Natural Language Processing I

1. 导论

WU Xiaokun 吴晓堃

xkun.wu [at] gmail

2022/03/11

什么是自然语言处理

人机对话

Dave Bowman: Open the pod bay doors, HAL.
HAL: I'm sorry Dave, I'm afraid I can't do that.

– Kubrick and Clarke 2001, A Space Odyssey.

Dave Bowman: HAL，请你打开太空舱的分离舱门。
HAL: 对不起，Dave，我不能这样做。

– Kubrick and Clarke 2001，《太空漫游》.

学科定义

自然语言处理 Natural Language Processing (NLP)

- 融合计算机科学、语言学、人工智能的**交叉学科**
  - NLP任务：用计算机处理人类语言
  - 主要研究方法是统计学和人工智能
- **终极目标**：让机器理解人类语言，实现**人机交流**
  - 从文本、语言中提取有用信息，辅助沟通交流
  - 通过语言进行人机交互

学科定义

自然语言处理 Natural Language Processing (NLP)

- 融合计算机科学、语言学、人工智能的**交叉学科**
  - NLP任务：用计算机处理人类语言
  - 主要研究方法是统计学和人工智能
- **终极目标**：让机器理解人类语言，实现**人机交流**
  - 从文本、语言中提取有用信息，辅助沟通交流
  - 通过语言进行人机交互

侧重语言学结构时，也称**计算语言学 Computational Linguistics**

- 使用计算机来辅助研究语言学

哪些任务可以被称为NLP?

- 打字，文本处理，拼写检查
- 语音识别，光学字符识别 OCR
- 文本分类，情感分析
- 机器翻译，同声传送
- 语音助手，聊天机器人
- 语言理解，机器问答，机器推断

NLP 的重要性

现代日常生活高度依赖信息交流

- 信息搜索，B站，抖音，外卖

NLP 的重要性

现代日常生活高度依赖信息交流

- 信息搜索，B站，抖音，外卖

高级语言系统是**人类智能**的主要特征之一

- 其他动物的声音不能称为语言系统

NLP 的重要性

现代日常生活高度依赖信息交流

- 信息搜索，B站，抖音，外卖

高级语言系统是**人类智能**的主要特征之一

- 其他动物的声音不能称为语言系统

很多信息时代的技术是由语言载体来支持的

- 网页排序，广告，推荐系统，语言翻译，问答

NLP 的重要性

现代日常生活高度依赖信息交流

- 信息搜索，B站，抖音，外卖

高级语言系统是**人类智能**的主要特征之一

- 其他动物的声音不能称为语言系统

很多信息时代的技术是由语言载体来支持的

- 网页排序，广告，推荐系统，语言翻译，问答

完美地实现语言理解**等价于实现人工智能**

Turing测试 [Turing 1950]

测试者与被测试者（一个人和一台机器）隔开的情况下，通过一些装置（如键盘）向被测试者随意提问。

为什么NLP很难?

语言演化的终极目标是**高效沟通**与**语义准确**之间的平衡

为什么NLP很难?

语言演化的终极目标是**高效沟通**与**语义准确**之间的平衡

- 古汉语：信息高度压缩，时代背景

> *生而不有（生养万物而不据为己有）*
> *为而不恃（竭尽全力而不自恃已能）*
> *–《道德经》*

为什么NLP很难?

语言演化的终极目标是**高效沟通**与**语义准确**之间的平衡

- 古汉语：信息高度压缩，时代背景

> *生而不有（生养万物而不据为己有）*
> *为而不恃（竭尽全力而不自恃已能）*
> *–《道德经》*

出生时一无所有？？？
做起事来有恃无恐？？？

为什么NLP很难?

语言演化的终极目标是**高效沟通**与**语义准确**之间的平衡

- 古汉语：信息高度压缩，时代背景

> *生而不有（生养万物而不据为己有）*
> *为而不恃（竭尽全力而不自恃已能）*
> *–《道德经》*

> 出生时一无所有？？？
> 做起事来有恃无恐？？？

- 语音：信号传递必然丢失信息
  - 在商场需要多大功率的喇叭？提示：分贝是对数比率

为什么NLP很难?

语言演化的终极目标是**高效沟通**与**语义准确**之间的平衡

- 古汉语：信息高度压缩，时代背景

> *生而不有（生养万物而不据为己有）*
> *为而不恃（竭尽全力而不自恃已能）*
> *–《道德经》*

> 出生时一无所有？？？
> 做起事来有恃无恐？？？

- 语音：信号传递必然丢失信息
  - 在商场需要多大功率的喇叭？提示：分贝是对数比率

人类对自己的语音系统太熟悉，就很难体会其复杂程度

示例：歧义

介词短语：高阶英语学习的最大难点

> *One morning I shot an elephant in my pajamas.*
> *How he got into my pajamas I don't know.*
>
> *– Groucho Marx, Animal Crackers, 1930*

示例：歧义

介词短语：高阶英语学习的最大难点

> *One morning I shot an elephant in my pajamas.*
> *How he got into my pajamas I don't know.*
>
> *– Groucho Marx, Animal Crackers, 1930*

- 多义词：方便的时候，给我打电话啊。

示例：歧义

介词短语：高阶英语学习的最大难点

> *One morning I shot an elephant in my pajamas.*
> *How he got into my pajamas I don't know.*
>
> *– Groucho Marx, Animal Crackers, 1930*

- 多义词：方便的时候，给我打电话啊。
- 断句：结婚的和尚未结婚的确实在干扰分词。

示例：歧义

介词短语：高阶英语学习的最大难点

> *One morning I shot an elephant in my pajamas.*
> *How he got into my pajamas I don't know.*
>
> *– Groucho Marx, Animal Crackers, 1930*

- 多义词：方便的时候，给我打电话啊。
- 断句：结婚的和尚未结婚的确实在干扰分词。
- 多音字：这种食物可以`zhì'ái`。-- “致癌”还是“治癌”？

歧义技术分析

The chef made her duck

- 厨师迫使她躲闪
  - 词类: “duck” 可以是**名词、动词**

歧义技术分析

The chef made her duck

- 厨师迫使她躲闪
  - 词类: “duck” 可以是**名词、动词**
- 二格、三格代词
  - possessive (“of her”)：厨师把**属于她**的鸭子烤了
  - dative (“for her”)：厨师**给她**做了烤鸭

歧义技术分析

The chef made her duck

- 厨师迫使她躲闪
  - 词类: “duck” 可以是**名词、动词**
- 二格、三格代词
  - possessive (“of her”)：厨师把**属于她**的鸭子烤了
  - dative (“for her”)：厨师**给她**做了烤鸭
- 厨师把**自己的**鸭子烤了
  - 共指: “她”可以指厨师自己

歧义技术分析

The chef made her duck

- 厨师迫使她躲闪
  - 词类: “duck” 可以是**名词、动词**
- 二格、三格代词
  - possessive (“of her”)：厨师把**属于她**的鸭子烤了
  - dative (“for her”)：厨师**给她**做了烤鸭
- 厨师把**自己的**鸭子烤了
  - 共指: “她”可以指厨师自己
- 厨师做了个小黄鸭雕塑
  - 词义：“make” 可以指“烤”，或“创作”

示例：机器翻译

你	说	他	不	行	,	你	行	你	上	啊
You	say	he	no	can	,	you	can	you	up	ah
You	say	he	can	not	,	you	can	you	on	ah

示例：机器翻译

你	说	他	不	行	,	你	行	你	上	啊
You	say	he	no	can	,	you	can	you	up	ah
You	say	he	can	not	,	you	can	you	on	ah

DeepL Translator

- https://www.deepl.com/translator#zh/en/
- 如何鉴定出好的机器翻译?
  - 图样图森破

示例：机器翻译

你	说	他	不	行	,	你	行	你	上	啊
You	say	he	no	can	,	you	can	you	up	ah
You	say	he	can	not	,	you	can	you	on	ah

DeepL Translator

- https://www.deepl.com/translator#zh/en/
- 如何鉴定出好的机器翻译?
  - 图样图森破

思考：如何正确翻译?

示例：内涵解析

黑话只在特定人群中使用

> 土匪：天王盖地虎！（你好大的胆！敢来气你的祖宗？）
> 杨子荣：宝塔镇河妖！（要是那样，叫我从山上摔死，掉河里淹死。）
> --《林海雪原》

示例：内涵解析

黑话只在特定人群中使用

土匪：天王盖地虎！（你好大的胆！敢来气你的祖宗？）
杨子荣：宝塔镇河妖！（要是那样，叫我从山上摔死，掉河里淹死。）
--《林海雪原》

翻译难度非常高

台词：Our master lords over tigers, Our pagoda seals river monsters.
机翻：The king of heaven covers the tiger, the pagoda to suppress the river demon.

示例：内涵解析应用

艺术源于生活。

> *老板：小伙子，好好干！只要这个月部门的业绩能达标，到时候嘛，你懂的~*
> *（说罢，还看了总监位置一眼。）*
> *–《万万没想到》*

示例：内涵解析应用

艺术源于生活。

> *老板：小伙子，好好干！只要这个月部门的业绩能达标，到时候嘛，你懂的~*
> *（说罢，还看了总监位置一眼。）*
> *–《万万没想到》*

> *张麻子：翻译出来给我听，什么XXX叫惊喜！什么XXX叫XXX惊喜！*
> *–《让子弹飞》*

示例：内涵解析应用

艺术源于生活。

> *老板：小伙子，好好干！只要这个月部门的业绩能达标，到时候嘛，你懂的~*
> *（说罢，还看了总监位置一眼。）*
> *–《万万没想到》*

> *张麻子：翻译出来给我听，什么XXX叫惊喜！什么XXX叫XXX惊喜！*
> *–《让子弹飞》*

> *……文艺作品中反映出来的生活却可以而且应该比普通的实际生活更高，更强烈，更有集中性，更典型，更理想，因此就更带**普遍性**。*
> *–毛泽东《在延安文艺座谈会上的讲话》，1942年*

示例：机器问答

主体身份误判

Q：第一次去动物园应该注意什么？
A：记得要食物—别只知道卖萌。

示例：机器问答

主体身份误判

```
Q：第一次去动物园应该注意什么？
A：记得要食物—别只知道卖萌。
```

暴力、淫秽、种族歧视、极右翼言论

- 案例

示例：机器问答

主体身份误判

```
Q：第一次去动物园应该注意什么？
A：记得要食物—别只知道卖萌。
```

暴力、淫秽、种族歧视、极右翼言论

- 案例

《客服人员标准礼貌用语》

```
x先生/小姐，非常感谢您为我们提供的宝贵意见，我们将尽快向有关部门反映，希望您继续对x的服务给予关注和支持。
```

示例：机器问答

主体身份误判

> Q：第一次去动物园应该注意什么？
> A：记得要食物—别只知道卖萌。

暴力、淫秽、种族歧视、极右翼言论

- 案例

《客服人员标准礼貌用语》

> x先生/小姐，非常感谢您为我们提供的宝贵意见，我们将尽快向有关部门反映，希望您继续对x的服务给予关注和支持。

> *热情，礼貌，但一问三不知。–《人民的名义》*

自然语言与编程语言

词汇量

编程语言中称为**关键字**

- C语言只有32个
- 变量名、函数名在编译后是没有本质区别的

词汇量

编程语言中称为**关键字**

- C语言只有32个
- 变量名、函数名在编译后是没有本质区别的

自然语言的词汇量可以是无限多

- 2013年版《通用规范汉字表》中收字8105个，不包括繁体字、异体字
- 新词被不断创造出来

结构化

编程语言是结构化的

```
class MLP(nn.Module):
    def __init__(self, name, dim):
        super().__init__()
        self.name = name
        self.dim = dim

mlp = MLP(name="单层感知机", dim=[20, 256, 10])
```

结构化

编程语言是结构化的

```
class MLP(nn.Module):
    def __init__(self, name, dim):
        super().__init__()
        self.name = name
        self.dim = dim

mlp = MLP(name="单层感知机", dim=[20, 256, 10])
```

自然语言处理任务：“模型选用单层感知机，其中隐藏层的维度是256”

- 中文分词：[单层 （的） 感知机 （模型）]
- 词类标注：“感知”不能拆解成“动词+名词”
- 命名实体识别：“隐藏层”是机器学习术语
- 指代消解：“其中”指代“单层感知机”

语法严格

编程语言不存在歧义性，否则无法编译执行

语法严格

编程语言不存在歧义性，否则无法编译执行

自然语言存在大量歧义

```
请解释下文中每个“意思”的意思。
领导：“你这是什么意思？”
阿呆：“没什么意思，意思意思。”
领导：“你这就不够意思了。”
阿呆：“小意思，小意思。”
领导：“你这人真有意思。”
阿呆：“其实也没有别的意思。”
领导：“那我就不好意思了。”
阿呆：“是我不好意思。”
```

容错性

编程语言中拼写错误导致无法编译或潜在bug

容错性

编程语言中拼写错误导致无法编译或潜在bug

自然语言很难避免拼写、语法错误

> 妍表究明，汉子的序顺并不定一能影像读阅。比如，当你看完这话句后，会发现这面里的字全是挫乱的。

- 但人类通常可以猜出意思
  - 省略常识，简略表达：高效沟通
  - 注意：主观推断不一定是好事，例如意识形态、吵架
- 注意：语序错乱不影响阅读是中文（分析语）的语法特色

易变性

编程语言有维护标准，更新非常缓慢

- C++标准：98、03、11、14
- 学习成本：人们更愿意淘汰旧语言

易变性

编程语言有维护标准，更新非常缓慢

- C++标准：98、03、11、14
- 学习成本：人们更愿意淘汰旧语言

自然语言是地区约定俗成的

- 古汉语和现代汉语差异巨大
- 拉丁语系分化

软件工具

HanLP

HanLP：面向生产环境的自然语言处理工具包

- 在线演示：https://hanlp.hankcs.com/
- 源码：https://github.com/hankcs/HanLP
  - 分词
  - 词类标注
  - 命名实体识别
  - 依存句法分析
  - 成分句法分析
  - 语义依存分析

实验：HanLP演示

```
HanLP/plugins/hanlp_demo/hanlp_demo/zh
```

- 源码：https://github.com/hankcs/HanLP
  - 分词
  - 词类标注
  - 命名实体识别
  - 依存句法分析
  - 成分句法分析
  - 语义依存分析

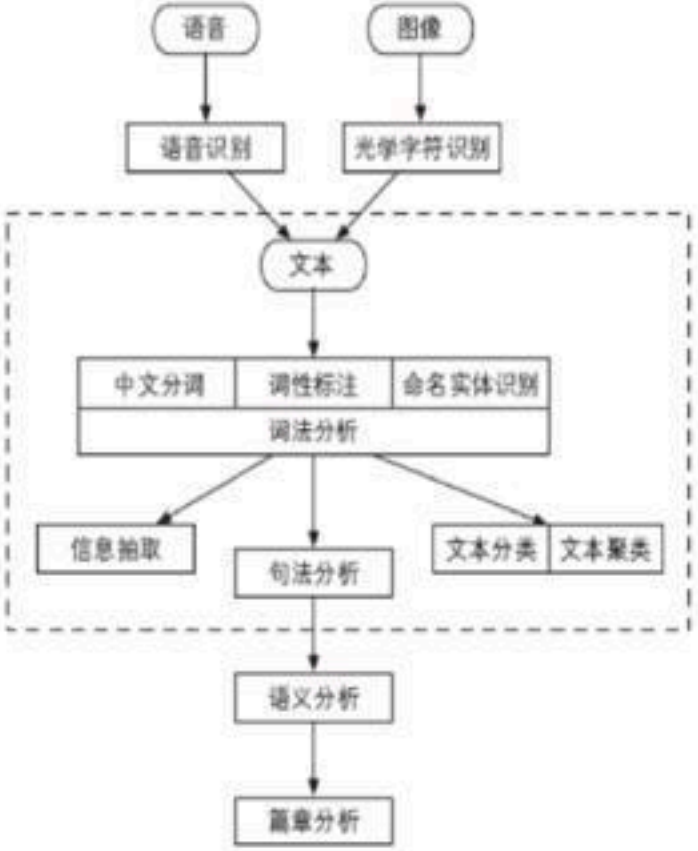
深度学习框架

TensorFlow：https://www.tensorflow.org/

PyTorch：https://pytorch.org/

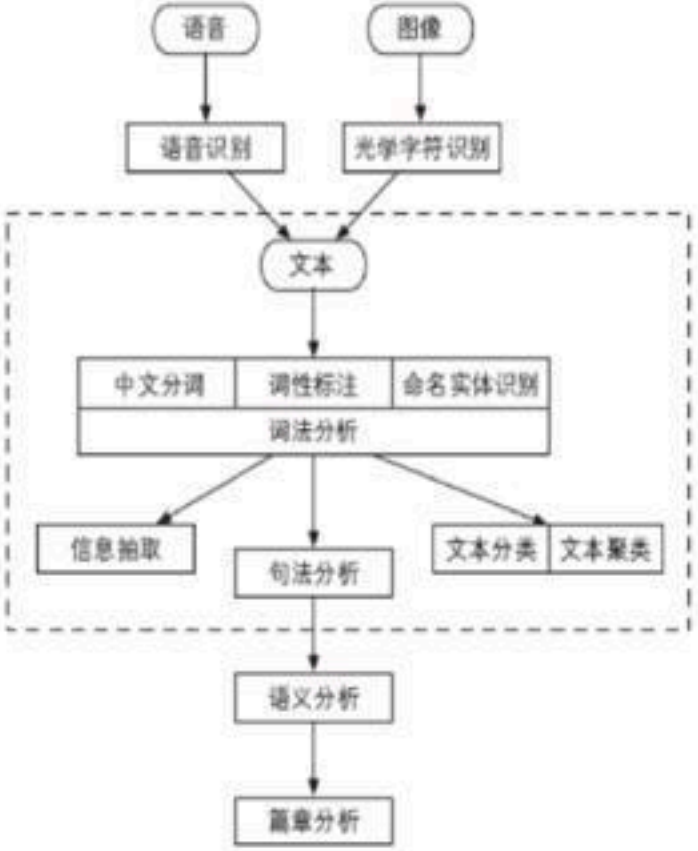
NLP 任务

NLP 任务层次



- 从语言、图像到文本
- 词法分析
  - 中文分词、词类标注、命名实体识别
- 句法分析
- 篇章分析

NLP 课题



- 信息抽取
- 文本分类、聚类
- 复杂任务
  - 自动问答、自动摘要、机器翻译、推断等

NLP 流派与历史

NLP 简史

时间	关键点	代表人物、技术
1940 - 1954	电子计算机发明，智能理论构建	Turing, Chomsky
1954 - 1970	形式化规则，逻辑理论，感知机	Prolog, Rosenblatt
1970 - 1980	HMM语音识别，语义和篇章建模	Jelinek
1980 - 1991	大规模规则知识库	WordNet (1985)
1991 - 2008	统计建模和机器学习	SVM, PageRank, 问答系统
2008 - now	大数据和深度学习	词嵌入，翻译，聊天

基于规则的专家系统

专家系统：基于规则，即由专家手工指定的确定性流程。

基于规则的专家系统

专家系统：基于规则，即由专家手工指定的确定性流程。

案例：波特词干算法 Porter stemming algorithm

IF	AND	后缀替换	例子
eed	辅音+元音同时出现	ee	agreed -> agree
ed	含辅音	空白	plastered -> plaster
ing	含辅音	空白	eating -> eat

基于规则的专家系统

专家系统：基于规则，即由专家手工指定的确定性流程。

案例：波特词干算法 Porter stemming algorithm

IF	AND	后缀替换	例子
eed	辅音+元音同时出现	ee	agreed -> agree
ed	含辅音	空白	plastered -> plaster
ing	含辅音	空白	eating -> eat

问题：维护（专家）成本高，难以拓展、更新

基于统计的学习方法

降低对专家的依赖，自动适应语言演化

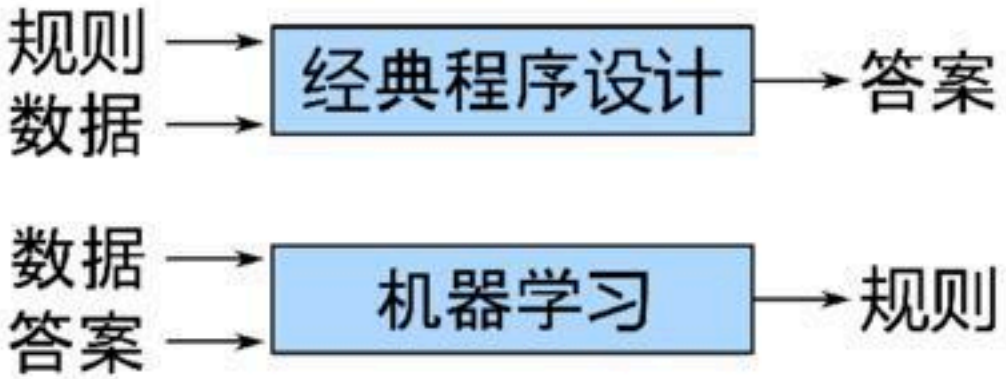
- 统计：在语料库上进行
- 语料库：泛指人工标注过的结构化文本

基于统计的学习方法

降低对专家的依赖，自动适应语言演化

- 统计：在语料库上进行
- 语料库：泛指人工标注过的结构化文本

编程范式



NLP 简史 I

时间	关键点	代表人物、技术
1940 - 1954	电子计算机发明，智能理论构建	Turing, Chomsky

Turing 1950, Computing Machinery and Intelligence

- Turing测试：人工智能的充分条件

NLP 简史 I

时间	关键点	代表人物、技术
1940 - 1954	电子计算机发明，智能理论构建	Turing, Chomsky

Turing 1950, Computing Machinery and Intelligence

- Turing测试：人工智能的充分条件

Chomsky 1957, Syntactic Structures

- 句子：语境无关的语法规则生成

NLP 简史 I

时间	关键点	代表人物、技术
1940 - 1954	电子计算机发明，智能理论构建	Turing, Chomsky

Turing 1950, Computing Machinery and Intelligence

- Turing测试：人工智能的充分条件

Chomsky 1957, Syntactic Structures

- 句子：语境无关的语法规则生成

Minsky 1951：首台模拟神经网络的机器

NLP 简史 II

时间	关键点	代表人物、技术
1954 - 1970	形式化规则，逻辑理论，感知机	Prolog, Rosenblatt

MIT AI, BASEBALL

- “如果句子中不含有其他动词，则 score 是一个动词， 否则是名词。”

NLP 简史 II

时间	关键点	代表人物、技术
1954 - 1970	形式化规则，逻辑理论，感知机	Prolog, Rosenblatt

MIT AI, BASEBALL

- “如果句子中不含有其他动词，则 score 是一个动词， 否则是名词。”

规则系统僵硬严格，被称为“玩具”

- 只能处理固定的问句，无法处理与或非逻辑等

NLP 简史 II

时间	关键点	代表人物、技术
1954 - 1970	形式化规则，逻辑理论，感知机	Prolog, Rosenblatt

MIT AI, BASEBALL

- “如果句子中不含有其他动词，则 score 是一个动词， 否则是名词。”

规则系统僵硬严格，被称为“玩具”

- 只能处理固定的问句，无法处理与或非逻辑等

Prolog (Programming in Logic) 1972：构建知识库及专家系统

NLP 简史 II

时间	关键点	代表人物、技术
1954 - 1970	形式化规则，逻辑理论，感知机	Prolog, Rosenblatt

MIT AI, BASEBALL

- “如果句子中不含有其他动词，则 score 是一个动词， 否则是名词。”

规则系统僵硬严格，被称为“玩具”

- 只能处理固定的问句，无法处理与或非逻辑等

Prolog (Programming in Logic) 1972：构建知识库及专家系统

Rosenblatt 1958：感知机

NLP 简史 III

时间	关键点	代表人物、技术
1970 - 1980	HMM语音识别，语义和篇章建模	Jelinek

Jelinek 1976, Continuous Speech Recognition by Statistical Methods

- 隐式Markov模型 HMM

NLP 简史 III

时间	关键点	代表人物、技术
1970 - 1980	HMM语音识别，语义和篇章建模	Jelinek

Jelinek 1976, Continuous Speech Recognition by Statistical Methods

- 隐式Markov模型 HMM

理想破灭导致第一次人工智能冬天

NLP 简史 IV

时间	关键点	代表人物、技术
1980 - 1991	大规模规则知识库	WordNet (1985)

专家系统再次兴起，商业化发展迅猛

- 根本原因：基于晶体管的集成电路技术成熟
- WordNet：大规模词汇数据库

NLP 简史 IV

时间	关键点	代表人物、技术
1980 - 1991	大规模规则知识库	WordNet (1985)

专家系统再次兴起，商业化发展迅猛

- 根本原因：基于晶体管的集成电路技术成熟
- WordNet：大规模词汇数据库

维护成本过高导致第二次人工智能冬天

NLP 简史 IV

时间	关键点	代表人物、技术
1980 - 1991	大规模规则知识库	WordNet (1985)

专家系统再次兴起，商业化发展迅猛

- 根本原因：基于晶体管的集成电路技术成熟
- WordNet：大规模词汇数据库

维护成本过高导致第二次人工智能冬天

LeCun 1989：深度卷积神经网络

NLP 简史 V

时间	关键点	代表人物、技术
1991 - 2008	统计建模和机器学习	SVM, PageRank, 问答系统

互联网的出现带来统计建模的热潮

- 海量数据

NLP 简史 V

时间	关键点	代表人物、技术
1991 - 2008	统计建模和机器学习	SVM, PageRank, 问答系统

互联网的出现带来统计建模的热潮

- 海量数据

领域专家作用减弱

- 特征工程：为统计模型设计特征，即将数据表达为计算机易于处理的格式

NLP 简史 VI

时间	关键点	代表人物、技术
2008 - now	大数据和深度学习	词嵌入，翻译，聊天

计算机算力提升带来神经网络的复兴

- 专用芯片：并行化矩阵计算
- 深度学习成为人工智能领域的核心技术

课程学习目标

课程计划

本课程计划讲解如下应用的实现：

- 情感分类
- 机器翻译
- 简单推断
- 聊天机器人

提取情感信息

成年人的对话常常带有内涵

- 《大明王朝1566》剧本

嘉靖帝：胡宗宪呢?
杨金水：他不是织造局的人（胡宗宪没贪钱）
嘉靖帝：吕芳呢?
杨金水：他是谁？（装傻，保吕芳；吕芳是你的忠心奴仆，怎么还怀疑他？）
嘉靖帝：就是杨金水他们口里的老祖宗，给你请六品顶戴的人！（吕芳不可能跟沈一石没有利益往来）
杨金水：有他，他在一百年前就死掉了。（那是很久之前的事，跟这次案件无关）。
嘉靖帝：你说了这么多人，为什么不说杨金水？（事已至此，你打算怎么办？）
杨金水：杨金水也死了。他害死了我，我已经把他也带走了。（曾经的杨金水已经死了，

情感分析

提取内涵的简单版本：**情感分析 sentiment analysis**

酒店评价

- 前台态度非常好！早餐很丰富，房间很干净。
- 结果大失所望，灯光昏暗，空间狭小，房间有霉味。

差评心理学

经历创伤后的文字

- 被动语态：强调疏远感
- 大量使用“we”：寻求公众同情、安慰

```
...we were ignored until we flagged down a waiter to get our
waitress...
```

差评心理学

经历创伤后的文字

- 被动语态：强调疏远感
- 大量使用“we”：寻求公众同情、安慰

```
...we were ignored until we flagged down a waiter to get our
waitress...
```

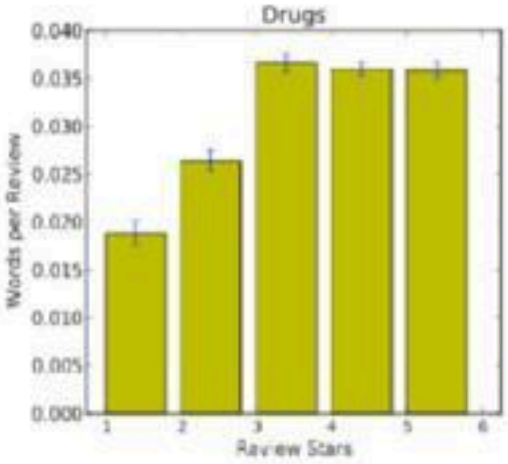
差评可以看作**创伤陈述**

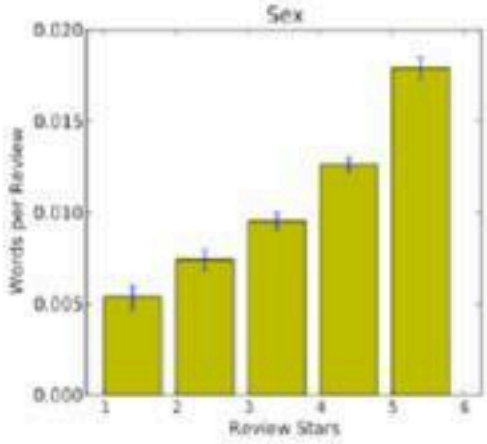
- 推论：人际互动（客服）很关键

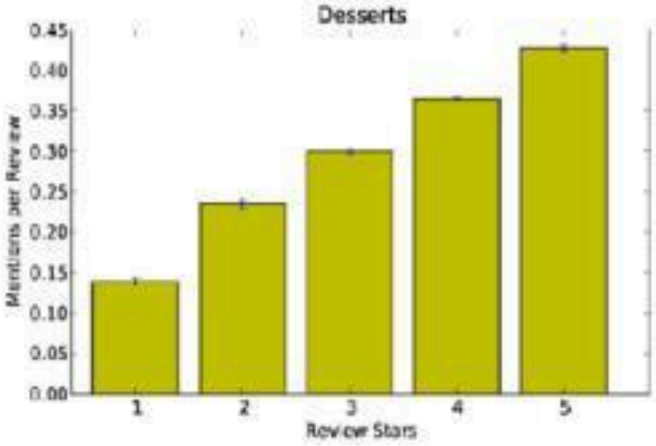
（英语）好评关键字

人的原始欲望：Drugs, Sex, and Dessert

Jurafsky 2014, Narrative framing of consumer sentiment in online restaurant reviews



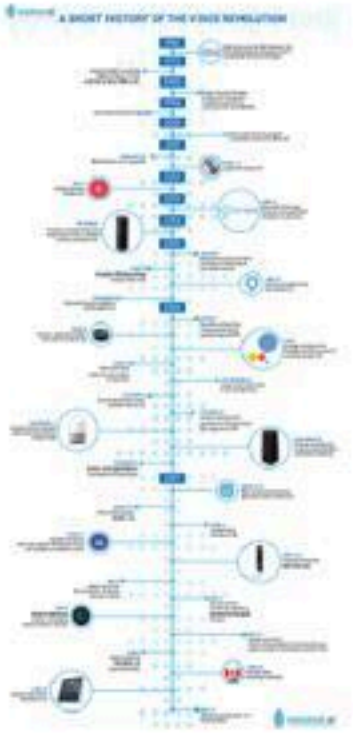




语音助手

Siri, Cortana, Echo

时间线：
https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/



语音助手的挑战

Siri vs. Cortana

Review

本章内容

NLP任务的困难原因。NLP任务的层次及课题。NLP的流派与历史。正则表达式。文本正则化。最小编辑距离及动态规划算法。

重点：NLP任务的层次及课题；正则表达式；文本正则化。

难点：最小编辑距离。

学习目标

- 理解NLP任务的困难原因。
- 掌握调用HanLP预训练模型解决NLP任务的方法。
- 理解NLP任务的几个层次，及其主要课题。
- 掌握使用正则表达式进行模板匹配与替换的方法。
- 了解ELIZA的实现原理。
- 理解文本正则化的几个方面。
- 了解最小编辑距离的动态规划解法。

问题

列举NLP任务的几个层次，及其主要课题。

举出至少三个NLP任务出错的例子，解释其原因，并尝试提出解决方案。

结合NLP任务列举文本正则化的重要步骤。

(*) 列举字符编辑的操作和成本，并计算最小编辑距离："leda"、"deal"。

课程项目

情感分析：毕设评语分类系统

- 现阶段：数据准备

(*) 模仿ELIZA的原理实现一个聊天机器人。

- 中文或英文
- 提示：使用正则表达式进行模板匹配与替换

Appendix

复杂内容表述力

现代汉语：由于A在B和C间造成的差值太大，相比于A在D和E间造成的差值以及F和G的差值，这个数值不该太明显。

英语：Because the difference between B and C caused by A is too large, compared with the difference between D and e caused by A and the difference between F and G, this value should not be too obvious.

德语：Da der durch A verursachte Unterschied zwischen B und C im Vergleich zu dem durch A verursachten Unterschied zwischen D und e und dem Unterschied zwischen F und G zu groß ist, sollte dieser Wert nicht zu offensichtlich sein.

文言：因甲致于乙、丙间差值太大也，比甲致于丁、癸及丑、亥之间差，其数不可太明。

可看出，关于这个复杂逻辑内容，文言文的表述力、简洁度、易理解度最强。

- 四个语言在这个测验中的表述力排名大概是文言>英语=现代汉语>=德语

信息密度

[Pellegrino 2011, Coupé 2019]