.8%
 好于 94% 科幻片
 好于 98% 灾难片

早在电影还没有正式上映前，各个点映场就纷纷传来“这是一部开启中国科幻电影元年”的赞誉声音。果真如此吗？从目前口碑、排片和票房来看，这部电影，称得上是中国科幻电影的开山之作。

其实从2009年开始，“中国硬科幻电影元年”的概念就一直频繁被提起。可十年过去了，中国的科幻电影依旧乏善可陈。呼声最高的《三体》电影迟迟未能上映，其他的电影不是特效五毛，就是逻辑硬伤。

发表回复

28

1

中国联通 4G 13:49 49%
 返回 电影流浪地球 +关注

转发 1.9万 评论 4506 赞 1.7万

按热度



赵今麦angel

#电影流浪地球#不怕严寒，不惧危难。大年初一，为了我们的家，一起作战吧！

温莎伊顿Duke等人 共51条回复 >

1-21 19:37

1905



张亦驰isolation

#电影流浪地球#用智慧冲破严寒，点燃希望！大年初一，冒险开启！

Lydia是只团子等人 共9条回复 >

1-21 20:30

755



FrantGwo郭帆

再转一个小破球的！#我和我的倔强# #电影流浪地球#

FrantGwo郭帆回复@现在徐建原来是徐健:就像“打死都不改稿之第三稿”一样.....

共32条回复 >

1-21 20:02

919



元气少女惠Rachel

真的好看哇 迪迦奥特曼 真的值得去电影院

转发

评论

赞

1.1 情感分析应用

- 舆情监控：挖掘重要的时间，通过分析得到大众的观点和看法
- 电影票房预测：用情感数据预测更有效
- 政治选举、消费信用调查、
- 股票市场预测：针对股票信息版中的帖子进行分析
- 产品排序、情绪流动与传播、情感维度、小说中的情绪进行分析和追踪、在线书评的社会影响、社会关系描述.....





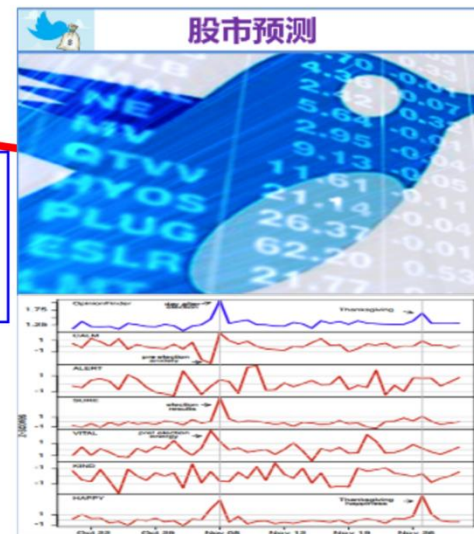
**社会
舆情**

**电子
商务**



**传统
行业**

**金融
领域**



1.2 情感分析研究

1、针对不同文本颗粒度的情感分析研究：篇章级、句子级和属性级

- **篇章级**：判别整篇文档表达了褒义还是贬义倾向（Pang et al.2002）
 - 这个任务有个基本的假设：认为一篇文档只对一个实体进行评价。
 - 但实际情况是，很多文档会同时对多个实体进行评价或比较。
- **句子级**：判别一个句子表达了褒义、贬义或中性的情感，中性情感不表达任何观点。
 - 主客观分类任务相关（Wiebe et al. 1999），但主观句不等于该句就是情感句或观点句。
 - *We bought the car last month and the windshield wiper has fallen off.*(客观句，隐含观点)
 - *I think he went home after lunch.*(主观句，不包含观点)

- 句子级情感分析也包含了假设：认为一个句子中含有褒义或贬义的观点，则认为整个句子中提到的事物都表达了相同的观点。
- 这显然不合理，所以要进行属性级分析。
- **属性级**：要素级，直接关注的是观点以及观点的目标（观点评价对象），观点评价对象是一个实体或者是该实体的属性。
 - (1) *Apple is doing very well in this poor economy.*
 - (2) *Although the service is not great, I still love this restaurant.*
 - 同时包含了褒义和贬义的观点
- 另外，观点还可分为：常规型观点和比较型观点
 - *Coke tastes very good.*
 - *Coke tastes better than Pepsi.*
 - 比较型观点比较多个实体的共同属性。

2、情感词典及其问题

- 情感词（观点词）：褒义、贬义和中性
 - 褒义情感词： **good, wonderful, amazing.....**
 - 贬义情感词： **bad, poor, terrible.....**
 - 包含情感的短语或成语： **cost an arm and a leg**
- 仅依赖情感词典不能精准的完成情感分析任务，其原因如下：
 - （1）情感词的倾向性会随着其应用领域以及所在**上下文的变化而变化**。
 - ***This camera sucks. (贬义)***
 - ***This vacuum cleaner really sucks. (褒义)***

(2) 一个句子中即使出现了情感词，也不一定会表达任何情感。

- 疑问句: *Can you tell me which Sony camera is good?*
- 条件句: *If I can find a good camera in the shop, I will buy it.*

有的也表达情感:

- *Does anyone know how to repair this terrible printer?*
- *If you are looking for a good car, get a Ford Focus.*

- (3) 讽刺句，这是一种很难处理的句式，无论其中是否出现情感词。
 - *What a great car! It stopped working in two days.*
 - 作者表达的意思是贬义
 - (4) 许多情况下，一个句子中可能不会出现任何情感词，这种句子在陈述事实，但是也隐含了作者的观点。
 - *This washer uses a lot of water.*
 - 隐晦地表达了贬义的情感。
- ### 3、辩论与评论分析
- 单一型帖子：评论或博客
 - 在线对话：辩论和讨论（交互型的情感或观点）
 - 自动发现在辩论过程中的各方立场，对参与者进行归类，挖掘支持和反对的观点，发现争论的焦点及核心问题
 - 这类研究对于社会学家非常有用

4、意图挖掘

- 意图：一个人或一群人试图遵循的行动步骤
- 应用：产品推荐、政治选举支持者发现
- 意图与情感的区别：
 - (1) 在一个含有意图倾向的句子中，作者通常会表达对于某一事物或实体的情感或情绪。
 - *I am dying to see Life of Pi.*
 - (2) 当一个人非常想得到某一东西的时候，他通常会对这个东西表达褒义的情感。
 - *I want to buy an iPhone8.*
 - (3) 有一些观点时通过描述意图的方式表达出来的。
 - *I want to throw this camera out of the window.*

• 5、垃圾观点检测与评论质量

- 垃圾观点检测:

- 垃圾观点发布 (**Opinion Spamming**): 存有不良目的或隐藏企图的人发布虚假评论欺骗情感分析系统
- 垃圾观点检测需要对用户行为进行分析, 所以不仅仅是**NLP**任务, 同时也是**数据挖掘**的任务。

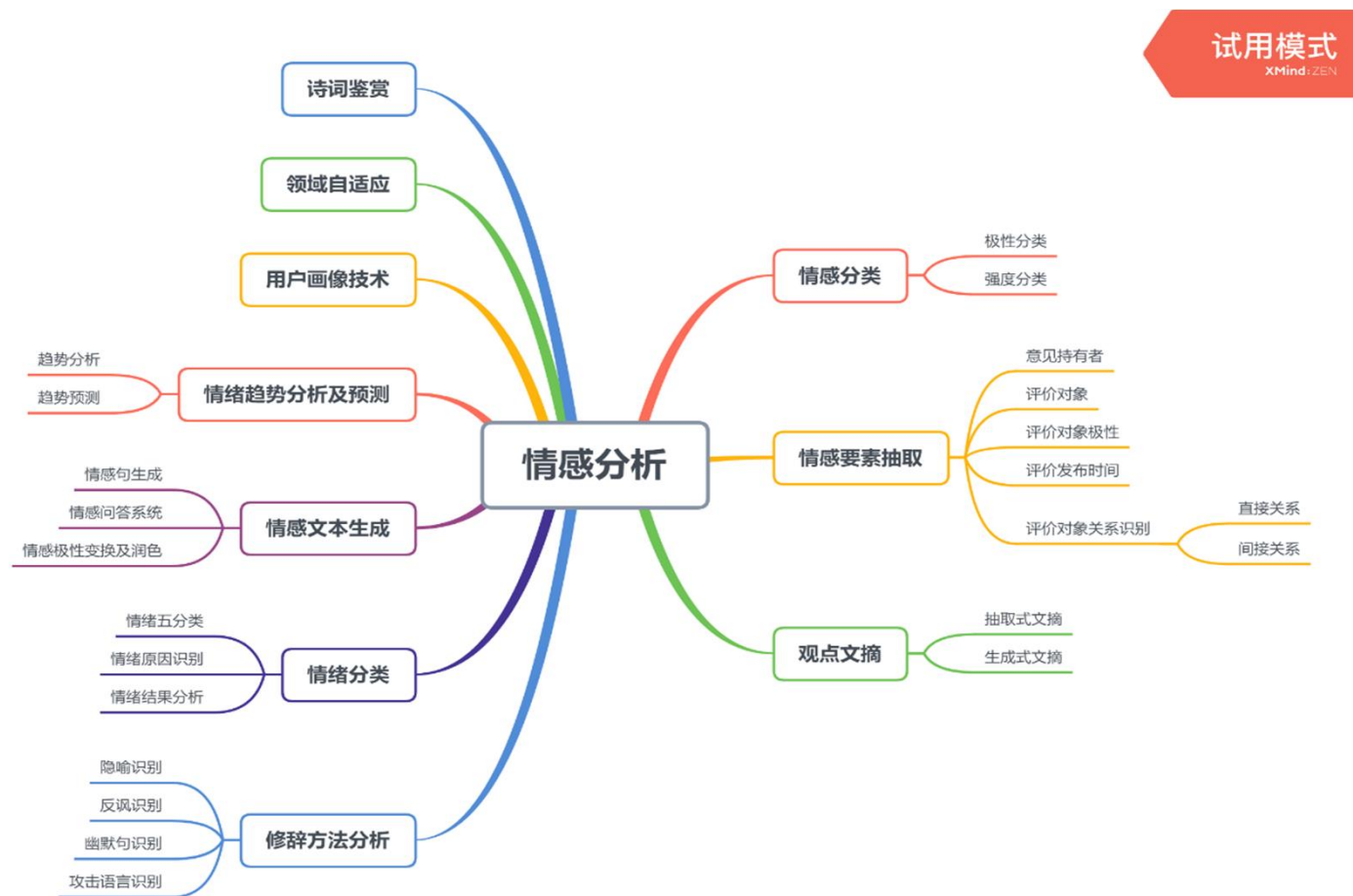
- 网络评论质量与效果评估:

- 任务目标是: 把高质量的用户评论识别出来, 并把他们展现在网页比较靠前的位置。用户首先读到这样的信息, 获得最有用的评论。

本章小结

- 情感分析是个迷你的自然语言处理任务
 - 情感分析不仅仅是**NLP**的子领域，它对于**NLP**的研究有很多扩展。
 - 情感分析涉及了词汇语义、指代消解、词义消歧、篇章分析、信息抽取以及语义分析等。
- 情感分析属于语义分析问题
 - 情感分析更限定范围，即不需要理解句子或篇章的全部语义信息，只需要**理解与情感观点有关的语义内容**
 - 在情感分析任务上，不能简单地直接运用通用自然语言处理方法，要加入**更深层次的语言理解和分析**。

在情感分析领域已有的研究



第二章 什么是情感分析

- 2.1 观点
 - 情感对象
 - 观点中的情感
 - 观点定义
 - 观点的理由和限定条件
 - 情感分析的目标和任务
- 2.2 观点摘要
- 2.3 感情、情绪和心情
- 2.4 观点的不同类型
- 2.5 作者视角和读者视角

- 情感和观点的重要特征就是 **主观性信息**
- 主要体现在这**3**个方面：
 - (1) 不同人有不同的经历，因此也就产生具有不同情感倾向的观点。
 - 同一个产品，不同的人会有不同的体验，会产生不同的评价结果
 - (2) 不同的人对相同的事物也会有不同看法
 - 股价下跌的看法
 - (3) 不同人也会有不同的兴趣和不同的意识形态
 - 研究针对多人的观点看法，比分析一个人的一个观点更重要
- **目前在产品和服务领域，观点比较容易处理。**
- 在社会和政治的评论方面难度较大，语言现象丰富，比如讽刺、反讽等，需要更多的背景知识才能更好的解决。

1、观点

2.1 观点定义

1、观点：包含情感、评价、态度等，使用情感这个词表示观点中暗含的褒义或贬义的感受。这里给出单一观点的定义

- 观点定义的四元组 (g, s, h, t)
 - g 代表观点评价对象（目标）
 - s 代表观点针对目标 g 所蕴含的情感，褒义或贬义
 - h 是观点持有者（持有此观点的人或组织）
 - t 是表达此观点的时间

举例

Posted by: John Smith

Date: September 10, 2011

“(1) *I bought a Canon G12 camera six months ago.* (2) *I simply love it.* (3) *The picture quality is amazing.* (4) *The battery life is also long.* (5) *However, my wife thinks it is too heavy for her.*”

- 每一个观点都有所评价的对象，在情感分析中找到评价对象很重要。
 - （1）一个句子如果有多个对象（通常是名词或名词短语），要去区分出每个对象所表达的不同观点分别是正面还是负面的；
 - （2）**Good, amazing, bad, poor**等这些表达情感的词语，与观点对象之间会有一些特定的句法关系，通过句法关系，可以设计出一些情感表达和观点评价对象的协同抽取算法

2、情感对象

- **情感对象的定义**：又称为**观点评价对象（目标）**，是观点所评价的实体、实体的一部分或实体的一个属性
- **实体的定义**：一个实体 e 可以是一个产品、服务、主题、个人、组织、论题或者事件。它可以用一个对 $e:(T, W)$ 来描述，其中 T 是一个层次关系，包含**组件、子组件**等。而 W 是 e 的一个**属性集合**。每个组件或子组件也有它自己的属性。
- 例：**Review A**
 - (1) *I bought a Canon G12 camera six months ago.*
 - (2) *I simply love it.*
 - (3) *The picture quality is amazing.*
 - (4) *The battery life is also long.*
 - (5) *However, my wife thinks it is too heavy for her.*

问题：请找出这个例子中的实体和属性

- 例: *Review A*

- (1) *I bought a **Canon G12** camera six months ago.*
- (2) *I simply love it. 👍*
- (3) *The **picture quality** is amazing. 👍*
- (4) *The **battery life** is also long. 👍*
- (5) *However, my wife thinks it is too heavy for her. 👎*

- 相机还有很多属性：相机尺寸、重量、曝光速度，还有一系列部件，如透镜、取景器和电池等

3 观点中的情感

- 定义：情感是观点中所蕴含的感受、态度、评价和情绪。

- 通常情感由一个三元组表示 (y, o, i) ; y 情感类型; o 情感倾向; i 情感强度

- 情感类型：

- (1) 理性情感：来源于理性推理、切实的信念和实用主义的态度，不包含任何主观的情绪。

- *eg. The voice of this phone is clear.*

- *eg. This car is worth the price.*

- (2) 感性情感：存在于人们深度的心理状态之中，来自于对实体的不可触及且情绪化的反应。

- eg. I am so angry with their service people.*

- eg. This is the best car ever.*

- 情感倾向：褒义（**positive**）、贬义（**negative**）、中性（**neutral**）
- 情感强度：情感副词加强或削弱情感的强度，如**more**，**less**
- 情感评分：常见的有离散化的评分
 - 评论5档（**1~5星**）
 - 感性正向（**5星**）
 - 理性正向（**4星**）
 - 中性（**3星**）
 - 理性负向（**2星**）
 - 感性负向（**1星**）



4 简化的观点定义

- 观点的又一定义：

定义一个五元组 (e, a, s, h, t)

- e 观点评价的目标实体
- a 实体 e 中一个观点评价的实体属性
- s 对实体 e 的 a 属性的观点中所包含的情感
- h 观点持有者
- t 观点发布的时间

- 此定义适合于：基于属性的情感分析 (aspect-based sentiment analysis)

- 分析句子： *I hate property tax increases.*

- 你认为哪个是实体？哪个是属性？
- 实体是地方政府，房产税就是属性。

5. 观点的理由和限定条件

- 定义：观点的原因就是引起或触发这个观点的缘故，或是对观点之所以如此的解释。
- 定义：观点的限定条件可以限制和约束观点中的含义。
- 分析下面句子的情感原因和限定条件：
 - *This car is too small for a tall person.*
 - *I do not like the picture quality of this camera because the pictures are quite dark.*
 - *The picture quality of this camera is not good for night shots as the pictures are quite dark.*
 - *I am six feet five inches tall. This car is too small for me.*

6 情感分析的目标和任务

- 情感分析目标：给定一个包含观点信息的文档 d ，找出 d 中所有的观点五元组 (e,a,s,h,t) 。对于更高级的分析，还可以找出情感原因和限定条件。
- 情感分析的关键任务：
 - (1) 实体抽取
 - 定义2.10 实体类别和实体表达：一个实体类别是一个唯一特定的实体，然而一个实体表达或实体提及是段落中实际表示一个实体类别的词或短语。
 - (2) 属性抽取
 - 定义2.11 属性类别和属性表达：实体的一个属性类别是这个实体的一个唯一特定的属性，而一个属性表达或属性提及是段落中实际表示一个属性的词或短语。

✓ 属性表达的类型

- ✓ 定义2.12 **显示属性表达**: 在观点文本中以名词、名词短语出现的属性表达称为显示属性表达。

- *The picture quality of this camera is great*

- ✓ 定义2.13 **隐式属性表达**: 不是名词或名词短语, 但也指明了某个属性的文本表达称为隐式属性表达。

- *This camera is expensive.*

- *This restaurant closes too early.*

• 属性抽取是很难的任务

- *The Vacuum cleaner does not get the crumbs out of thick carpets.*

- 这里的属性比较复杂:

- 属性1是吸尘器吸走碎屑的能力, 属性2是对厚地毯的吸力。

• 给定一个观点文档集合D，情感分析的主要任务如下：

- (1) 实体抽取和消解
- (2) 属性抽取和消解
- (3) 观点持有者抽取和消解
- (4) 时间抽取和标准化
- (5) 属性的情感分类和回归
- (6) 生成观点五元组
- (7) 观点原因抽取和消解
- (8) 观点限定条件的抽取和消解

- 练习：找出观点五元组中的每一个值

Review B: posted by bigJohn Date: September 15, 2011

(1) I bought a Samsung camera and my friend brought a Canon camera yesterday.

(2) In the past week, We both used the cameras a lot.

(3) The photos from my Samy are not clear for night shots, and the battery life is short too.

(4) My friend was very happy with his camera and loves its picture quality.

(5) I want a camera that can take good photos.

(6) I am going to return it tomorrow.

• 练习:

*Review B: posted by **bigJohn** Date: **September 15, 2011***

*(1) I bought a **Samsung camera** and my friend brought a **Canon camera** yesterday.*

(2) In the past week, We both used the cameras a lot.

*(3) The photos from my **Samy** are not clear for night shots, and the **battery life** is short too. 🖐*

*(4) My friend was very happy with his camera and loves its **picture** quality. 👍*

*(5) I want a camera that can take good **photos**.*

(6) I am going to return it tomorrow.

例子

“我今年天让入手诺基亚5800，把玩不到24小时，目前感觉5800屏幕很好，操作也很方便，通话质量也不错，但是外形有些偏女性化，不适合男生。这些都是小问题，最主要的问题是电池不耐用，只能坚持一天，反正我觉得对不起这个价格。”



- 外形
- 电池



- 屏幕
- 操作
- 通话质量



2.2 观点摘要 (Opinion summary)

- 定义2.14: 基于属性的观点摘要:

对于实体 e 的基于属性的观点摘要结果具有如下形式:

General: 对实体 e 持正面观点的人数

对实体 e 持负面观点的人数

属性1: 对实体 e 的属性1持正面观点的人数

对实体 e 的属性1持负面观点的人数

.....

属性 n : 对实体 e 的属性 n 持正面观点的人数

对实体 e 的属性 n 持负面观点的人数

- 特点: 不同于传统的文本摘要, 观点摘要反映了用户正面或负面观点的数量

- 一种基于属性的观点摘要示例:

Digital_camera_1:

Feature: **picture quality**

Positive: 253

<individual review sentences>

Negative: 6

<individual review sentences>

Feature: **size**

Positive: 134

<individual review sentences>

Negative: 10

<individual review sentences>

...

✓ 还可以考虑：观点原因和
观点限定条件的类似摘要，
这个任务比较困难！

2.3 感情、情绪与心情

- 情绪类的情感(emotional sentiment):
 - 感情 (affect)
 - 情绪 (emotion)
 - 心情 (mood)
- 在心理学、哲学和社会学里，都有很多研究。
 - 侧重：心理状态、理论化的定义、基本情绪、生理反应等.....
- 在**NLP**和情感分析里，**侧重表达这些感受和情绪的语言词汇。**
 - 概念上的和研究上的一些困惑：不同的理论研究者对其常有不同的定义，甚至相互间对情绪、心情和感情的定义也不统一。

2.3.1 心理学的感情、情绪和心情

- 定义:

- 感情: 感觉或情绪, 尤指那些由面部表情或肢体语言所表达的情感。
- 情绪: 下意识的而不是经由意识作用产生的一种精神状态, 常伴有生理变化。
- 心情: 思维或情绪活动的一种状态。
- 感觉: 一种意识的情感状态, 例如从情绪、情感或渴望中产生的状态。

- 几个定义之间的关系:

- **感情**常定义为一个神经生理学的状态, 是情绪和心情最简单而原始 (未经反射) 的一种直觉反应。
- **情绪**是感情的一种表现, 是快速而强烈的感受, 心情是散漫而较久的感受。
 - 如愤怒的情绪-->烦躁的心情 (持续)
 - 情绪常由某个事件引起, 往往针对特定对象, 情绪有点属于理性的情感或者观点的范畴。
 - 心情往往由多个事件引起, 没有特定的原因或对象

情绪分类

- 心理学的理论学家对情绪的分类：
 - Arnold (1960)——anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
 - Ekman et al.(1982)——anger, disgust, fear, joy, sadness, surprise
 -
 - **Parrott(2001)——anger, fear, joy, love, sadness, surprise**



- ✓ 电影《头脑特工队》主人公大脑里的五种拟人化的情感：**欢乐、忧伤、愤怒、讨厌、害怕**

基本情绪	二级情绪	三级情绪
surprise	surprise	Amazement, astonishment
fear	horror	Alarm, fear, fright, horror, hysteria ...
	nervousness	Anxiety, apprehension, distress, dread, suspense...

• 人机情绪交互网络（**HUMAINE,2006**）

- 情绪标注和表示语言：**48**中情绪分为正面和负面的倾向

Negative and forceful	Negative and passive	Quite positive	Negative thoughts	Positive thoughts	Reactive
Anger Annoyance Contempt	Boredom Despair Disappointment	Calm Content Relaxed	Doubt Envy Frustration	Courage Hope Pride	Interest Politeness Surprised

- 在实际的应用系统中，这种分类方式很有价值。

情感分析中的感情、情绪和心情

- 对于情感分析任务：侧重研究这些感觉在自然语言中如何表达，以及如何识别的问题。
- 在选择语法和词汇时，通常有如下几种表达情绪和心情的方法：
 - 1、使用如**love, disgust, angry** 和 **upset**等描述情绪和心情的情感词
 - 2、描述情绪相关的行为
 - 3、使用强度词，如**very, so, extremely, really**等，强调词可能重复**very very**
 - 4、使用最高级，如：**This car is simply the best.**
 - 5、使用贬义词、褒义词以及表示反讽的词
 - 6、使用咒骂、诅咒、侮辱、责怪、指责和威胁性质的文本表达。

情绪的定义

- 定义：情绪是一个五元组 (e, a, m, f, t)
 - e 目标实体
 - a 是 e 的属性，是情绪所抒发的目标
 - m 是情绪的类型或是一个情绪类型与强度等级所组成的元组
 - f 是情绪的接受者
 - t 是情绪的表达时间
- 例： ***I am so upset with the manager of the hotel.***
 - 句子中的实体是 **hotel**，属性是 **the manager of the hotel**，情绪是愤怒，情绪的接收者是 **I**，如果知道时间可以加入到五元组中

情绪的致因

- 观点的原因 和 情绪的致因
 - *eg. I am so mad with the manager of the hotel because he refused to refund my booking fee.*
- 情绪的致因发现往往是隐含的，心情的致因在文中更是难以发现的。
 - *I felt a bit sad this morning.*
 - *There is sadness in her eyes.*
 - *I love this car.*

2.4 观点的不同类型

• 2.4.1 常规型观点(regular opinion)和比较型观点

- 常规型观点：直接观点和 间接观点

- 直接观点：直接对实体或实体属性表述的观点。

- *eg. The picture quality is great.*

- 间接观点：间接对实体或实体属性表述的观点。

- 如医药领域： *After injection of the drug, my joints felt worse.*

- 间接地表达了对药物的负面观点

- 如其他领域：

- *After switching to this laptop, my eyes felt much better.*

- 比较型观点：对两个或更多实体之间的相同或不同点进行比较，表达了观点持有者的一个偏好，以及包含了观点持有者在面对一些实体的共有属性时的偏好。

- *eg. Coke tastes better than Pepsi.*

- **2.4.2 主观的和隐含在事实中的观点**
- 主观观点 (**subjective opinion**)
 - 在主观陈述中表达的常规型观点和比较型观点
- 隐含在事实中的观点 (**fact-implied opinion**)
 - 通常是一个客观或真实的事实陈述中蕴含的常规型或比较型观点。
 - (1) 隐含在个人事实中的观点

The battery life of this phone is longer than my previous Samsung phone.
 - (2) 隐含在非个人事实中的观点

The unemployment rate came down last week.
Google made more money than Yahoo! Last month.

2.5 作者视角和读者视角

- 从不同的角度来审视一个观点：
 - *This car is too small for me.*
 - 作者视角：车子比较小，负面观点。
 - 读者视角：可能是正面观点
 - *Google's profits went up by 30%.*
 - 作者视角：正面观点。
 - 竞争者的视角：负面观点
- 在实际系统中，根据需求来判断观点。

2.6 小结

- 什么是观点 和 情感？
- 情感分析的主要任务是什么？
- 典型的观点摘要？
- 如何研究感情、情绪和心情？