自然语言处理、计算与理解

宗成庆

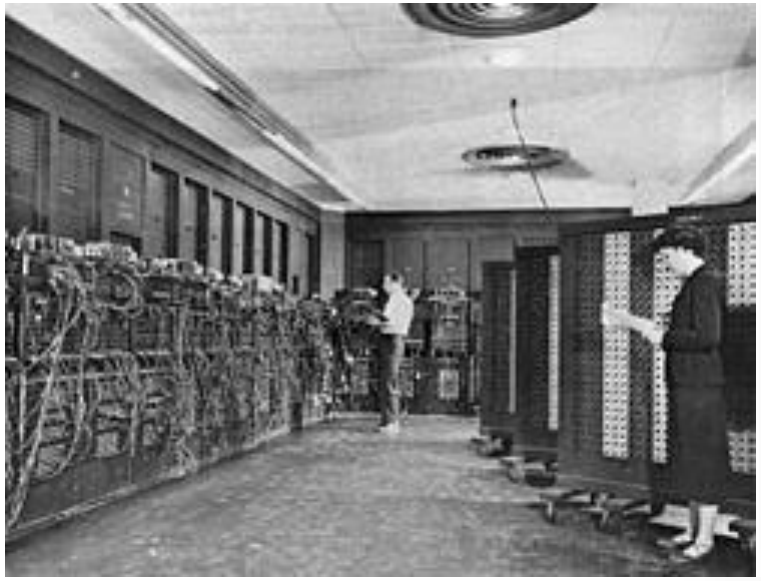
中国科学院自动化研究所
模式识别国家重点实验室

cqzong@nlpr.ia.ac.cn

内容提要

1. 引言
2. NLP方法概述
3. 深度学习方法应用
4. 讨论与结语

1. 引言



1946年，世界上第一台计算机ENIAC诞生

Warren Weaver (July 17, 1894 – Nov. 24, 1978)

- ✧ 信息论先驱
- ✧ 1920至1932年Wisconsin大学数学教授
- ✧ 1932至1955年担任Rockefeller Institute自然科学部主任

- ◆ A. D. Booth 数学物理学家，二战中参与计算机研制，在程序化计算机研究中成绩卓著；
- ◆ 1947年3月至9月，曾在普林斯顿大学参与 John von Neumann 研究组，后来曾在伦敦大学工作。

1. 引言

诺伯特·维纳 (Norbert Wiener) (1894年11月26日～1964年3月18日)

[Reproduced by permission of the Rockefeller Foundation Archives]

March 4, 1947

Dear Norbert:

I was terribly sorry, when in Cambridge recently, that I got unavoidably held up by several unexpected jobs, and did not get a chance to see you.

One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.

I wondered if it were unthinkable to design a computer which would translate

Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Have you ever thought about this? As a linguist and expert on computers, do you think it is worth thinking about?

Cordially,

Warren Weaver.

Professor Norbert Wiener
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

WW:AEB

1. 引言

达特茅斯学院 (Dartmouth College)
(成立于1769年)

左起：摩尔、麦卡锡、明斯基、
赛弗里奇(Oliver Selfridge)、所罗门诺夫

人工智能夏季研讨会（大茅斯会议, 1956）
Summer Research Project on Artificial Intelligence (Dartmouth Conference)

1. 引言

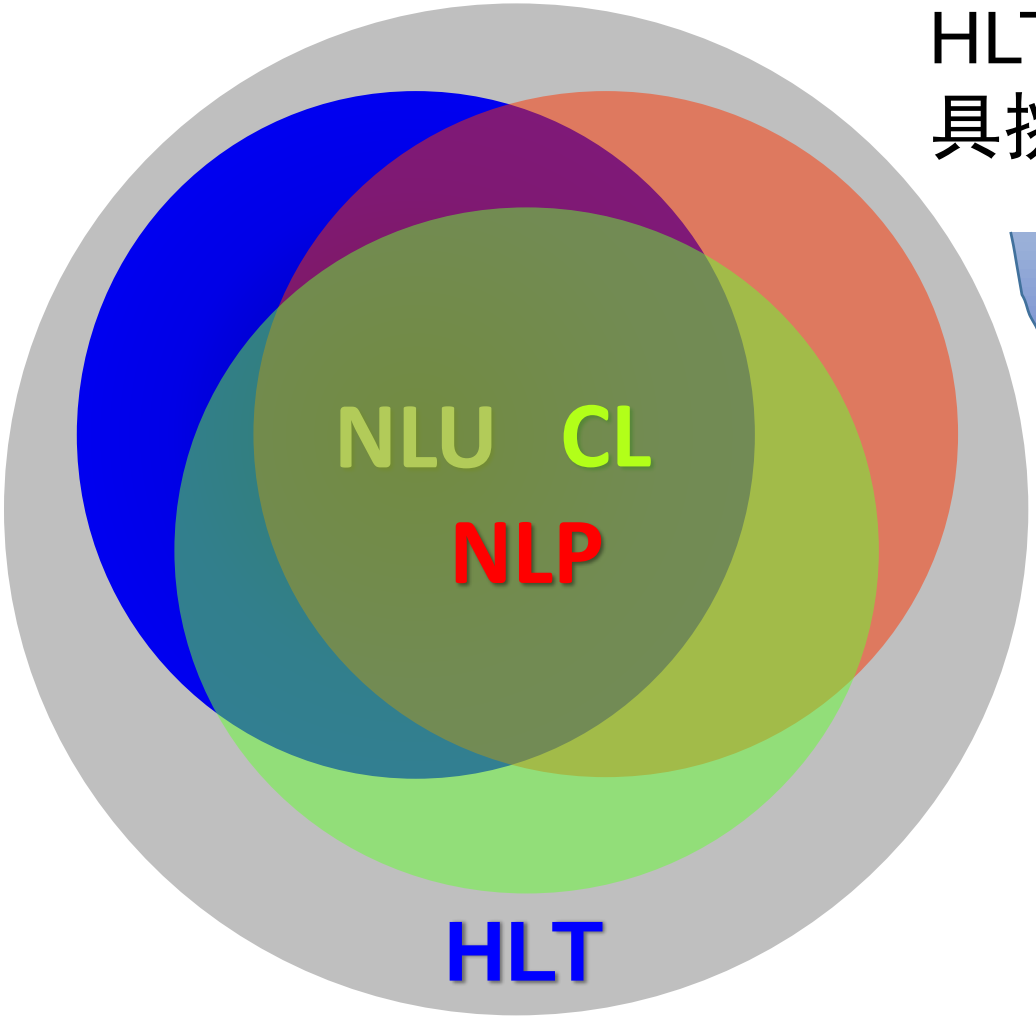
- **自然语言理解**(natural language understanding, NLU)是人工智能最重要的研究方向之一
- **计算语言学** (Computational Linguistics, CL)

  1960S，形成相对独立的学科。1962年国际计算语言学学会(ACL)成立，1965年国际计算语言学委员会(ICCL)成立，1966年“计算语言学”首次出现在美国国家科学院ALPAC报告里
- **自然语言处理** (Natural Language Processing, NLP)

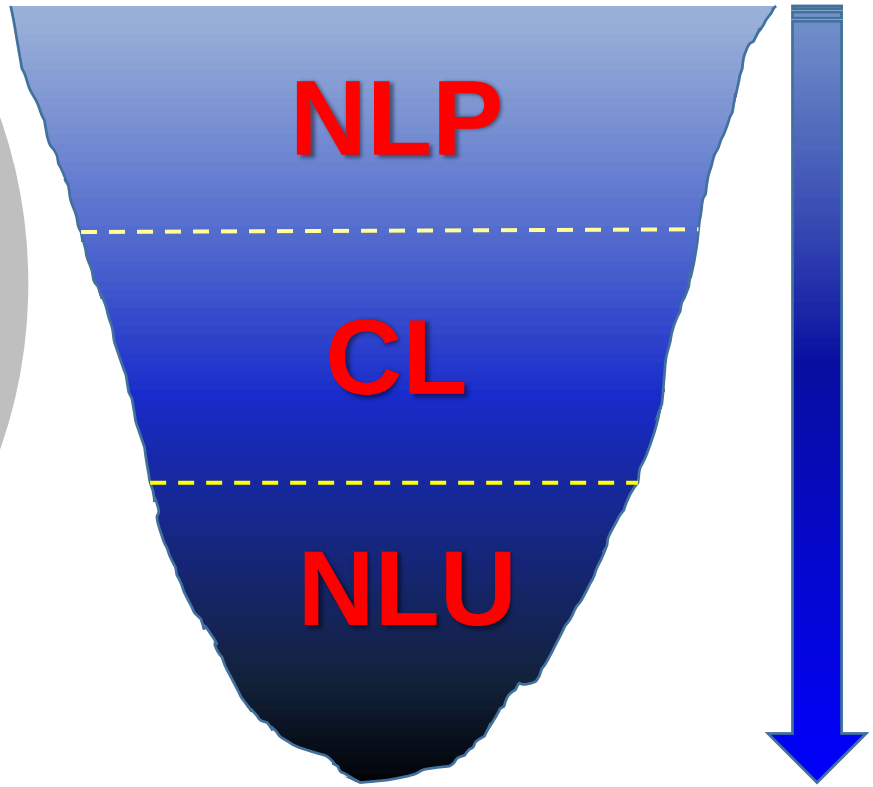
  1980S，面向计算机网络和移到通信，从系统实现和语言工程的角度开展语言信息处理方法的研究。专门针对中文的语言信息技术研究成为中文信息处理

NLU、CL和NLP统称为人类语言技术(Human Language Technology, HLT)

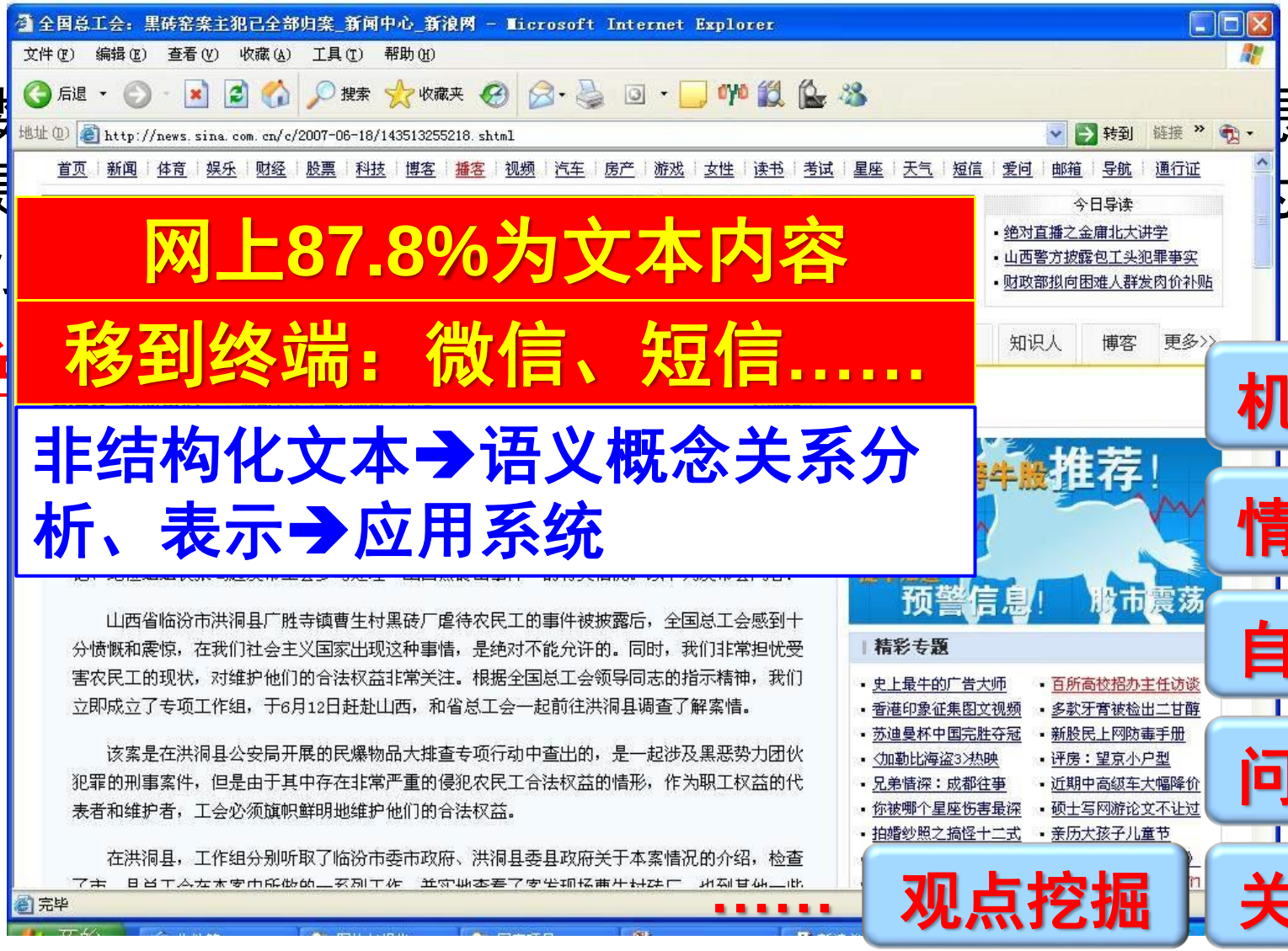
1. 引言



HLT 是当前人工智能领域最具挑战性的研究方向之一。



1. 引言



……

机器翻译

情感分析

自动摘要

问答系统

观点挖掘

关系抽取

1. 引言

全球数万亿网页，
80%非汉语文字

出境游人数破亿，前
20出境游目的地
有12种语言

64个国家和地区
44亿人口
50多种语言

1. 引言

ICML’2015

6-11 July 2015, Lille, France

1. 引言


Panel Discussion
Deep Learning Workshop ICML'15
Kevin Murphy
Google
Neil Lawrence
U. Sheffield
Dennis Hassabis
Google DeepMind

1. 引言

At DL 2015, Neil Lawrence said …

“NLP is kind of like a rabbit in the headlights of the deep learning machine, waiting to be flattened.”

A Professor of Machine Learning
at the University of Sheffield

1. 引言

1. 引言

问题与挑战

- 大量的未知现象
  如：高山，埃博拉，奥特
- 无处不在的歧义词汇
  如：苹果, 粉丝, Bank
- 复杂或歧义结构比比皆是
  喜欢乡下的孩子。　　Time flies like an arrow.
- 普遍存在的缩略和隐喻表达
  要把权力装进制度的笼子；老虎苍蝇一起打。
  破四旧，除四害；消灭一切牛鬼蛇神。

1. 引言

问题与挑战

- **跨语言语义概念不对等**

  如：馒头： steamed bread

We do chicken right.

我们做鸡的权利。 (Google Translate, 2016.11.4.)

我们是烹鸡专家。 (百度翻译, 2016.11.4.)

NLP要解决的问题是从大量不确定性中寻找确定性结论，很多背景知识和常识性知识是隐含的，是在语义和概念层面上进行的表示、处理和变换。

内容提要

1. 引言
2. NLP方法概述
3. 深度学习方法应用
4. 讨论与结语

2. NLP方法概述

2.1 基本方法

- **理性主义方法**：1957～1980S
  - 词法分析，句法方法，语义分析
  - 词典、规则**—基于规则的方法**
- **经验主义方法**：～1950S，1980S～
  - 训练样本
  - 统计模型**—基于统计的方法**

2. NLP方法概述

- **以机器翻译为例**

  给定英语句子：

  There is a book on the desk.

  将其翻译成汉语。

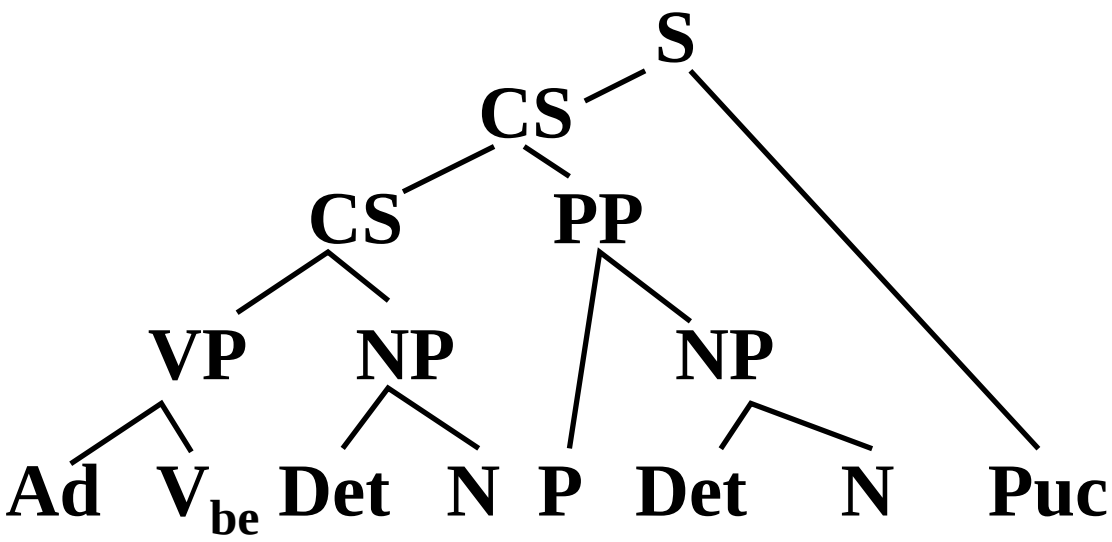
2. NLP方法概述

➢ 基于规则的方法

✧ 对英语句子进行词法分析

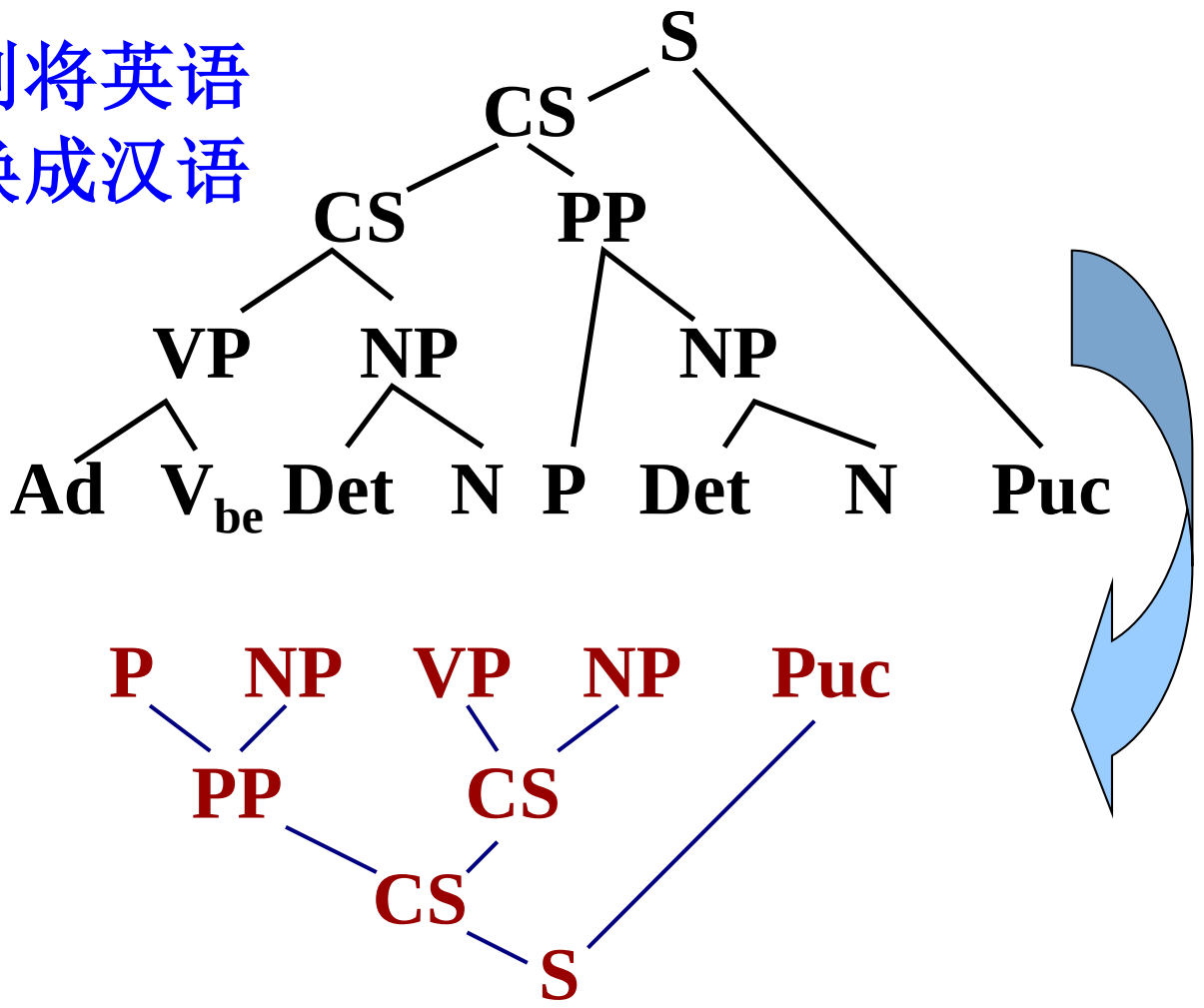
There/Ad is/V_{be} a/Det book/N on/P the/Det desk/N ./Puc

✧ 对英语句子进行句法结构分析



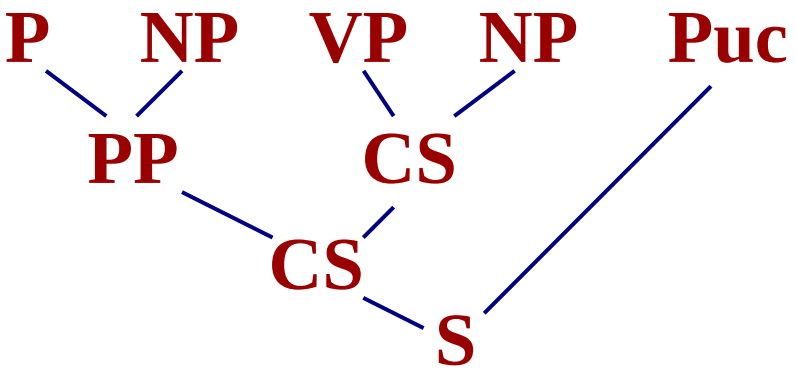
2. NLP方法概述

- 利用转换规则将英语句子结构转换成汉语句子结构



2. NLP方法概述

✧ 根据转换后的句子结构，利用词典和生成规则生成翻译的结果句子



#a, Det, 一

#book, N, 书; V, 预订

#desk, N, 桌子

#on, P, 在 X 上

#There be, V, 有

输出译文：

在桌子上有一本书。

基于规则的NLP方法的基本步骤：
词法分析（汉语分词） → 句法分析 → 语义分析（词义消歧等） → 语言生成

2. NLP方法概述

➢基于统计的方法

给定源语言句子：$E = e_1^m \equiv e_1 e_2 \cdots e_m$

将其翻译成目标语言句子：$C = c_1^l \equiv c_1 c_2 \cdots c_l$

根据贝叶斯公式：$P(C|E) = \dfrac{P(C) \times P(E|C)}{P(E)}$

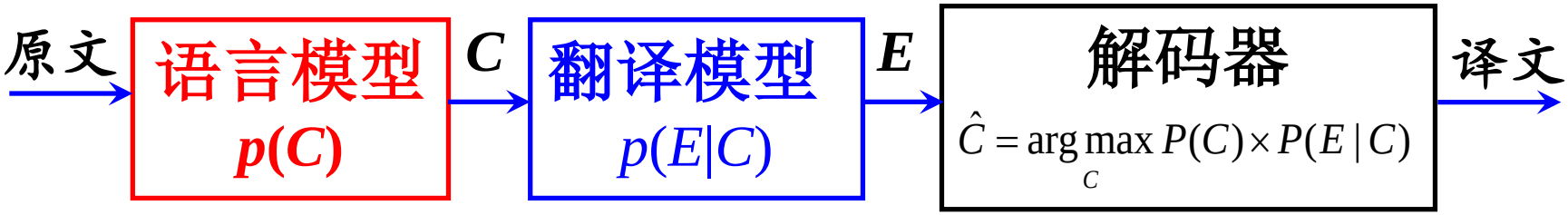
$$\hat{C} = \arg\max_C P(C) \times P(E|C)$$

语言模型
(Language model, LM)

翻译模型
(Translation model, TM)

2. NLP方法概述

构建解码器(decoder)，快速搜索最优翻译候选：



◆三个关键问题：

- 估计语言模型概率 $p(C)$；
- 估计翻译模型概率 $p(E|C)$；
- 快速有效地搜索候选译文C，使$p(C)\times p(E|C)$最大。

◆主要任务：

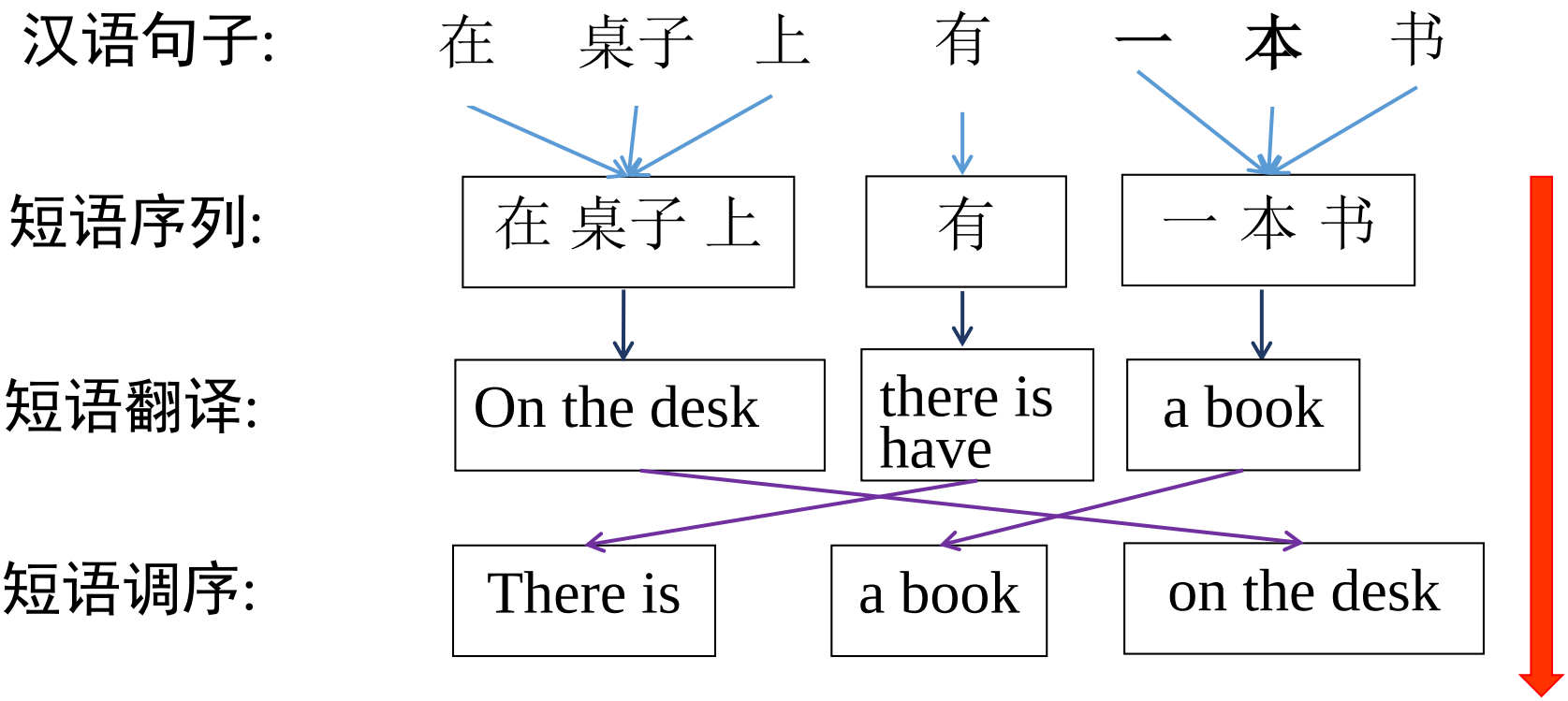
- 收集大规模双语句子对、目标语言句子
- 参数训练与模型优化

2. NLP方法概述

人类 共 有 二十三 对 染色体 。	humans have a total of 23 pairs of chromosomes .
澳洲 重新 开放 驻 马尼拉 大使馆	australia reopens embassy in manila
中国 大陆 手机 用户 成长 将 减缓	growth of phone users in mainland china to slow
… …	… …
外交 人员 搭乘 第五 架 飞机 返国	diplomatic staff will take the fifth plane home .
驻 南韩 美军 三千人 奉命 冻结 调防	us freezes transfer of 3,000 troops in south korea
姚明 感慨 NBA 的 偶像 来 得 太 快	yao ming feels nba stardom comes too fast

双语句对

2. NLP方法概述


汉语句子:
在 桌子 上 有 一 本 书
短语序列:
在 桌子 上
有
一 本 书
短语翻译:
On the desk
there is
have
a book
短语调序:
There is
a book
on the desk
英语译文:
There is a book on the desk.

2. NLP方法概述

2.2 常用的统计模型和开源工具

- 感知机(perceptron)：二类分类
- *k*-近邻法(*k*-nearest neighbor, *k*-NN)：多类分类问题
- 朴素贝叶斯法(naïve Bayes)：多类分类问题
- 决策树(decision tree)：多类分类问题
- 最大熵(maximum entropy)：多类分类问题
- 支持向量机(support vector machine, SVM)：二类分类
- 条件随机场(conditional random field, CRF)：序列标注
- 隐马尔可夫模型(hidden Markov model, HMM)：标注

2. NLP方法概述

开源工具：

- **条件随机场：**
  - ✧**CRF++ （C++版）：**
    http://crfpp.googlecode.com/svn/trunk/doc/index.html
  - ✧**CRFSuite （C语言版）：**
    http://www.chokkan.org/software/crfsuite/
  - ✧**MALLET (Java版，通用的NLP工具包，包括分类、序列标注等机器学习算法)：** http://mallet.cs.umass.edu/
  - ✧**NLTK (Python版，通用的NLP工具包，很多工具是从MALLET中包装转成的Python接口)：** http://nltk.org/

2. NLP方法概述

- 贝叶斯分类器：http://www.openpr.org.cn
- 支持向量机(LibSVM)：
  http://www.csie.ntu.edu.tw/~cjlin/libsvm
- 隐马尔可夫模型：http://htk.eng.cam.ac.uk/
- 最大熵：
  - ✧ **OpenNLP**：http://incubator.apache.org/opennlp/
  - ✧ **Malouf**：http://tadm.sourceforge.net/
  - ✧ **Tsujii**：http://www-tsujii.is.s.u-tokyo.ac.jp/~tsuruoka/maxent/
  - ✧ 张乐：http://homepages.inf.ed.ac.uk/lzhang10/maxent.html
  - ✧ 林德康：http://webdocs.cs.ualberta.ca/~lindek/downloads.htm

2. NLP方法概述

2.3 应用举例

❶由字构词的汉语自动分词

(Character-based Chinese word segmentation)

Nianwen Xue(薛念文) and **S. Converse**, 2002, The 1st SIGHAN Workshop.

基本思想：将分词过程看作是字的分类问题。该方法认为，每个字在构造一个特定的词语时都占据着一个确定的构词位置(即词位)。假定每个字只有4个词位：词首(B)、词中(M)、词尾(E)和单独成词(S)，那么，每个字归属一特定的词位。

2. NLP方法概述

(1) 上海/ 计划/ 到/ 本/ 世纪/ 末/ 实现/ 人均/ 国内/ 生产/ 总值/ 五千美元/ 。/

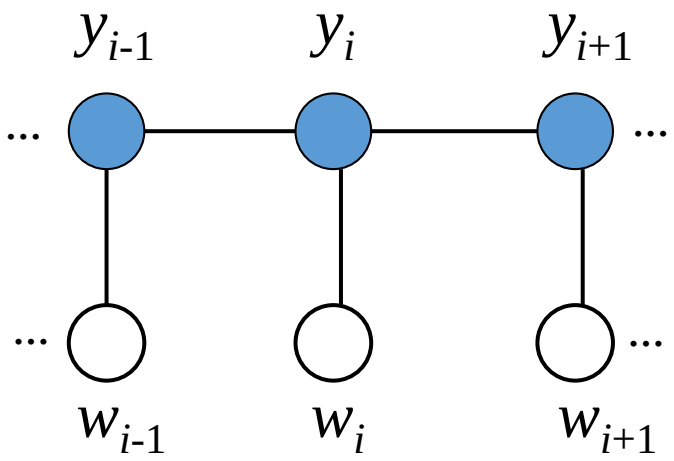
(2) 上/B 海/E 计/B 划/E 到/S 本/S 世/B 纪/E 末/S 实/B 现/E 人/B 均/E 国/B 内/E 生/B 产/E 总/B 值/E 五/B 千/M 美/M 元/E 。/S

在字标注过程中，对所有的字根据预定义的特征进行词位特征学习，获得一个概率模型，然后在待切分字串上，根据字与字之间的结合紧密程度，得到一个词位的分类结果，最后根据词位定义直接获得最终的分词结果。

- **工具：**
  - 支持向量机 (SVM)
  - 条件随机场 (CRF)

2. NLP方法概述

基于条件随机场(CRF)的识别方法：序列标注



三个问题：

①特征选取

②参数训练

③解码

$$p(Y \mid X, \boldsymbol{\lambda}) = \frac{1}{Z(X)} \exp(\lambda_j \cdot F_j(Y, X))$$

$Z(X)$为归一化因：$Z(X) = \sum_Y \exp(\lambda_j \cdot F_j(Y, X))$

特征函数：$F_j(Y, X) = \sum_{i=1}^{n} f_j(y_{i-1}, y_i, X, i)$

2. NLP方法概述

上/B 海/E 计/B 划/E 到 本 世 纪 ……

↑ B, E, M, S ?

- 当前字的前后 n 个字 (如 $n=\pm 2$)
- 当前字左边字的标记
- 当前字在词中的位置

……

Urheen 汉语自动分词系统：

http://www.nlpr.ia.ac.cn/cip/software.htm

2. NLP方法概述

❷**词义消歧** (word sense disambiguation, WSD)

(1)他**打**鼓很在行。
(2)他会**打**家具。
(3)他把碗**打**碎了。
(4)他在学校**打**架了。
(5)他很会与人**打**交道。
(6)他用土**打**了一堵墙。
(7)用面**打**浆糊贴对联。
(8)他**打**铺盖卷儿走人了。
(9) 她会用毛线**打**毛衣。
(10)他用尺子**打**个格。
(11)他**打**开了箱子盖。
(12)她**打**着伞走了。
(13)他**打**来了电话。
(14)他**打**了两瓶水。
(15)他想**打**车票回家。
(16)他以**打**鱼为生。

2. NLP方法概述

每个词表达不同的含意时其上下文（语境）往往不同，不同的词义对应不同的上下文，因此，如果能够将多义词的上下文区别开，其词义自然就明确了。

他/P 很/D 会/V 与/C 人/N **打/V** 交道/N 。/PU

-2 -1 **0** +1 +2

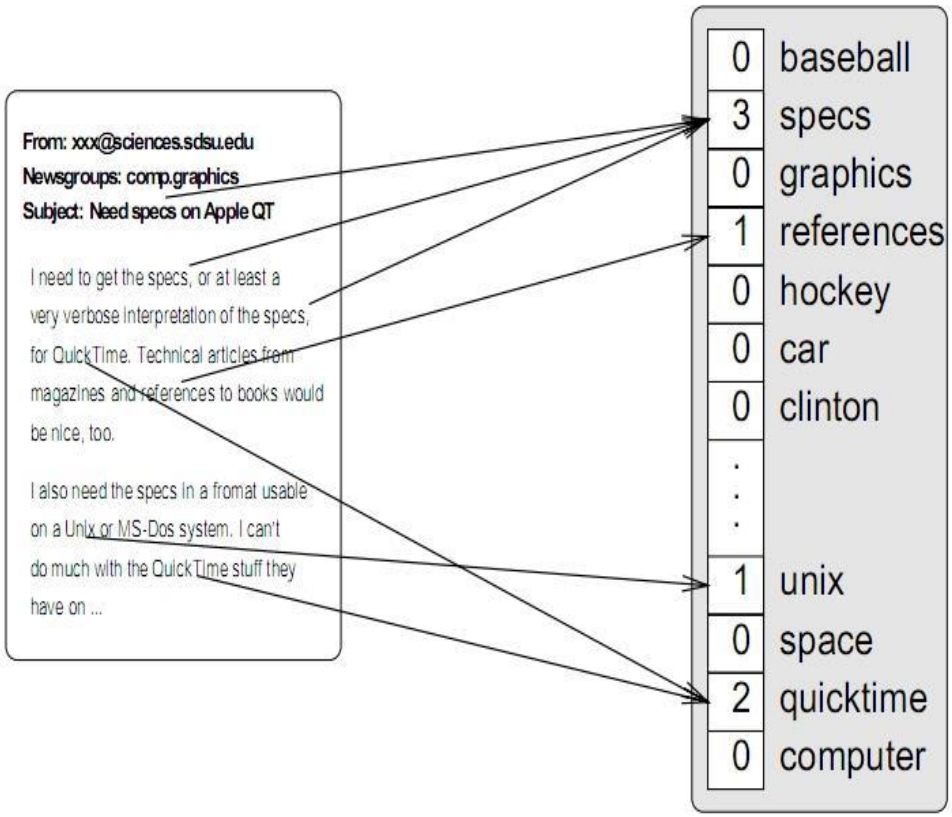
基本的上下文信息：词、词性、位置

- **上下文表示：**词袋模型(bag of word, BOW)或称向量空间模型(vector space mode, VSM)
- **分类器：**贝叶斯、条件随机场(CRF)、最大熵 ……

2. NLP方法概述

❸文本分类 (text classification, TC)

- 文本表示：VSM



- 布尔变量(是否出现)

$$\omega_{ki} = \begin{cases} 1, & \text{if } t_i \text{ exits in } D_k \\ 0, & \text{Otherwise} \end{cases}$$

- 词频(term frequency, TF)

$$\omega_{ki} = tf_{ki}$$

- 倒排文档频率：$\omega_i = \log \frac{N}{df_i}$

(inverse document frequency, IDF)

- TF-IDF: $\omega_i = tf_{ki} \times \log \frac{N}{df_i}$

2. NLP方法概述

- 训练样本（带类别标签的文本）

计算机

IBM、微软、Google等一批国际著名计算机公司昨天下午的中国计算机大会上展示了他们最新严重的 ……

联想公司笔记本电脑……

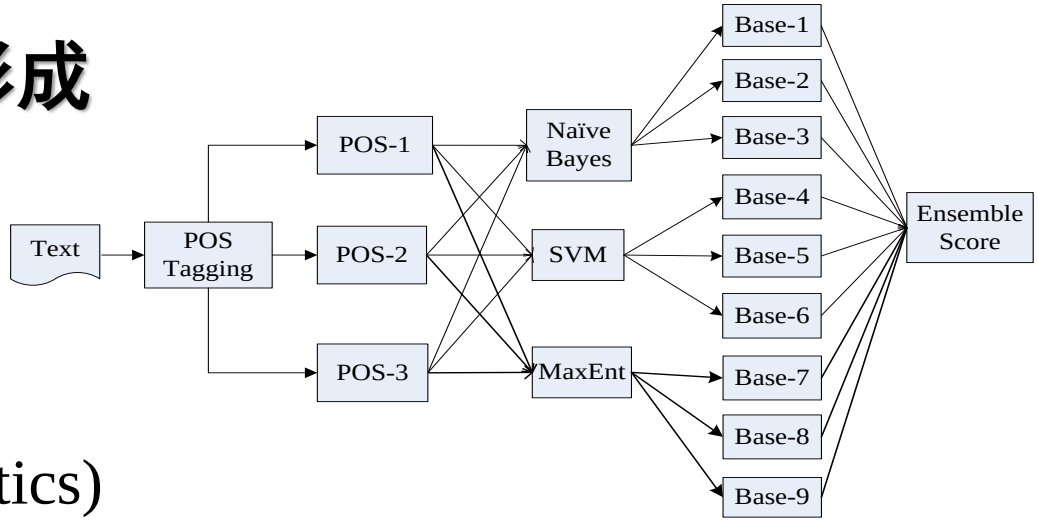
体育

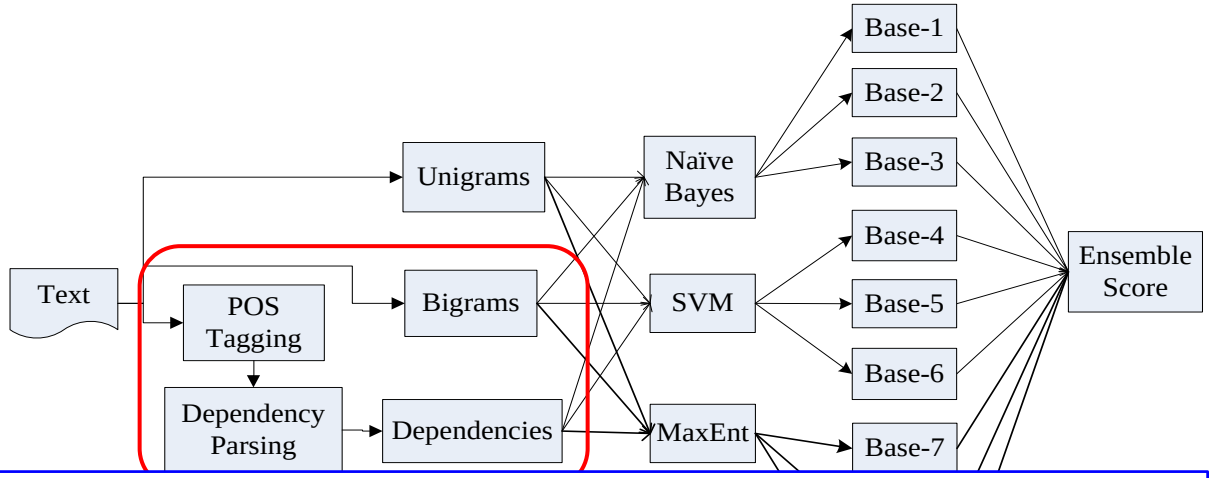
国际田径锦标赛将于8月16日在北京奥林匹克体育中心赛….

第五届东亚运动会中国军团奖牌总数创新高，男女排球双双夺冠 …

2. NLP方法概述

- **词汇抽取、词汇表形成**
- **特征选择**
  - 文档频率
  - 互信息信息增益
  - CHI (Chi-square statistics)
- **分类器设计**
  - SVM
  - 贝叶斯 ……
  - 组合分类器



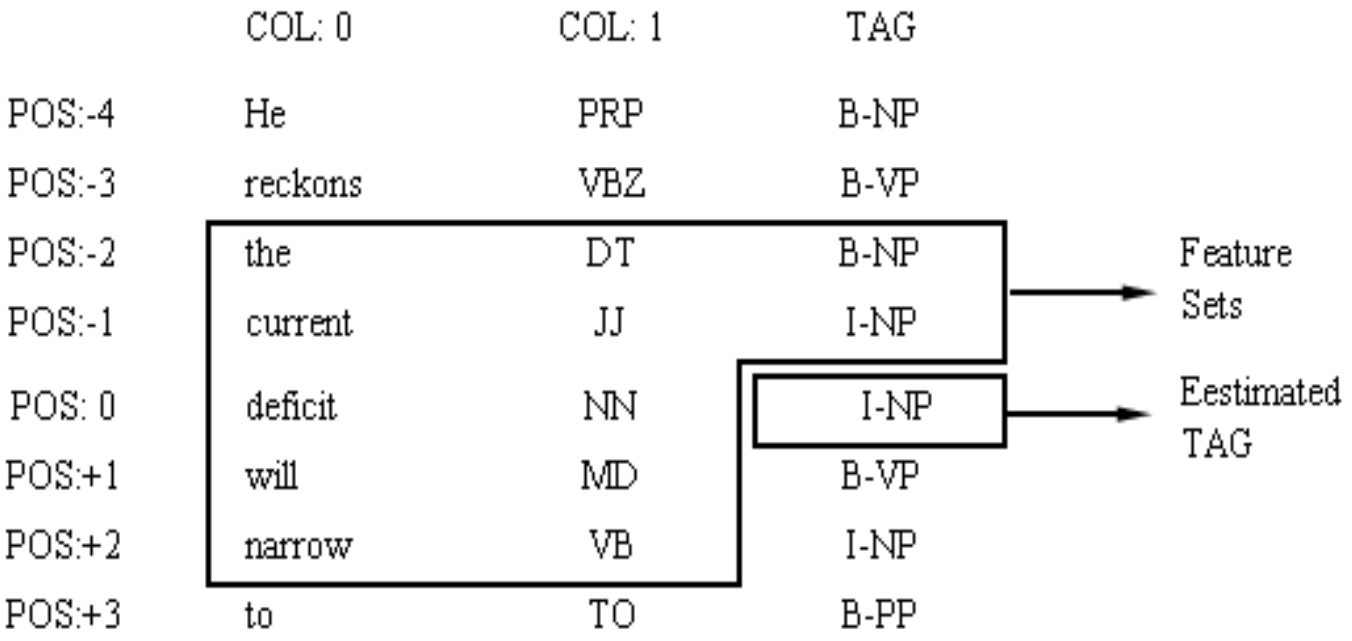


R. Xia, C. Zong, and S. Li. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181(2011): 1138–1152

2. NLP方法概述

- **其他任务：**

语块识别 → 命名实体识别 → 词性标注 → 指代消解 → 语义角色标注 → 依存句法分析 → 篇章单元识别 → 篇章关系识别 → 情感分类 …



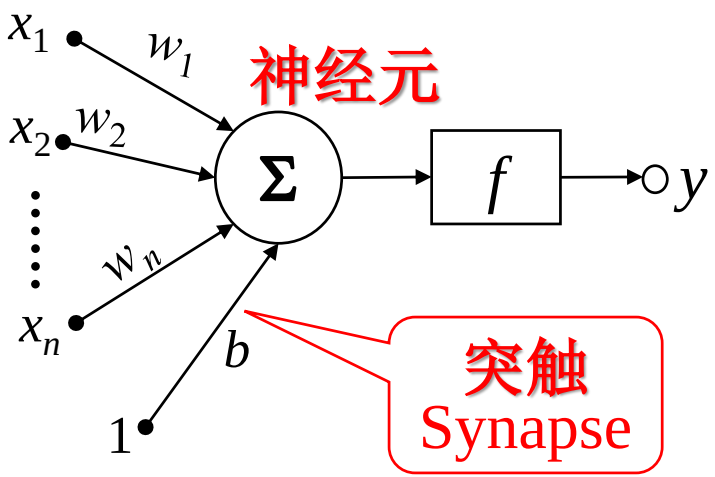
特征表示 + 分类器

内容提要

1. 引言
2. NLP方法概述
3. 深度学习方法应用
4. 讨论与结语

3. 深度学习方法应用

3.1 概要



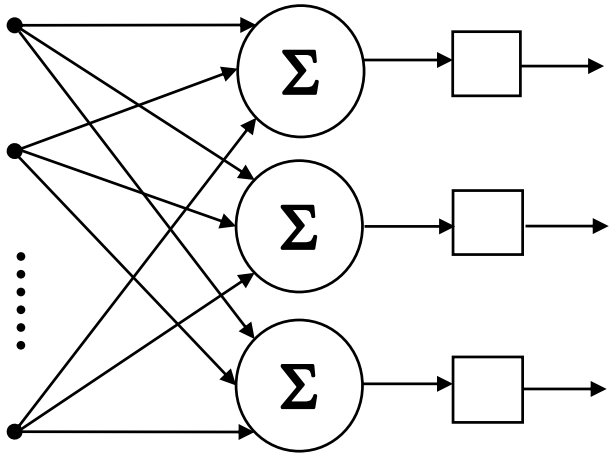
- 神经元数学描述
  - $x_1 \sim x_n$ 为输入向量的各分量
  - $w_1 \sim w_n$ 为权值
  - b 为偏置
  - f 为传递函数，通常为非线性函数
  - y 为输出
  - 数学表示：$y = f(\vec{W} \bullet \vec{X}' + b)$

$\vec{W}$ 为权值向量；$\vec{X}$ 为输入向量，$\vec{X}'$ 为 $\vec{X}$ 的转置。

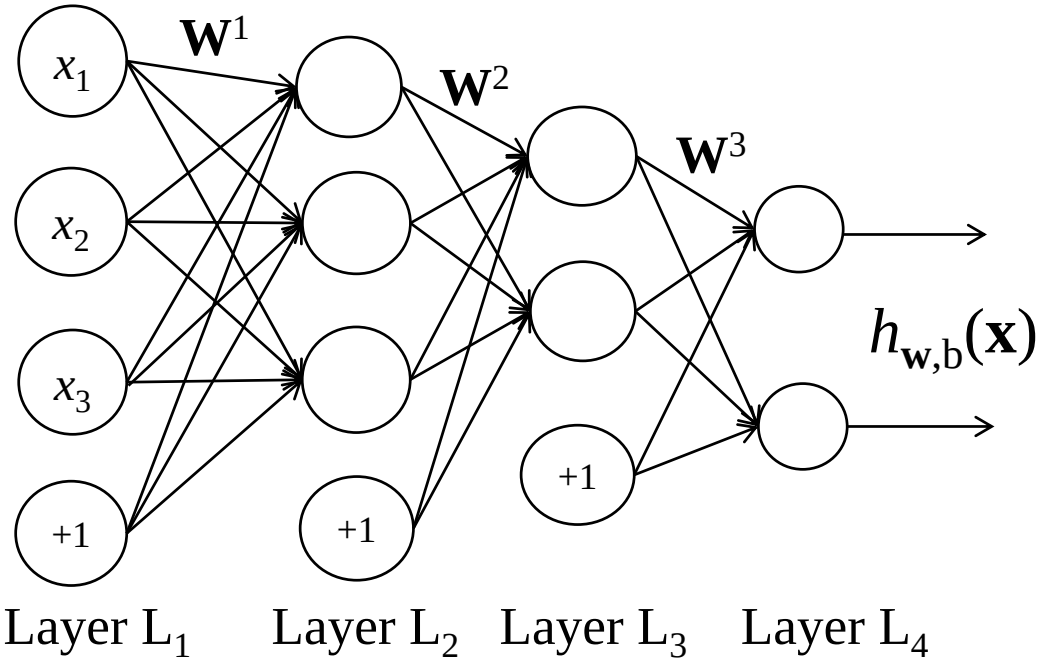
3. 深度学习方法应用

- **神经网络(1982，Hopfield神经网络模型；1984,连续时间的Hopfield神经网络模型)**

  - 有限个神经元
  - 所有神经元的输入都是同一个向量 $\vec{X}$
  - 网络输出也是一个向量，维数等于神经元的个数

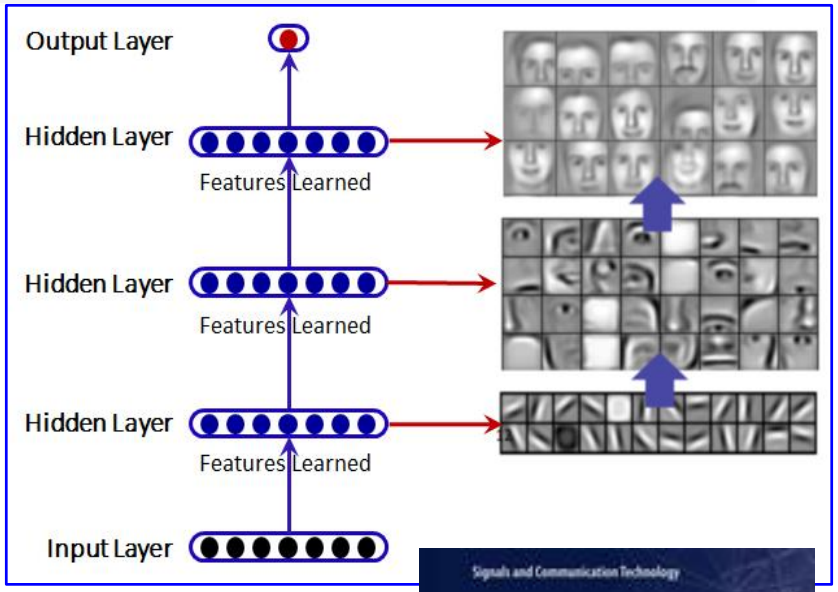


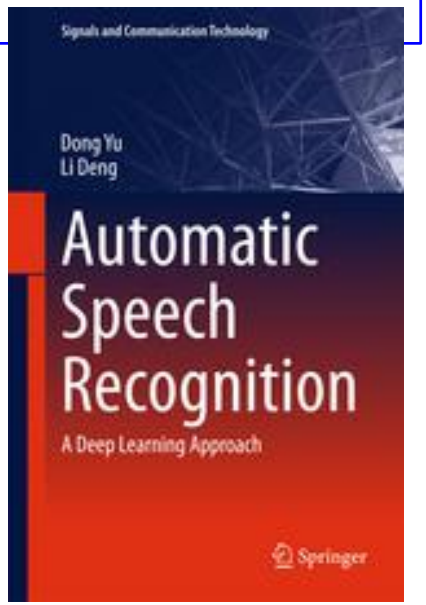
单层网络



3. 深度学习方法应用

- 基于深层(前向多层)神经网络的学习通过校正训练样本，对各个层的权重进行调整(learning)
- 2006 年 G. E. Hinton (辛顿)等人使用受限玻尔兹曼机(restricted Bolzman machine)进行逐层无监督训练方法，率先在图像识别上获得了突破
- 2009年DNN在语音识别中获得成功应用





3. 深度学习方法应用

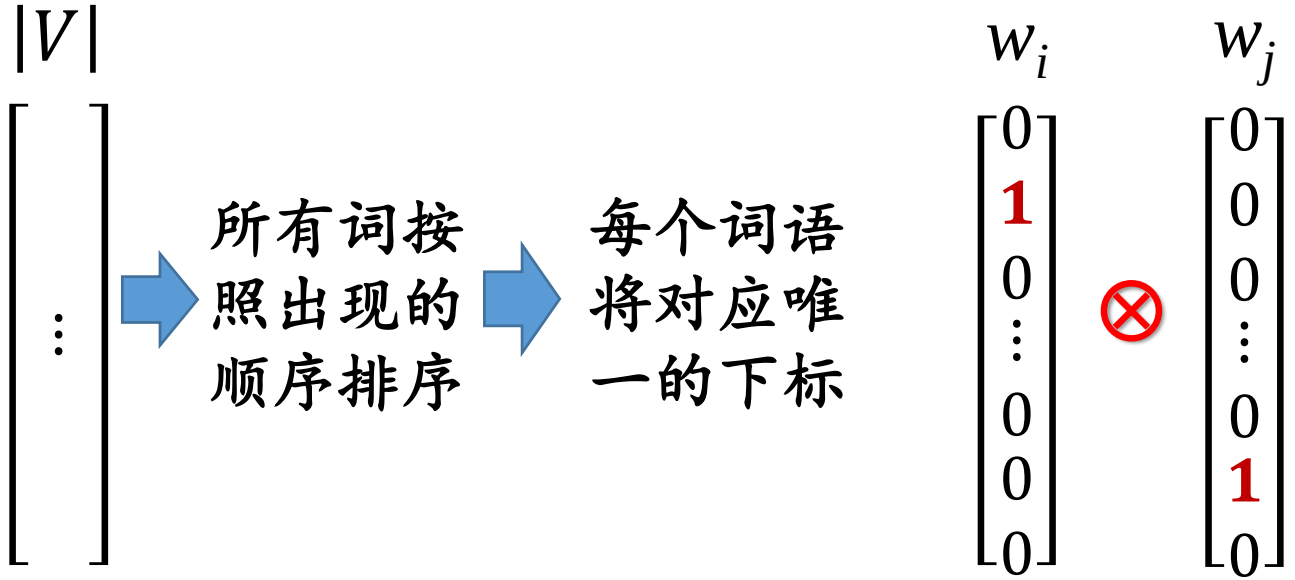
NLP中常用的几种网络：

- 前馈神经网络
- 循环神经网络
- （递归）自编码器
- 递归神经网络
- 卷积神经网络

3. 深度学习方法应用

3.2 词向量表示

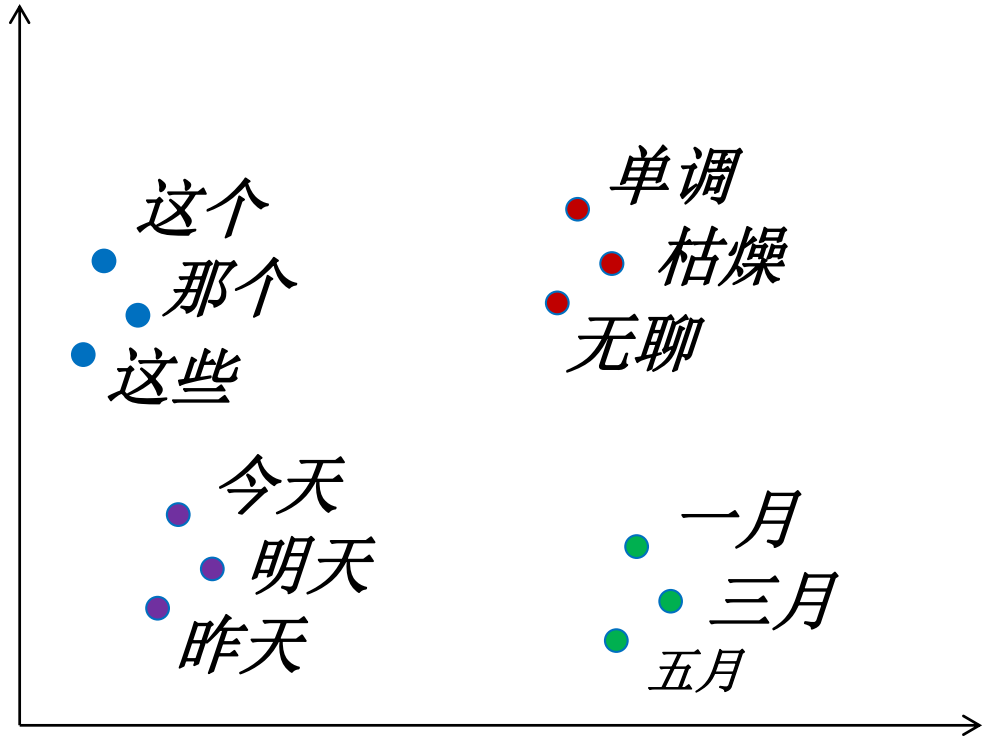
- **one-hot词义表示法：**



任意两个词之间的相似度都为0！

3. 深度学习方法应用

构建词语-实数的向量表示



枯燥

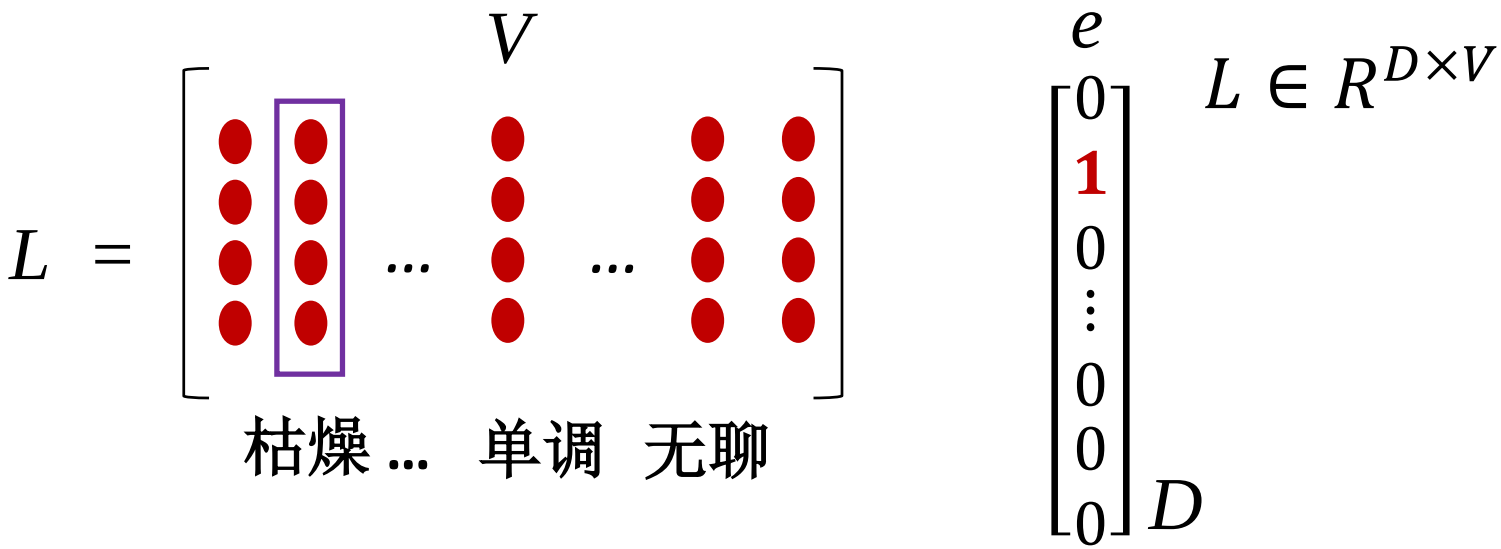
$$\begin{bmatrix} 0.24 \\ 0.15 \\ 0.42 \\ 0.51 \\ 0.21 \end{bmatrix}$$

无聊

$$\begin{bmatrix} 0.25 \\ 0.12 \\ 0.39 \\ 0.46 \\ 0.26 \end{bmatrix}$$

低维、稠密的连续实数空间

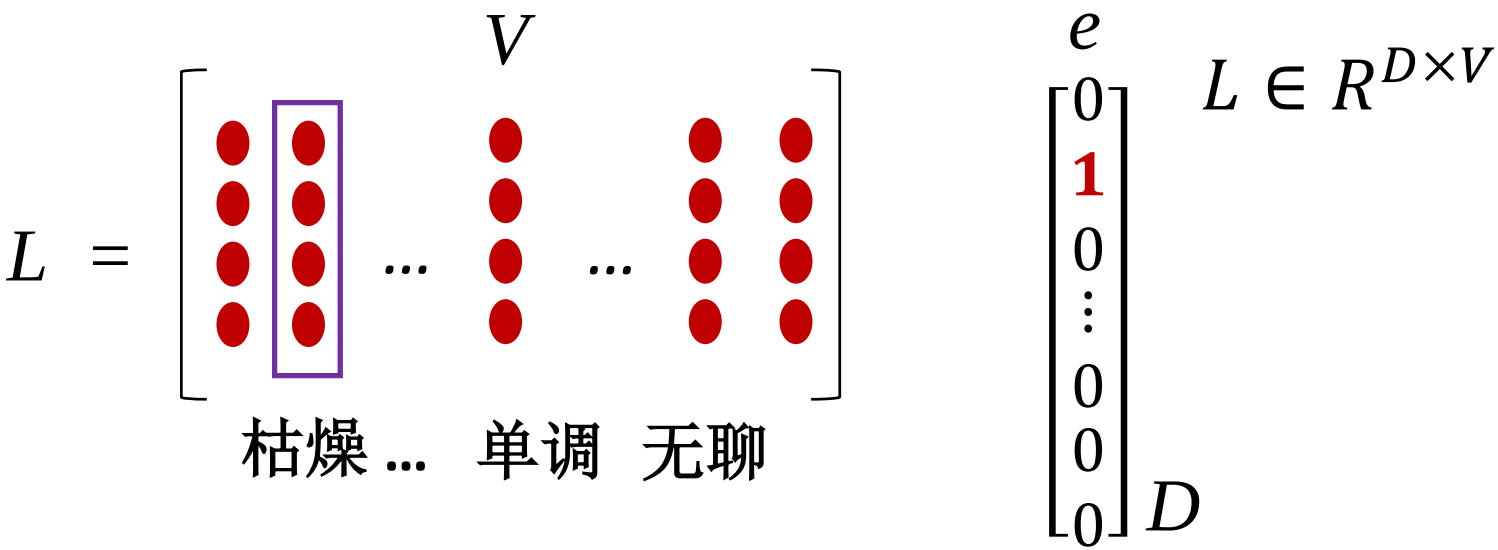
3. 深度学习方法应用



- **通常称为 look-up table**

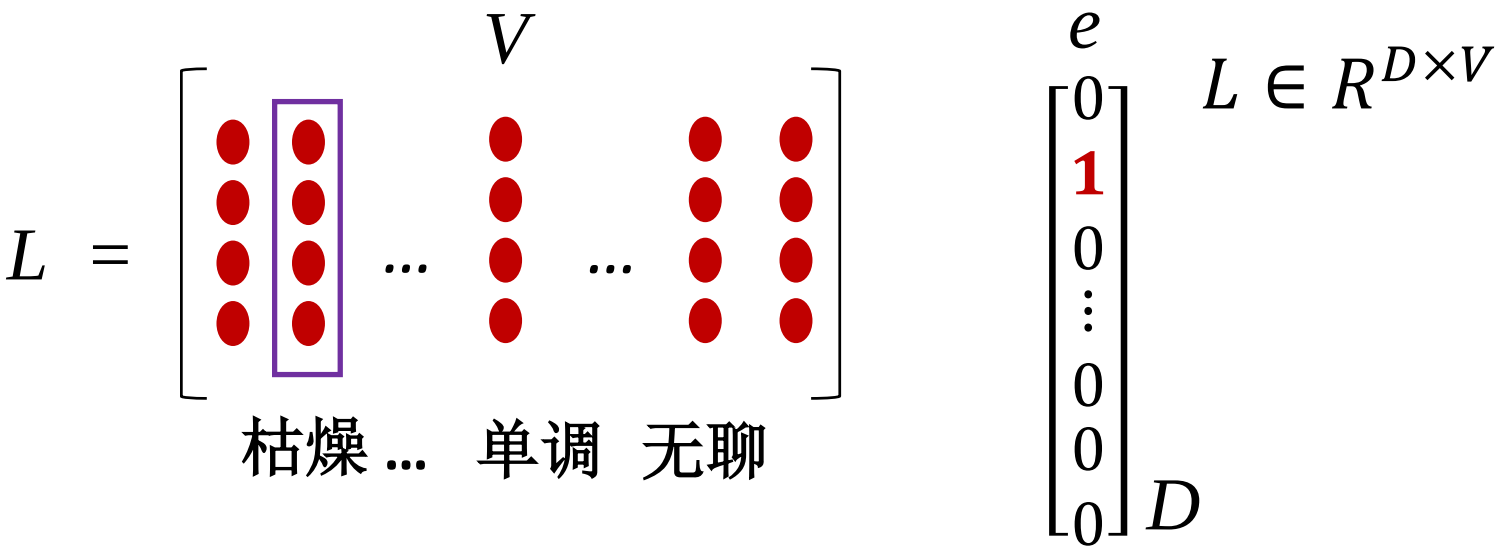
  对L右乘一个词的one-hot表示 e，得到该词的低维、稠密的实数向量表达：$x = Le$

3. 深度学习方法应用



- **词表规模V 和词向量维度D 的确定：**
  - **V 的确定：①训练数据中所有词； ②出现频率高于某个阈值的所有词； ③前V个频率最高的词。**
  - **D 的确定：超参数，人工设定，一般从几十到几百。**

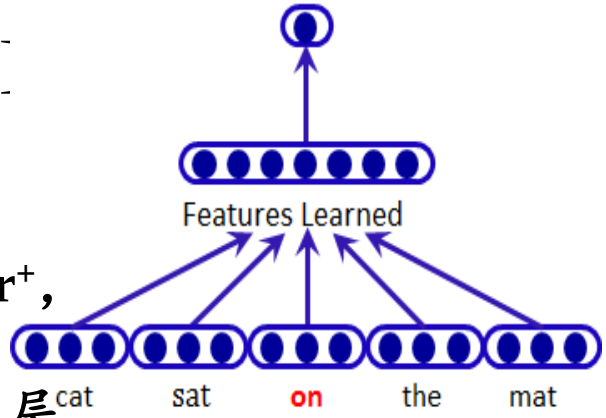
3. 深度学习方法应用



- 如何学习L?
  - 通常先随机初始化，然后通过目标函数优化词的向量表达（**e.g.** 最大化语言模型似然度）

3. 深度学习方法应用

①替换中间词的方法[Collobert et al., 2011]

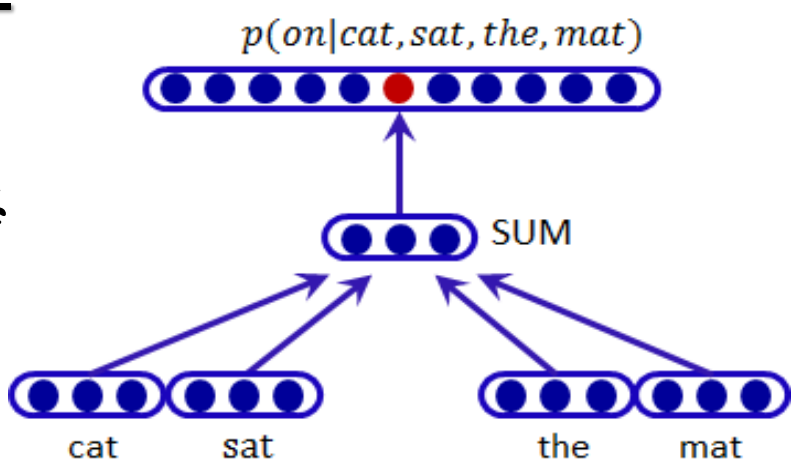


- 将词表中的每个词随机初始化一个向量
- 用大规模语料训练优化该向量
  - 从训练数据中随机选一个窗口大小为n的片段$\mathbf{phr}^+$，作为正例
  - 将$\mathbf{phr}^+$对应的词向量拼接，作为神经网络的输入层
  - 经过一个隐含层后得到得分f^+，表示该片段是否为一个正常的自然语言片段
  - 将窗口中间的词随机替换成词表中的另一个词，得到一个负例片段$\mathbf{phr}^-$及负例的打分f^-
  - 用排序合页损失（ranking hinge loss）作为损失函数，使正例的得分f^+至少比负例的得分f^-大1
  - 对该损失函数求导得到梯度，使用反向传播的方式学习神经网络各层的参数，同时更新正负例样本中的词向量，从而将语义相似的词映射到向量空间中相近的位置。

[Collobert et al., 2011]Collobert, Roman et al. 2011. Natural Language Processing (almost) from Scratch. J. Mach. Learn. Res., pp. 2493-2537

3. 深度学习方法应用

②用周围词预测中间词的方法 – 连续词包模型(CBOW)



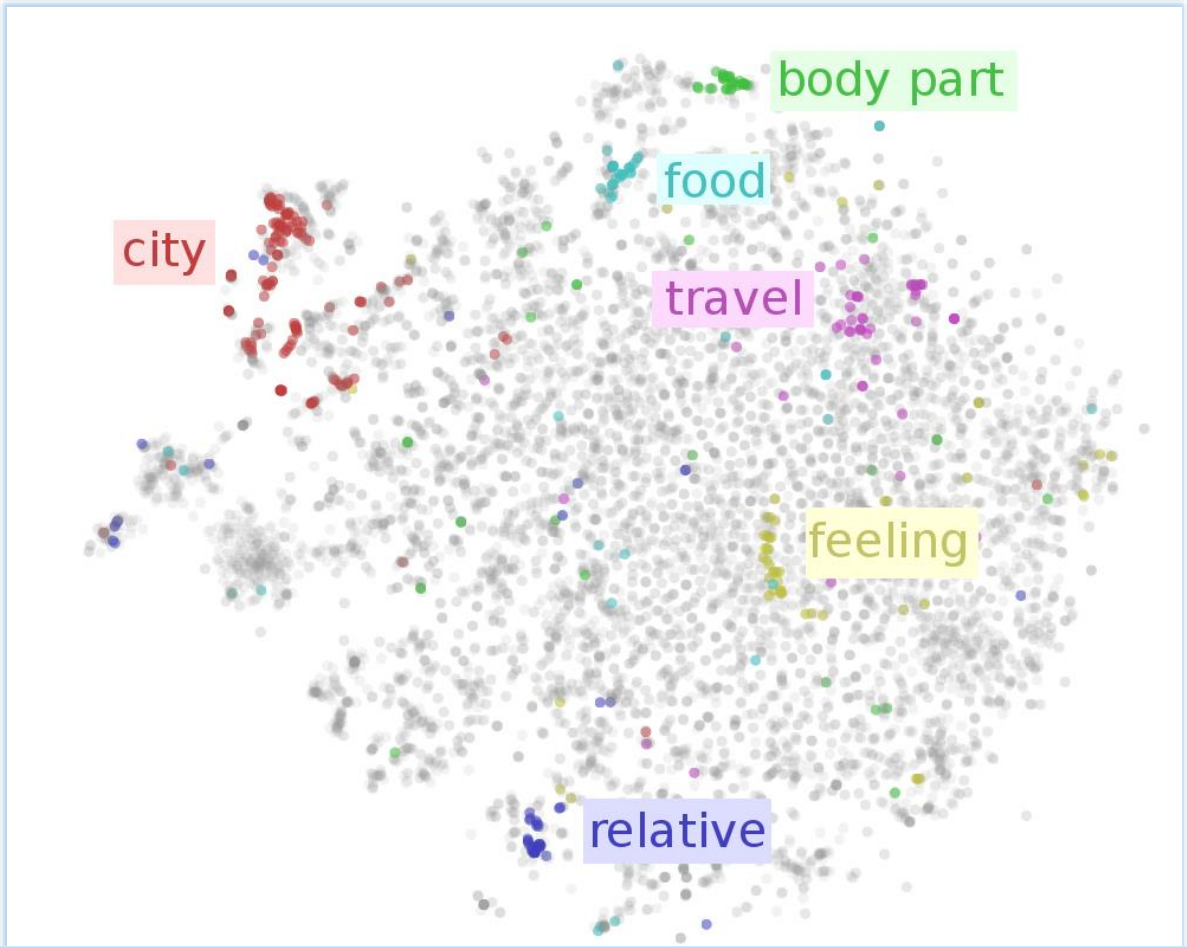
- 将相邻的词向量直接相加得到隐层，用隐层预测中间词的概率
- 连续 skip-gram 模型：通过中间词预测周围词的概率

[Mikolov et al., 2013]Mikolov, Tomas, K. Chen et al . 2013. Efficient estimation of word representation in vector space. arXiv preprint arXiv: 1301.3781, 2013

- **skip-gram with negative-sampling (SGNS) *vs*. SVD**

[Levy and Goldberg, 2014]Levy, Omer and Yoav Goldberg. 2014. Neural Word Embedding as Implicit Matrix Factorization. *Proc. NIPS*

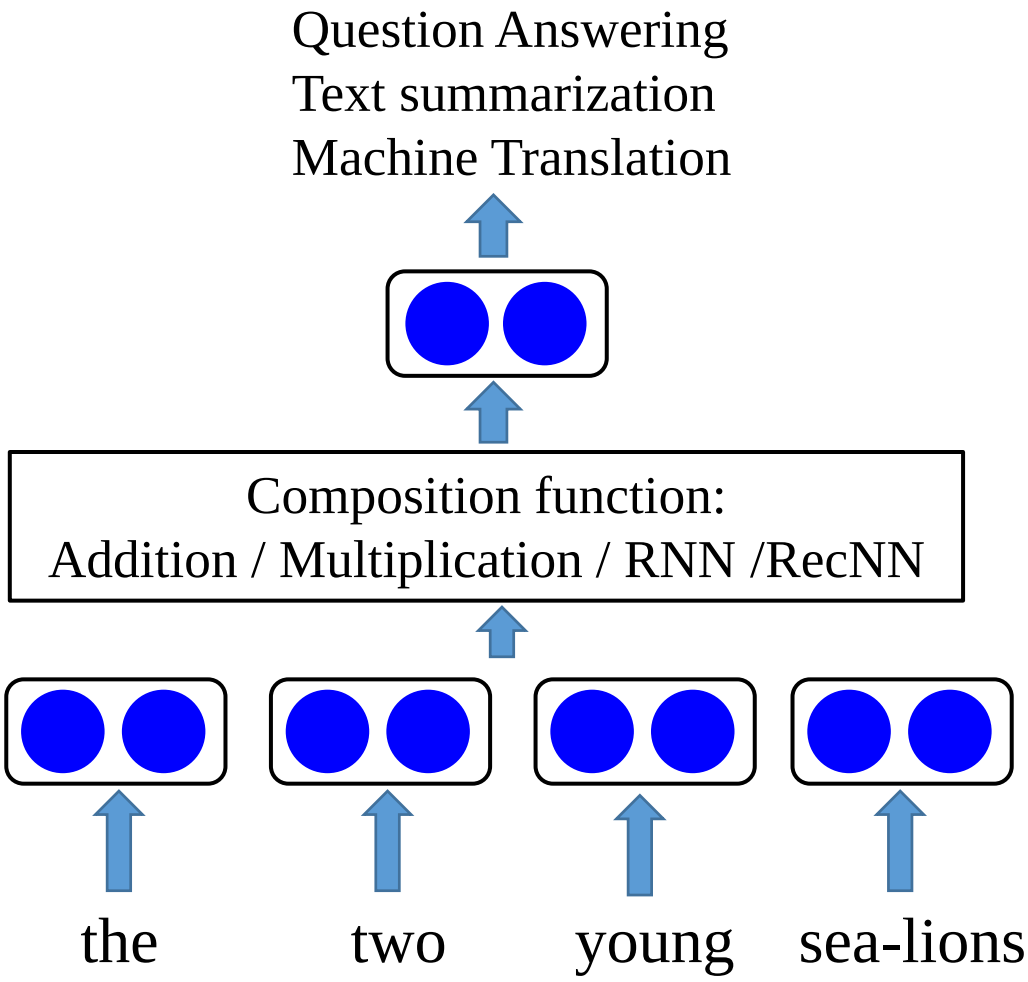
3. 深度学习方法应用



在低维、稠密的实数向量空间中，相似的词聚集在一起，在相同的历史上下文中具有相似的概率分布！

3. 深度学习方法应用

3.3 句子语义表示



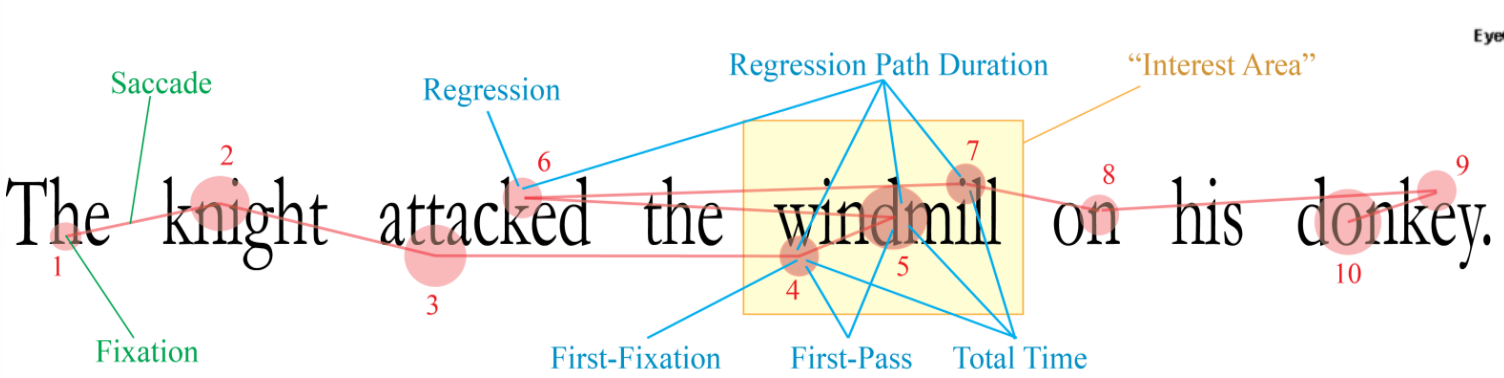
应用

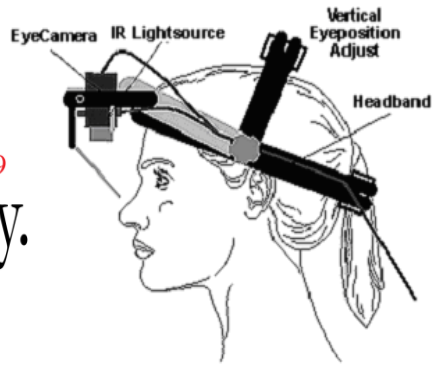
句子向量表示

词向量表示

文字序列

3. 深度学习方法应用





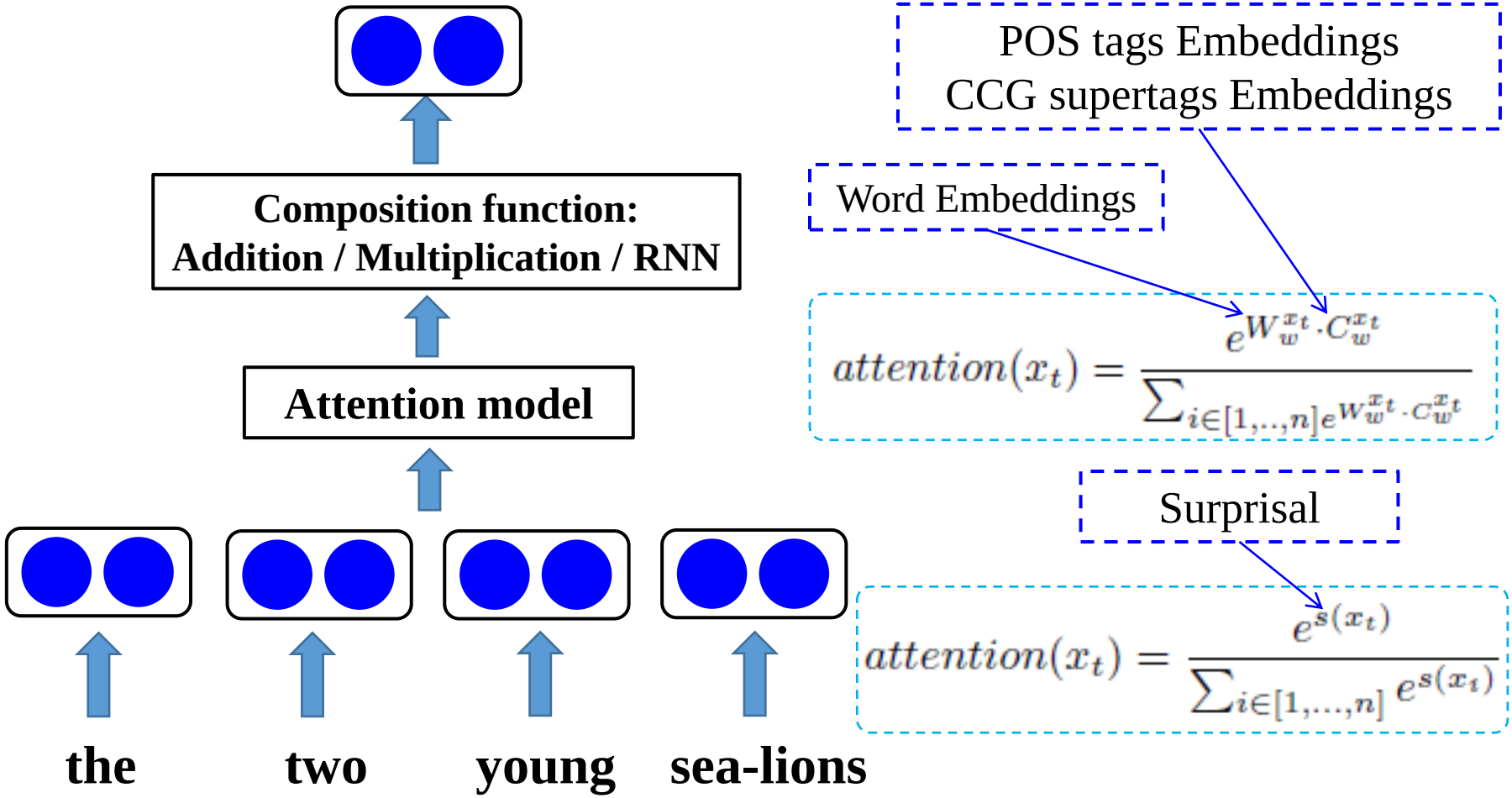
$$Surprisal(x_t) = -log(P(x_t|x_1, ..., x_{t-1}))$$

The Dundee human reading time corpus

Reading time/word	the	two	young	Sea-lions	took
RTfpass	27.2	138.7	155.5	314.8	169.3
RTgopast	27.2	138.7	178.4	426.1	180.9
RTrb	27.2	138.7	155.5	339.2	169.3

Surprisal (惊异度)
POS tag (词性)
CCG supertag (范畴)
Word length
Word frequency
……

3. 深度学习方法应用


POS tags Embeddings
CCG supertags Embeddings
Word Embeddings
$attention(x_t) = \frac{e^{W_w^{x_t} \cdot C_w^{x_t}}}{\sum_{i \in [1,..,n]} e^{W_w^{x_t} \cdot C_w^{x_t}}}$
Surprisal
$attention(x_t) = \frac{e^{s(x_t)}}{\sum_{i \in [1,...,n]} e^{s(x_t)}}$
Composition function:
Addition / Multiplication / RNN
Attention model
the
two
young
sea-lions

3. 深度学习方法应用

实验对比

Pearson rank correlation on SemEval textual similarity datasets. The bold scores in each row are the best result in the PP model column and the SCBOW model column, respectively. `Base` denotes baseline model.

	Additive	Semeval-Best	PP		SCBOW			
			Base	ATT-SUR	Base	ATT-SUR	ATT-POS	ATT-CCG
MSRpar	0.423	0.734	0.476	**0.487**	**0.429**	0.425	0.414	0.419
MSRvid	0.505	0.880	0.774	**0.801**	0.620	0.666	0.702	**0.734**
OnWN	0.634	0.727	0.714	**0.724**	0.687	0.680	0.695	**0.696**
SMTeurop	0.527	0.567	0.481	**0.504**	0.537	0.549	0.538	**0.552**
SMTnews	0.476	0.609	0.652	**0.659**	0.523	0.524	0.541	**0.557**
2012 Average	0.513	0.703	0.619	**0.635**	0.559	0.569	0.578	**0.592**
FNWN	0.286	0.582	0.476	**0.503**	0.378	0.370	0.383	**0.392**
OnWN	0.518	0.843	0.738	**0.753**	0.584	**0.627**	0.609	0.583
headlines	0.649	0.784	0.733	**0.748**	0.693	0.705	0.704	**0.711**
2013 Average	0.484	0.736	0.649	**0.668**	0.552	**0.567**	0.565	0.562
OnWN	0.612	0.875	0.812	**0.820**	0.686	**0.713**	0.705	0.693
deft-forum	0.360	0.531	0.540	**0.553**	0.400	0.411	0.406	**0.421**
deft-news	0.706	0.785	0.739	**0.746**	0.724	**0.735**	0.726	0.733
headlines	0.647	0.784	0.707	**0.719**	0.652	0.655	0.666	**0.668**
images	0.583	0.837	0.805	**0.808**	0.660	0.667	0.733	**0.761**
tweets	0.648	0.792	0.769	**0.776**	0.708	0.689	**0.754**	0.723
2014 Average	0.593	0.767	0.729	**0.737**	0.638	0.645	0.665	**0.666**
ans-forums	0.356	0.739	0.691	**0.694**	0.476	**0.514**	0.491	0.512
ans-students	0.677	0.788	0.781	**0.782**	0.723	0.722	0.715	**0.732**
belief	0.570	0.772	0.773	**0.783**	0.584	0.591	0.600	**0.603**
headlines	0.672	0.842	0.764	**0.773**	0.713	0.723	0.720	**0.727**
images	0.672	0.871	0.837	**0.841**	0.740	0.746	0.773	**0.787**
2015 Average	0.487	0.802	0.769	**0.775**	0.647	0.659	0.660	**0.672**
answer	0.339	0.692	0.670	**0.673**	**0.398**	0.396	0.379	0.391
deadlines	0.642	0.827	0.699	**0.710**	0.683	0.700	**0.705**	0.701
plagiarism	0.615	0.841	0.802	**0.819**	0.708	0.711	**0.736**	**0.736**
postediting	0.747	0.867	0.828	**0.831**	**0.708**	0.703	0.700	0.706
question	0.487	0.747	0.535	**0.559**	0.580	0.627	0.623	**0.652**
2016 Average	0.566	0.795	0.707	**0.718**	0.615	0.627	0.629	**0.637**

3. 深度学习方法应用

- 注意力模型是否与人阅读是的眼动数据一致呢？

Pearson/Spearman rank correlation between attention by our models and human reading time. The asterisk means significantly for p-values < 0.0001.

	RTfpass	RTgopast	RTrb
ATT-SUR	0.399*/0.228*	0.381*/0.199*	0.401*/0.234*
ATT-POS	0.385*/0.307*	0.370*/0.268*	0.386*/0.302*
ATT-CCG	0.430*/0.347*	0.412*/0.306*	0.431*/0.342*

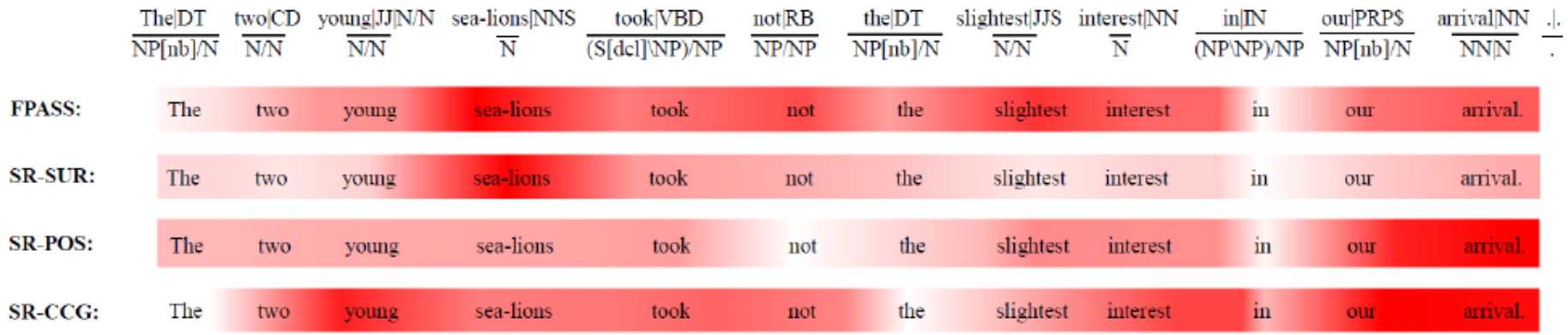
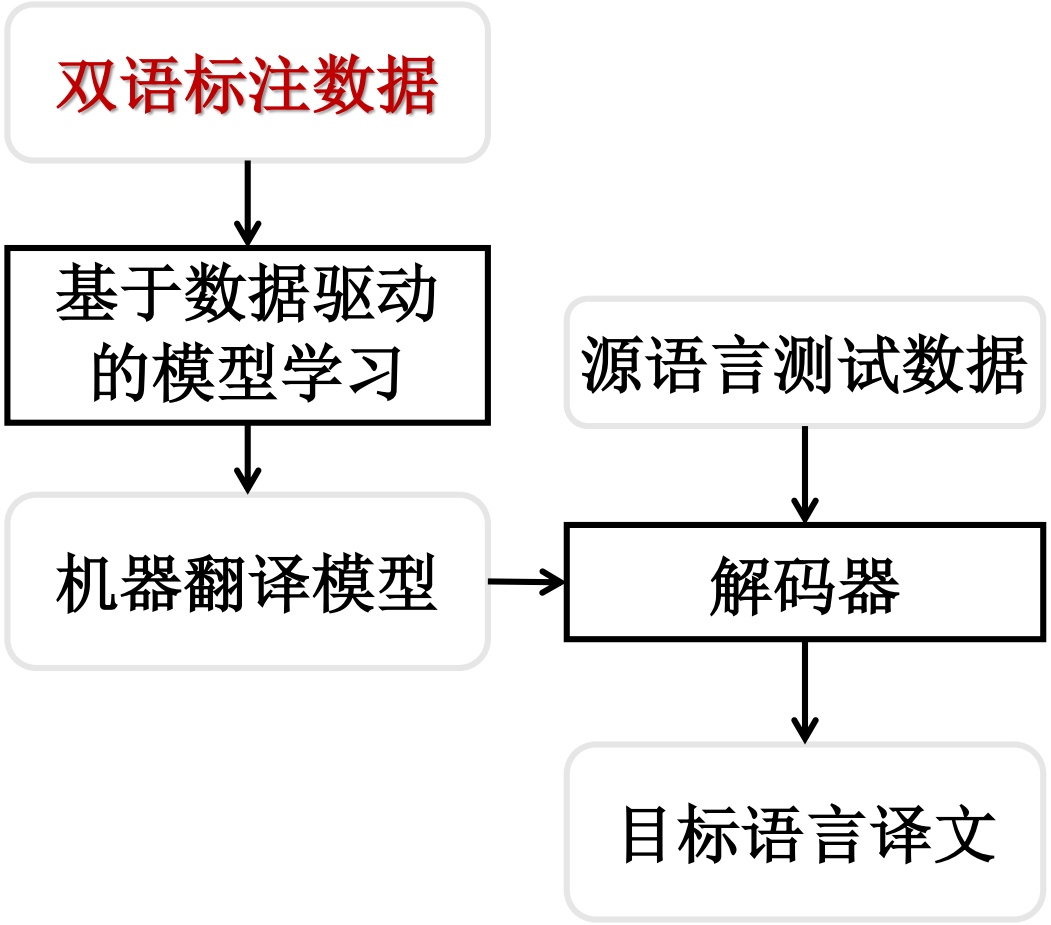


Figure 3: An example sentence from the Dundee corpus with the corresponding POS tag and CCG supertag, together with heatmaps of human first-pass reading time and attention calculated by our models with surprisal, POS and CCG, respectively.

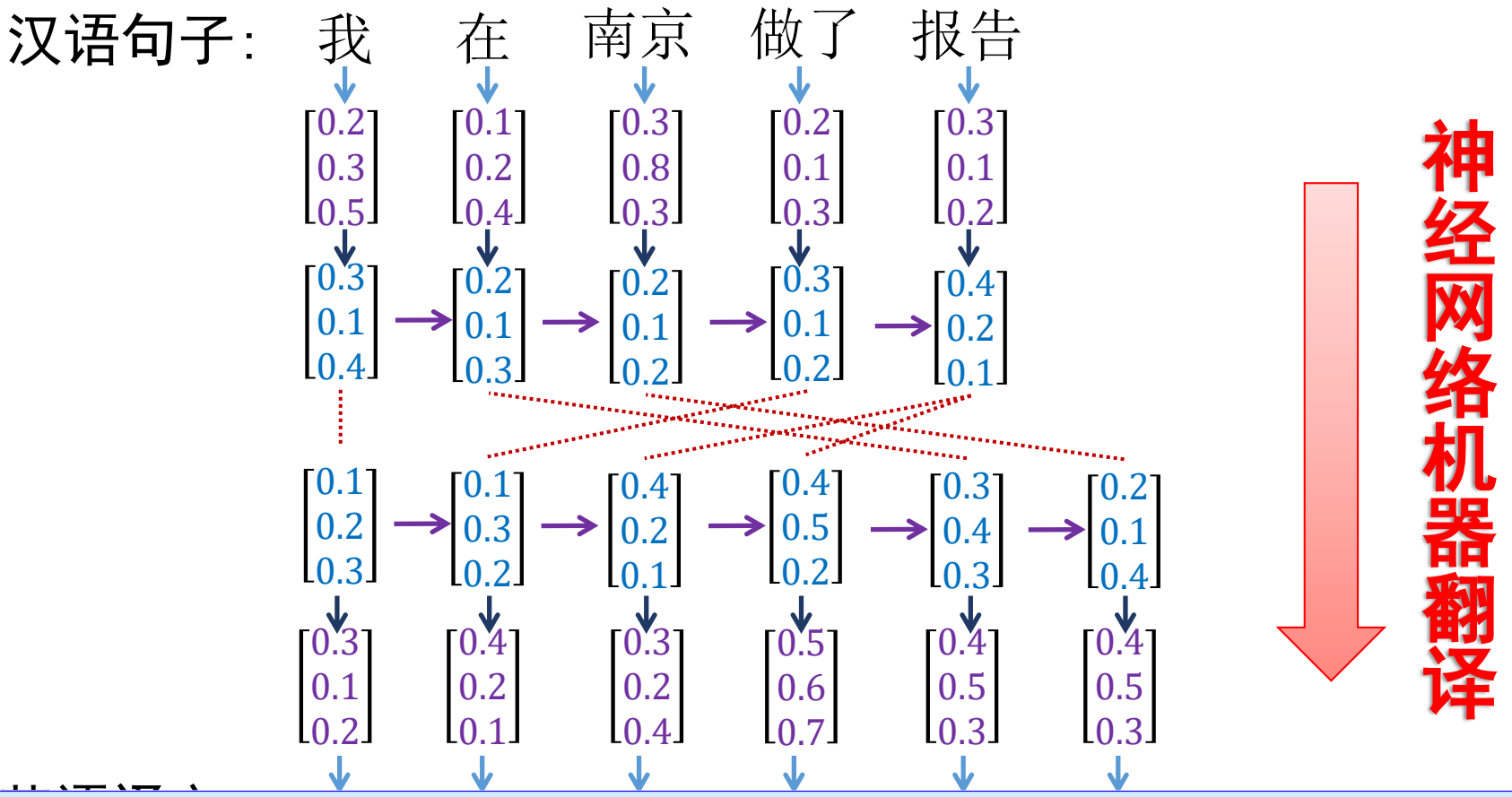
3. 深度学习方法应用

3.4 基于神经网络的机器翻译方法



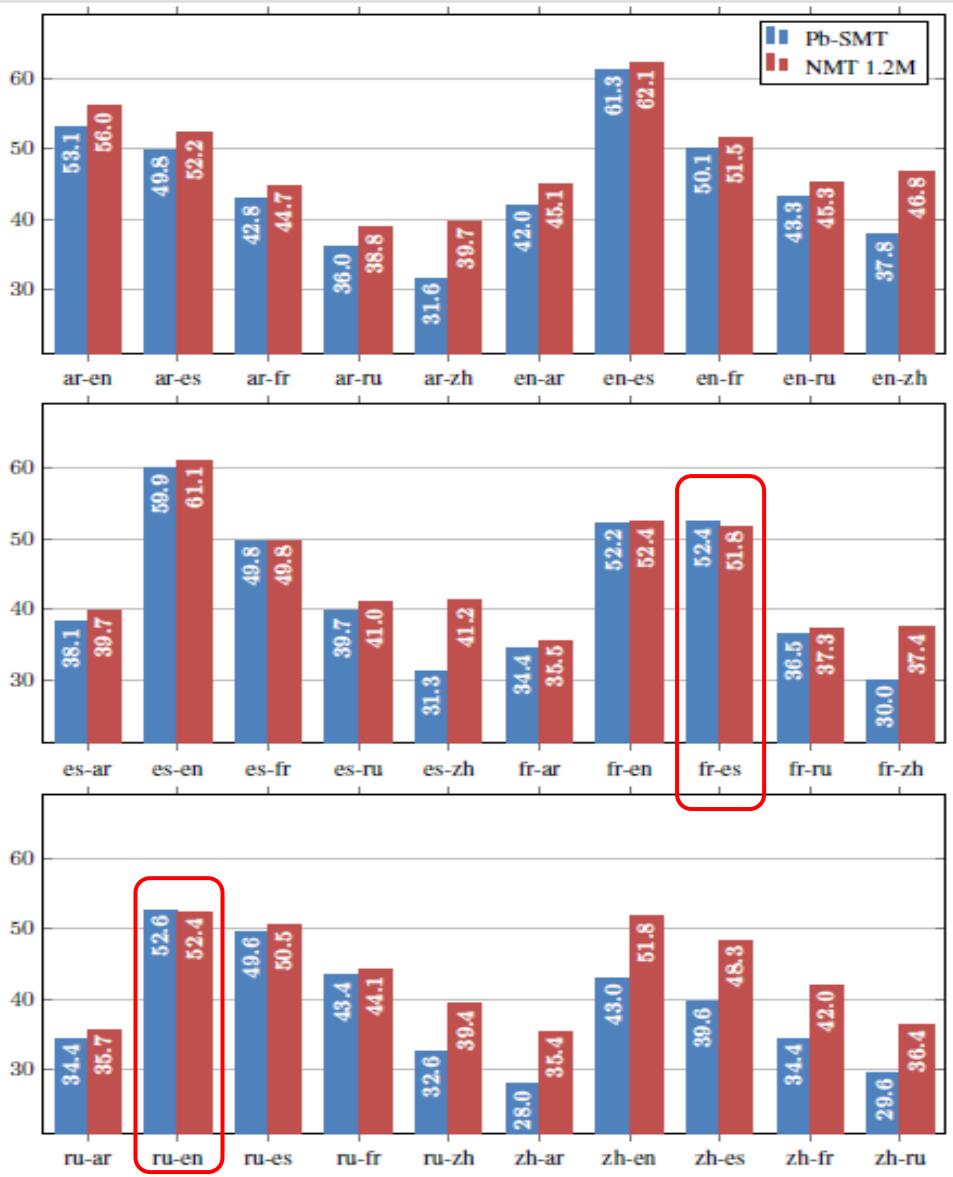
机器翻译框架

3. 深度学习方法应用



$$P(e_i) \approx P(e_i|e_1 \cdots e_{i-1}, f) \quad \text{目标函数: } L = \sum_i log\big(P(e_i)\big)$$

3. 深度学习方法应用



神经机器翻译几乎全面超越传统的统计翻译方法。

[Junczys-Dowmunt et al, 2016]

3. 深度学习方法应用

双语标注数据匮乏 ➡ 大量集外词无法翻译

悍然 对 南联盟 进行 了 大规模 轰炸

⬇ 悍然 、南联盟 是集外词

UNK 对 UNK 进行 了 大规模 轰炸

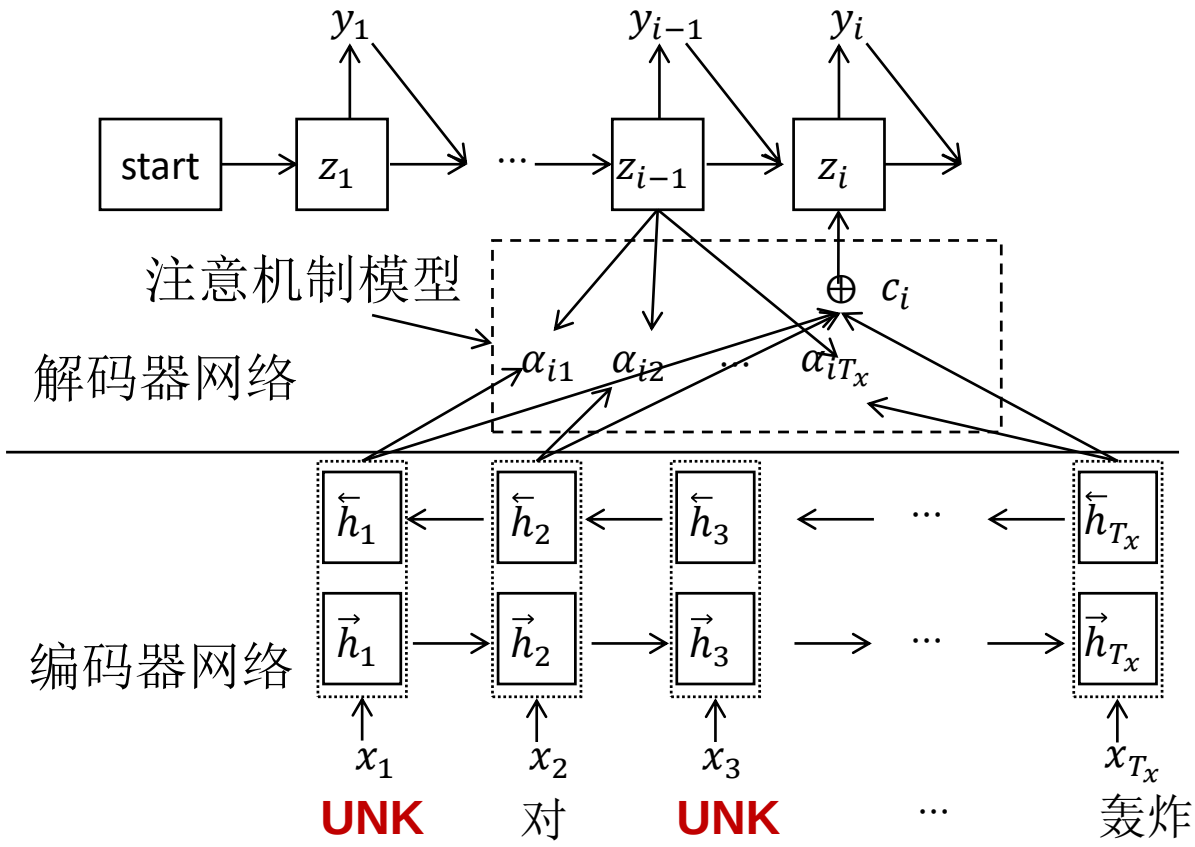
问题：句子结构不完整，影响整个句子的翻译

3. 深度学习方法应用

基于深度神经网络的翻译模型

目标语言句子生成

源语言语义表示



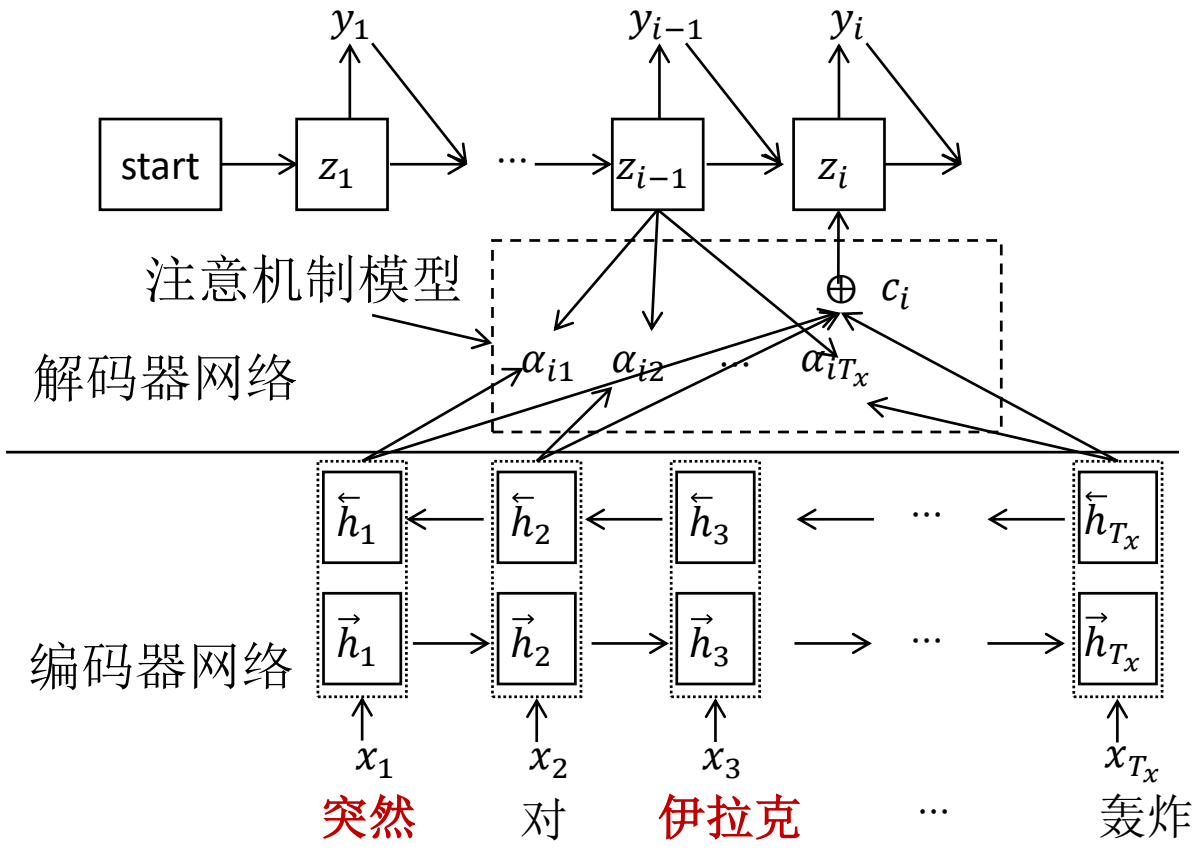
问题：源语言句子语义无法正确表示

3. 深度学习方法应用

解决方案：语义替换

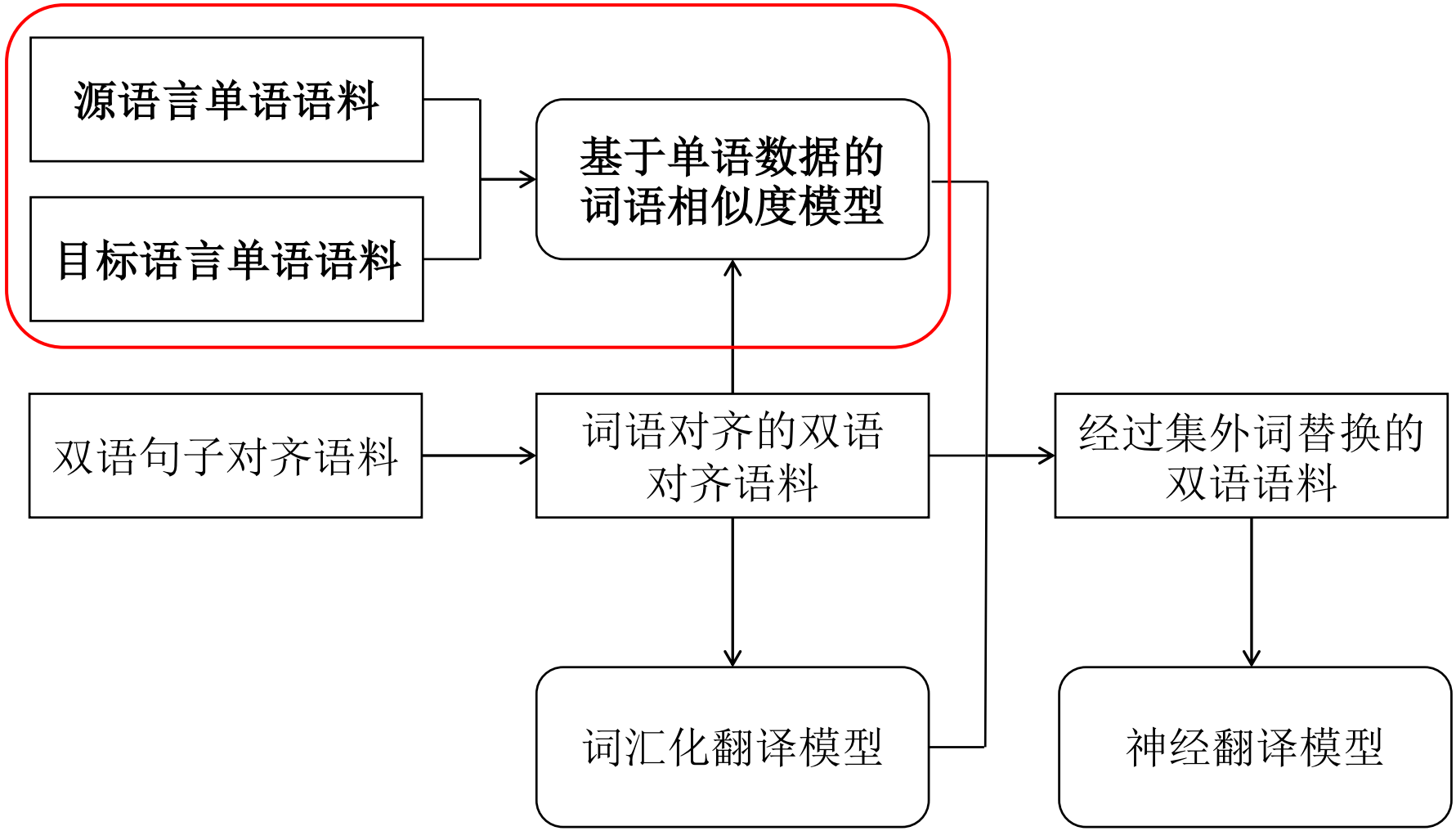
目标语言句子生成

源语言语义表示



悍然 → 突然　　南联盟 → 伊拉克

3. 深度学习方法应用


源语言单语语料
目标语言单语语料
基于单语数据的
词语相似度模型
双语句子对齐语料
词语对齐的双语
对齐语料
经过集外词替换的
双语语料
词汇化翻译模型
神经翻译模型

3. 深度学习方法应用

System	03 (dev)	04	05	06	Average
Bahdanau et al. (2015)	25.65	28.94	25.13	27.86	26.90
Luong et al. (2015)	27.63	30.02	26.42	28.72	28.20
Ours	**29.85**	**33.08**	**28.95**	**32.31**	**31.05**

平均>4.0的BLEU提升！

X. Li, J. Zhang and C. Zong. Towards Zero Unknown Word in Neural Machine Translation. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, New York, USA, July 12-15, 2016, pp. 2852-2858

内容提要

1. 引言
2. NLP方法概述
3. 深度学习方法应用
4. 讨论与结语

4. 讨论与结语

◆ 深度学习不等于深度理解

缉毒警察

Name　　Time

张小五/ 从警/ 20多年/ 来/ ，/ 历尽/ 千辛万苦/ ，/ 下/ 无数/ 战功/ ，/ 曾/ 被/ 誉为/ 孤胆英雄/ 。‖/ 然而/ ，/ 谁/ 也/ 未/ 曾/ 想到/ ，/ 就是/ 这样/ 一位/ 曾/ 让/ 毒贩/ 闻风丧胆/ 的/ 铁骨/ 英雄/ 竟然/ 为了/ 区区/ 小利/ 而/ 铤而走险/ ，/ 悔恨/ 之下/ 昨晚/ 在/ 家/ 开枪/ 自毙/ 。

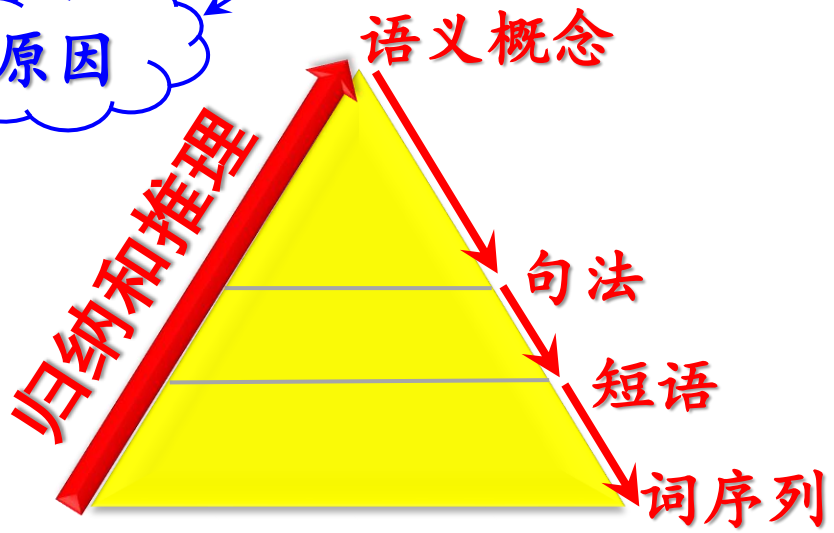
死去　　原因

① 分词（96%）

② 命名实体识别（90%）

③ 实体关系抽取（85%）

④ 语义角色标注（70-82%）



4. 讨论与结语

◆ 深度学习不等于深度理解

夫人穿着很得体，举止优雅，左臂上挂着一个暗黄色的皮包，右手领着一只白色的小狗，据说是京巴。

英文译文(Google Translate, 2016.10.19)：

Lady wearing a very decent, elegant manner, his left arm hanging on a dark yellow bag, his right hand led a white dog, is said to be Beijing Pakistan.

?

4. 讨论与结语

◆ 深度学习不等于深度理解

夫人穿着很得体，举止优雅，左臂上挂着一个暗黄色的皮包，右手领着一只白色的小狗，据说是局长夫人。

英文译文(Google Translate, 2016.10.19)：

Lady wearing a very decent, elegant manner, his left arm hanging on a dark yellow bag, his right hand led a white dog, is said to be his wife.

4. 讨论与结语

◆ **深度学习不等于深度理解**

①夫人穿着很得体，举止优雅。

The lady was well dressed and elegant.

②她左臂上挂着一个暗黄色的皮包，右手领着一只白色的小狗。

Her left arm hanging on a dark yellow bag, his right hand led a white dog.

③据说她是局长夫人。

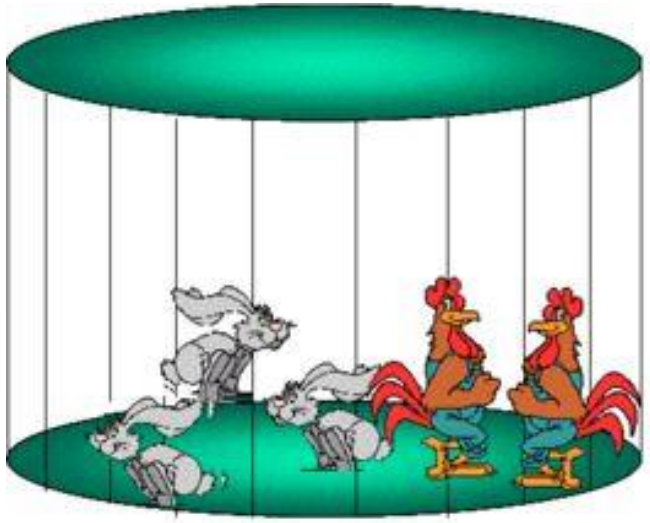
She is said to be the wife of the Secretary.

4. 讨论与结语

◆ 关于常识学习

一群鸡和兔子，放在同一个笼子里，上面有35个头，下面有94只脚，问有多少只鸡、多少只兔？

- 焦点词确定
- 常识获取
- 交互学习



4. 讨论与结语

ACL-IJCNLP'2015 会议论文投稿情况:

- 长文: 692篇，录用173篇
- 短文: 648篇，录用145篇

1340 篇投稿
318 篇录用

目前70%以上的NLP论文采用深度学习方法

第一作者为中国(大陆)人的论文数量:

- 长文：59/173 = 34.1%
- 短文：59/145 = 40.7%
- 合计：118/318 = 37.1%

（误差：±1）

4. 讨论与结语

People go to west, I go to east.

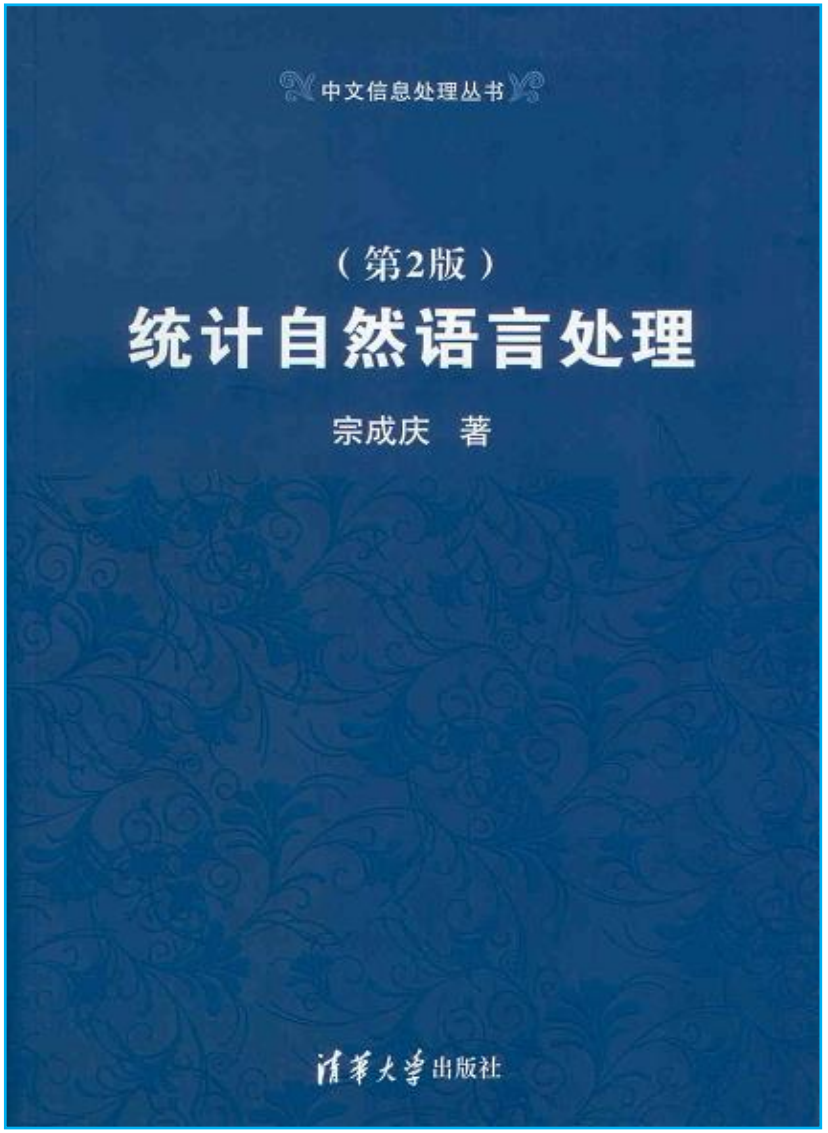
- LIU Bing, University of Illinois at Chicago

深度学习方法可能过时，机器学习不会过时。

- 周志华

谢谢！

Thanks!


中文信息处理丛书
（第2版）
统计自然语言处理
宗成庆 著
清华大学出版社