计算语言学

第9讲 机器翻译II

刘群
中国科学院计算技术研究所
liuqun@ict.ac.cn

中国科学院研究生院2002~2003学年第二学期课程讲义

大纲

- 基于翻译记忆的机器翻译方法
- 基于模板（模式）的机器翻译方法
- 双语语料库对齐技术
  - 句子对齐
  - 词语对齐
- 机器翻译的评价

翻译记忆方法 1

- 翻译记忆方法（Translation Memory）是基于实例方法的特例；
- 也可以把基于实例的方法理解为广义的翻译记忆方法；
- 翻译记忆的基本思想：
  - 把已经翻译过的句子保存起来
  - 翻译新句子时，直接到语料库中去查找
    - 如果发现相同的句子，直接输出译文
    - 否则交给人去翻译，但可以提供相似的句子的参考译文

翻译记忆方法 2

- 翻译记忆方法主要被应用于计算机辅助翻译（CAT）软件中
- 翻译记忆方法的优缺点
  - 翻译质量有保证
  - 随着使用时间匹配成功率逐步提高
  - 特别适用于重复率高的文本翻译，例如公司的产品说明书的新版本翻译
  - 与语言无关，适用于各种语言对
  - 缺点是匹配成功率不高，特别是刚开始使用时

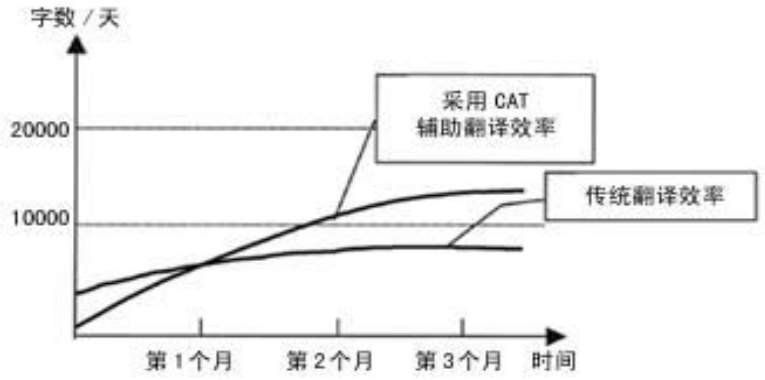
翻译记忆方法 3

- 计算机辅助翻译（CAT）软件已经形成了比较成熟的产业
  - TRADOS
    - 号称占有国际CAT市场的70%
    - Microsoft、Siemens、SAP等国际大公司和一些著名的国际组织都是其用户
  - 雅信CAT
    - 适合中国人的习惯
    - 产品已比较成熟
  - 国际组织： LISA（Localisation Industry Standards Association）
- 面向用户：专业翻译人员
- 数据交换：LISA制定了TMX（Translation Memory eXchange）标准。

翻译记忆方法 4

- 完整的计算机辅助翻译软件除了包括翻译记忆功能以外，还应该包括以下功能
  - 多种文件格式的分解与合成
  - 术语库管理功能
  - 语料库的句子对齐（历史资料的重复利用）
  - 项目管理：
    - 翻译任务的分解与合并
    - 翻译工作量的估计
  - 数据共享和数据交换

翻译记忆方法 5



基于模板(模式)的机器翻译方法 1

- 基于模板（Template）或者模式（Pattern）的机器翻译方法通常也被看做基于实例的机器翻译方法的一种延伸
- 所谓“翻译模板”或者“翻译模式”可以认为是一种颗粒度介于“翻译规则”和“翻译实例”之间的翻译知识表示形式
  - 翻译规则：颗粒度大，匹配可能性大，但过于抽象，容易出错
  - 翻译实例：颗粒度小，不易出错，但过于具体，匹配可能性小
  - 翻译模板（模式）：介于二者之间，是一种比较合适的知识表示形式
- 一般而言，单语模板（或模式）是一个常量和变量组成的字符串，翻译模板（或模式）是两个对应的单语模板（或模式），两个模板之间的变量存在意义对应关系

基于模板(模式)的机器翻译方法 2

- 模板举例：
  - 这个 X 比 Y 更 Z。
  - The X is more Z than Y.
- 模板方法的主要问题
  - 对模板中变量的约束
  - 模板抽取
  - 模板的冲突消解

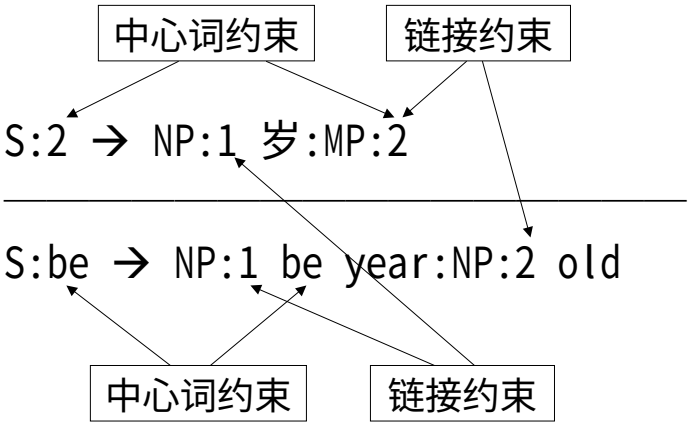
Pattern-Based CFG for MT 1

- Koichi Takeda, Pattern-Based Context-Free Grammars for Machine Translation, Proc. of 34th ACL, pp. 144-- 151, June 1996
- 给出了翻译模式的一种形式化定义，并给出了相应的翻译算法以及算法复杂性的理论证明

Pattern-Based CFG for MT 2

- 每个翻译模板由一个源语言上下文无关规则和一个目标语言上下文无关规则（这两个规则称为翻译模板的骨架），以及对这两个规则的中心词约束和链接约束构成；
- 中心词约束：对于上下文无关语法规则中右部（子结点）的每个非终结符，可以指定其中心词；对于规则左部（父结点）的非终结符，可以直接指定其中心词，也可以通过使用相同的序号规定其中心词等于其右部的某个非终结符的中心词；
- 链接约束：源语言骨架和目标语言骨架的非终结符子结点通过使用相同的序号建立对应关系，具有对应关系的非终结符互为翻译。

Pattern-Based CFG for MT 3



Pattern-Based CFG for MT 3

- 翻译的过程分为三步：
  - 使用源语言CFG骨架分析输入句子s
  - 应用源语言到目标语言的CFG骨架的链接约束，生成一个译文CFG推导序列
  - 根据译文CFG推导序列产生译文
- 模板排序的启发式原则：
  - 对于源文CFG骨架相同的模板，有中心词约束的模板优先于没有中心词约束的模板；
  - 对于同一跨度上的两个结点，比较其对应的模板的源文CFG骨架，非终结符少的模板优先于非终结符多的模板；
  - 中心词约束被满足的结点优先于中心词约束不被满足的结点；
  - 对于一个输入串而言，分析步骤越短（推导序列越短）越优先。

Pattern-Based CFG for MT 4

- 模板库的获取：假设T是一组翻译模板，B是双语语料库，<s,t>是一对互为翻译的句子
  - 如果T能够翻译句子s为t，那么do nothing；
  - 如果T将s译为t'（不等于t），那么：
    - 如果T中存在<s,t>的推导Q，但这个推导不是最优解，那么给Q中的模板进行实例化；
    - 如果不存在这种推导，那么加入适当的模板，使得推导成立；
  - 如果根本无法翻译s（分析失败），那么将<s,t>直接加入到模板库中。

模板的自动提取

- 利用一对实例进行泛化
  - Jaime G. Carbonell, Ralf D. Brown, Generalized Example-Based Machine Translation http://www.lti.cs.cmu.edu/Research/GEBMT/
- 利用两对实例进行比较
  - H. Altay Guvenir, Ilyas Cicekli, Learning Translation Templates from Examples Information Systems, 1998
  - 张健，基于实例的机器翻译的泛化方法研究，中科院计算所硕士论文，2001

通过泛化实例得到翻译模板

- 已有实例：
  - Karl Marx was born in Trier, Germany in May 5, 1818.
  - 卡尔·马克思于1818年5月5日出生在德国特里尔城。
- 泛化：
  - <Person> was born in <City> in <Date>
  - <Person>于<Date>出生在<City>
- 对齐
  - <Person> ⇔ <Person>
  - <City> ⇔ <City>
  - <Date>⇔<City>

通过比较实例得到翻译模板

- 已有两对翻译实例：
  - 我给玛丽一支笔 ⇔ I gave Mary a pen.
  - 我给汤姆一本书 ⇔ I gave Tom a book.
- 双侧单语句子分别比较，得到：
  - 我 给 #X 一 #Y #Z ⇔ I give #W a #U.
- 查找变量的对应关系：
  - #X ⇔ #W
  - #Y ⇔ ϕ
  - #Z ⇔ #U

实例库的匹配 1

- 实例匹配的目的是将输入句子分解成语料库中实例片断的组合，这是基于实例的机器翻译的关键问题之一，实例匹配的各种方法有很大的差异，还没有那种做法显示出明显的优势；
- 实例库匹配的效率问题：由于实例库规模较大，通常需要建立倒排索引；
- 实例库匹配的其他问题：
  - 实例片断的分解：
  - 实例片断的组合：

实例库的匹配 2

- 实例片断的分解
  - 实例库中的句子往往太长，直接匹配成功率太低，为了提高实例的重用性，需要将实例库中的句子分解为片断
  - 几种通常的做法：
    - 按标点符号分解
    - 任意分解
    - 通过组块分析进行分解

实例库的匹配 3

- 实例片断的组合
  - 一个被翻译的句子，往往可以通过各种不同的实例片断进行组合，如何选择一个最好的组合？
  - 简单的做法：
    - 最大匹配
    - 最大概率法：选择概率乘积最大的片断组合
  - 有点像汉语词语切分问题

片断译文的选择

- 由于语料库中一个片断可能有多种翻译方法，因此存在片断译文的选择问题；
- 常用的方法：
  - 根据片断上下文进行排歧；
  - 根据译文的语言模型选择概率最大的译文片断组合

实例库的对齐

- 实例库又称双语语料库（Bilingual Corpus）或平行语料库（Parallel Corpus）
- 双语语料库对齐的级别
  - 篇章对齐
  - 段落对齐
  - 句子对齐
  - 词语对齐
  - 短语块对齐
  - 句法结构对齐
- 基于实例的机器翻译中实例库必须至少做到句子级别的对齐

不同对齐级别的差异

- 段落对齐和句子对齐
  - 要求保持顺序（允许局部顺序的调整）
  - 只有一个层次
- 词语对齐和短语块对齐
  - 不要求保持顺序
  - 只有一个层次
- 句法结构对齐
  - 不要求保持顺序
  - 多层次对齐

句子对齐 1

汉语	英语	模式
1995年初我来成都的那天，没想到会是在一个冬季的漆黑的日子。	I little thought when I arrived in Chengdu in the dark, dark days of winter, early in 1995, that I would still be here more than five years later.	1:1
那时我也根本没有想到会在这儿呆上五年，也不知道我会遇到一位成都的女儿，并且后来还娶她为妻。一个完全陌生的家庭接纳了我，我也因此成为成都的一部分。	I little knew that I would meet one of Chengdu's daughters, and later marry her, thus acquiring a whole new family who embraced me as one of them, and thus I became part of this place.	2:1

句子对齐 2

对于篇章对齐（或者段落对齐）的一对文本(S,T)：

$$S = s_1 ... s_m, T = t_1 ... t_n$$

定义其对齐为A={A_1,…,A_k}，其中A_i称为一个句珠(Bead)：

$$A_i = (S_i, T_i) = (s_{a_{i-1}+1} ... s_{a_i}, t_{b_{i-1}+1} ... t_{b_i}),$$

其中$a_0 = 0 < ... < a_{i-1} < a_i < ... < a_k = m, b_0 = 0 < ... < b_{i-1} < b_i < ... < b_k = n$

整个对齐的概率为：

$$P(A) = \prod_{i=1}^{n} P(A_i)$$

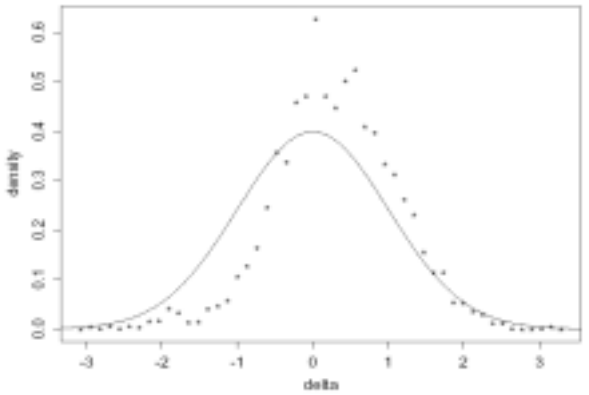
基于长度的句子对齐 1

- 基本思想：源语言和目标语言的句子长度存在一定的比例关系
- 用两个因素来估计一个句珠的概率
  - 源语言和目标语言中句子的长度
  - 源语言和目标语言中的句子数（对齐模式）

$$P(A_i) = P(S_i, T_i)$$
$$\approx P(l_{Si}, l_{Ti}) \times P(m_{Si}, m_{Ti})$$

基于长度的句子对齐 2

- 根据统计，随机变量$X=l_{Ti}/l_{Si}$服从正态分布



基于长度的句子对齐 3

- 设通过语料库统计得到X的期望为c，方差为v^2，那么随机变量δ将服从[0,1]正态分布：

$$\delta = \frac{X - c}{v} = \frac{l_T - cl_S}{vl_S} \sim N(0,1)$$

- 根据正态分布公式可以计算出(直接查表)：

$$P(l_S, l_T) = P(\delta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta^2}{2}}$$

基于长度的句子对齐 4

- 对齐模式的概率$P(m_S,m_T)$可以通过对语料库的统计得到。
- 下面是Gale & Church根据UBS语料库的统计结果：

Category	Frequency	Prob(match)
1-1	1167	0.89
1-0 or 0-1	13	0.0099
2-1 or 1-2	117	0.089
2-2	15	0.011
	1312	1.00

基于长度的句子对齐 5

- 最优路径的搜索：采用动态规划算法
- 定义$P(i,j)=P(s_1 \ldots s_i,t_1 \ldots t_j)$

$$P(i,j) = \max_{x,y}\{P(i-x, j-y) + Score(s_{i-x+1} \ldots s_i, t_{j-y+1} \ldots t_j)\}$$

- 最优对齐为P(m,n)所对应的路径

基于长度的句子对齐 6

- 优点
  - 不依赖于具体的语言；
  - 速度快；
  - 效果好
- 缺点
  - 由于没有考虑词语信息，有时会产生一些明显的错误
- 讨论
  - 长度计算可以采用词数或者字节数，没有明显的优劣之分

基于词的句子对齐 1

- 基本思想：互为翻译的句子对中，含有互为翻译的词语对的概率，大大高于随机的句子对
- 用两个因素来估计一个句珠的概率
  - 源语言和目标语言中互译词语的个数
  - 源语言和目标语言中的句子数（对齐模式）

$$P(A_i) = P(S_i, T_i)$$
$$\approx P(w_{Si}, w_{Ti}) \times P(m_{Si}, m_{Ti})$$

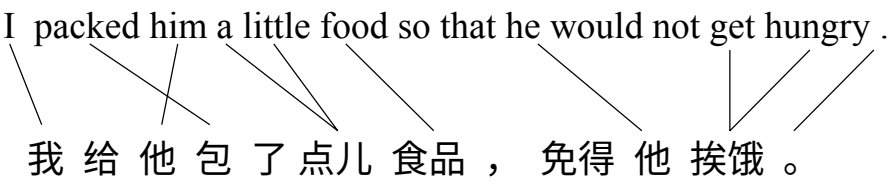
基于词的句子对齐 2

- 优点
  - 可以充分利用词语互译信息，提高正确率
- 缺点
  - 单独使用时，正确率有时低于基于长度的方法（取决于词典的规模质量等）
  - 时空开销大
- 讨论
  - 对于同源的语言（英语和法语，汉语和日语）可以利用词语同源信息而不使用词典

句子对齐小结

- 句子对齐的语料库是基于语料库的机器翻译的基础；
- 综合采用基于长度的方法和基于词汇的方法可以取得较好的效果；
- 句子对齐可以取得很高的正确率，已经达到实用水平。

词语对齐 1



- 特点：
  - 保序性不再满足
  - 对齐模式复杂：一对多、多对一、多对多都非常普遍

词语对齐 2

- 困难：
  - 翻译歧义：一个词出现两个以上的译词
  - 双语词典覆盖率有限：非常普遍的现象
  - 位置歧义：出现两个以上相同的词
  - 汉语词语切分问题
  - 虚词问题：虚词的翻译非常灵活，或没有对译词
  - 意译问题：根本找不到对译的词

词语对齐 3

- 一般而言，一个单词对齐的模型可以表述为两个模型的乘积：
  - 词语相似度模型(word similarity model)
  - 位置扭曲模型(word distortion model)

  用公式表示如下：

$$Score(e_i, c_j) = S(e_i, c_j) \times D(i, j)$$

词语相似度模型 1

- 翻译概率：IBM Model 1

$$S(e_i, c_j) = p(c_j \mid e_i) = \frac{\text{语料库中}e_i\text{翻译成}c_j\text{的次数}}{\text{语料库中}e_i\text{出现的次数}}$$

- T-Score：

$$T - score(\mathbf{e}, \mathbf{c}) = \frac{N_{\mathbf{ec}} \times Total - N_{\mathbf{c}} \times N_{\mathbf{e}}}{Total \times \sqrt{Total}}$$

  N_c：语料库中单词c出现的词数

  N_e：语料库中单词e出现的词数

  N_{ec}：语料库中单词e和单词c互译的词数

词语相似度模型 2

- 戴斯系数（dice coefficient）
  设S_1和S_2分别是两个集合，则这两个集合的戴斯系数可以通过如下公式计算

$$Dice(S_1, S_2) = \frac{2 | S_1 \cap S_2 |}{| S_1 | + | S_2 |}$$

- 把汉语词理解为汉字的集合，戴斯系数就是两个词中相同的汉字占两个词汉字总数的比例。考虑到汉字表意性，这种方法在计算汉语词相似度时有较好的效果
- 计算汉语词c和英语词e的相似度：
  - 先用英语词e查英汉词典，得到所有的汉语对译词；
  - 计算所有对译词和c的戴斯系数，取其中的最大值。

词语相似度模型 3

- 互信息（mutual information）
  通过两个事件X和Y各自出现的概率为p(X)和p(Y)，他们联合出现的概率为p(X,Y)，这两个事件之间共同的互信息量定义为：

$$I(X, Y) = -\log_2 \frac{p(X)p(Y)}{p(X, Y)}$$

- 当两个事件相互独立时，互信息量为0；
- 当两个事件倾向于同时出现时，互信息量为正；
- 当两个事件倾向于互相排斥时，互信息量为负；
- 利用互信息作词语相似度计算效果较差。

词语相似度模型 4

- ϕ^2方法：利用联立表（contingency table）

	Wt+	Wt-
Ws+	31,950(a)	12,004(b)
Ws-	4,793(c)	848,330(d)

$$\phi^2 = \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

- ϕ^2方法的效果比较好

词语相似度模型 5

- 对数似然比（ Log Likelihood Ratio,LLR ）

$$LLR = \log L(p_1,k_1,n_1) + \log L(p_2,k_2,n_2) - \log L(p,k_1,n_1) - \log L(p,k_2,n_2)$$

其中： $\log L(p,n,k) = k \log p + (n-k)\log(1-p)$

$$k_1 = f(w_t,w_s), k_2 = f(w_t,\neg w_s), n_1 = f(w_s), n_2 = f(\neg w_s)$$

$$p_1 = p(w_t \mid w_s) = \frac{k_1}{n_1}, p_2 = p(w_t \mid \neg w_s) = \frac{k_2}{n_2}, p = p(w_t) = \frac{k_1+k_2}{n_1+n_2}$$

对数似然比在使用中比较有效，在训练语料库规模较小时尤为明显

词语相似度模型 6

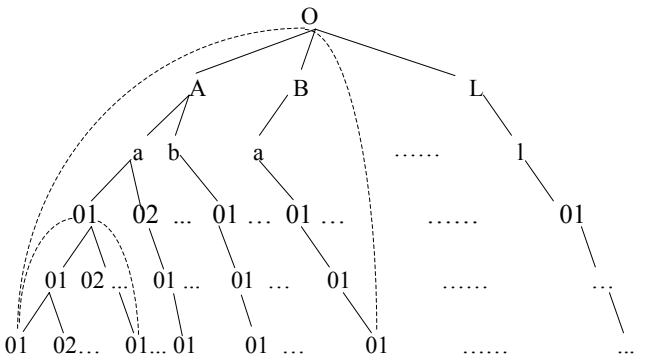
- 概念相似度
  利用某种形式的义类词典（Thesaurus），计算两个词语对应的概念之间的相似度

$$Sim(p_1, p_2) = \frac{\alpha}{d + \alpha}$$

其中d是概念p_1、p_2之间的距离，一般用概念层次体系中两个结点之间的距离来计算
α是一个可条件的参数

词语相似度模型 7

《同义词词林》的概念层次体系



虚线用于标识某上层结点到下层结点的路径

位置扭曲模型 1

- 绝对扭曲模型：IBM Model 2

$$d(j \mid i, m, l)$$

l：源语言句子长度
m：目标语言句子长度
i：源语言词语位置
j：目标语言词语位置

位置扭曲模型 2

- 相对偏移模型

$$dis(i, j) = \min(|L|, |R|)$$
$$L = |s_i - s_{i-1}| - |t_j - t_{j-1}|$$
$$R = |s_i - s_{i+1}| - |t_j - t_{j+1}|$$
$$d(i, j) = \begin{cases} d1 & if\ dis(i, j) = 0 \\ d2 & if\ dis(i, j) = 1 \\ d3 & if\ dis(i, j) = 2 \\ d4 & if\ dis(i, j) >= 3 \end{cases}$$

s_i是源语言e_i单词的位置
t_j是目标语言单词c_j的位置
s_i跟t_j对齐
s_{i-1}跟t_{j-1}对齐
s_{i+1}跟t_{j+1}对齐

位置扭曲模型 3

- 基于HMM的扭曲模型
  - 将每个对齐看作状态，对齐位置之间的转移是状态的转移，该对齐处的单词对作为输出。这样就可以将对齐问题映射到HMM上

词语对齐小结

- 词语对齐比句子对齐困难得多；
- 词语对齐主要使用一个词语相似度模型和一个位置扭曲模型；
- 词语对齐的副产品：双语词典抽取
  - 贪心算法：每次抽取可能性最高的词对；
  - 词语抽取和词语对齐反复迭代
  - 可以抽取多词单元（n元组）

机器翻译评价 1

- 最早的机器翻译评价：ALPAC报告
- 机器翻译评价的常用指标
  - 忠实度（Adequacy）：译文在多大程度上传递了源文的内容；
  - 流利度（Fluency）：译文是否符合目标语言的语法和表达习惯；
  - 信息度（Informative）：用户可以从译文中获得信息的程度（通过选择题评分）
- 绝对评价和相对评价

机器翻译评价 2

- 人工评价
  - 准确
  - 成本极高
  - 不能反复使用
- 自动评价
  - 准确率低
  - 成本低
  - 可以反复使用

机器翻译评价 3

- 机器翻译的评价一直是机器翻译研究领域中一个备受关注的问题；
- 机器翻译的自动评价越来越引起重视
  - “评测驱动”成为自然语言处理研究的一个主要动力
  - 大规模语料库的出现、各种机器翻译算法的提出，使得开发过程中频繁的评测成为必需
  - 开发过程中频繁的评测只能通过采用自动评测方法

机器翻译的自动评测

- 完全匹配方法
  - 与参考译文完全相同的译文才被认为是正确的
  - 显然该标准过于严格，不适用
- 编辑距离方法
- 基于测试点的方法
- 基于N元语法的方法

基于编辑距离的机器翻译评测 1

- 编辑距离定义：

  从候选译文到参考译文，所需要进行的插入、删除、替换操作的次数
- 举例说明：
  - 源文：She is a star with the theatre company.
  - 机器译文：她 是 与 剧院 公司 的 一 颗 星 。
  - 参考译文：她 是 剧团 的 明星 。
  - 编辑距离：6
    - 插入：与 公司 一 颗
    - 替换：剧团→剧院 明星→星

基于编辑距离的机器翻译评测 2

- 单词错误率：编辑距离除以参考译文中单词数
  - 这个指标是从语音识别中借鉴过来的。
  - 由于语音识别的结果语序是不可变的，而机器翻译的结果语序是可变的，显然这个指标存在一定的缺陷。
- 与位置无关的单词错误率：计算编辑距离时，不考虑插入、删除、替换操作的顺序
  - 也就是说，候选译文与参考译文相比，多出或不够的词进行删除或插入操作，其余不同的词进行替换操作。
  - 这个指标与单词错误率相比，允许语序的变化，不过又过于灵活。

基于测试点的机器翻译评测 1

- 俞士汶等，机器翻译译文质量自动评估系统，中国中文信息学会1991年论文集，pp. 314～319
- 基本思想
  - 对于每一个句子，孤立测试点，简化测试目标（模拟人类标准化考试的办法）
  - 对于每一个句子，采用一种TDL语言描述的BNF去与译文匹配，匹配成功则正确，否则错误
  - 大批量出题，全面评价机器翻译译文质量

基于测试点的机器翻译评测 2

- 测试点分组：
  单词、词组、词法、语法（初、中、高级）
- 测试点示例：
  - 源文：I am a student.
  - 测试：译文中出现“学生/大学生”为正确
  - 源文：I bought a table with three dollars.
  - 测试：“买”出现在“美元”之后为正确
  - 源文：I bought a table with three legs.
  - 测试：“买”出现在“腿”之前为正确

基于测试点的机器翻译评测 3

- 优点：
  - 全自动
  - 实验证明，评价结果是可信的
  - 可以按照人类专家的要求进行单项评测
- 缺点
  - 题库的构造需要具有专门知识的专家，并且成本较高

基于N元语法的机器翻译评测 1

- Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu, Bleu: a Method for Automatic Evaluation of Machine Translation, IBM Research, RC22176 (W0109-022) September 17, 2001
- 基本思想
  - 用译文中出现的N元组和参考译文中出现的N元组相比，计算匹配的N元组个数与候选译文的N元组总个数的比例
  - 允许一个源文有多个参考译文，综合评分

基于N元语法的机器翻译评测 2

源文：党指挥枪是我党的行动指南。
候选译文：

- It is a guide to action which ensures that the military always obeys the command of the party
- It is to insure the troops forever hearing the activity guidebook that party direct

参考译文：

- It is a guide to action that ensures that the military will forever heed party commands
- It is the guiding principle which guarantees the military forces always being under the command of the party
- It is the practical guide for the army to heed the directions of the party

基于N元语法的机器翻译评测 3

- 两个改进：
  - 对于候选译文中某个n元接续组出现的次数，如果比参考译文中出现的最大次数还多，要把多出的次数“剪掉”（不作为正确的匹配）。
  - 为了避免“召回率”过低的问题，BLEU的评价标准又对比参考译文更短的句子设计了“惩罚因子”。

基于N元语法的机器翻译评测 4

- BLEU的总体评价公式如下：

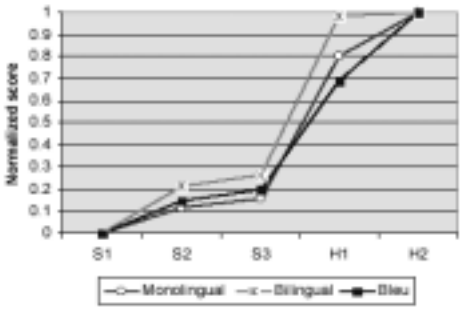
$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right).$$

其中，p_n是出现在参考译文中的n元词语接续组占候选译文中n元词语接续组总数的比例，$w_n = 1/N$，N为最大的n元语法阶数（实际取4）。

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \le r \end{cases}.$$

其中c为候选译文中单词的个数，r为参考译文中与c最接近的译文单词个数。

基于N元语法的机器翻译评测 5



其中S1、S2、S3分别是三个不同的机器翻译系统提供的译文，H1和H2是两个人类翻译者提供的译文。蓝线是BLEU系统评测的结果，红线是只懂目标语言的人类专家提供的评测结果，绿线是同时懂源语言和目标语言的人类专家提供的评测结果。

基于N元语法的机器翻译评测 6

- 这种方法比较好地模拟了人对机器翻译结果的评价
  - 对于低质量译文比高质量译文的评价跟准确；
  - 评价结果与只懂目标语言的人的评价结果更接近（相对于懂双语的人而言）
- 优点
  - 全自动
  - 可以提供多种参考译文综合考虑，结果更全面
  - 容易构造测试集，不需要专门知识

复习思考题

- 利用《圣经》双语语料库实现一个词语对齐系统，并从中抽取出一部包含多词单元的双语词典。