

专利机器翻译的现状与展望

王米利 庄 丽 丁冬冬 和晓萍 李晶晶

(云南民族大学 基础数学与计算机学院, 云南 昆明 650031)

【摘 要】文章回顾机器翻译研究的历史, 介绍典型的机器翻译方法, 包括: 基于规则、基于统计以及基于实例的机器翻译方法; 针对机器翻译的研究现状, 详细介绍和分析了专利机器翻译发展现状, 最后对专利机器翻译进行总结和展望。

【关键词】机器翻译; 基于规则; 基于统计; 基于实例; 专利机器翻译

近几年来, 自然语言处理的研究已经成为热点, 而机器翻译作为自然语言研究领域的一个重要分支, 同时也是人工智能领域的一个课题, 同样为大家所关注。纵观机器翻译的研究历史, 从上个世纪 40 年代英国工程师 Booth 和美国工程师 Weaver 提出利用计算机进行翻译的想法, 到 50 年代欧美国家投入大量的人力、物力致力于机器翻译的研究, 再到 60 年代 ALPAC 置疑报告的提出, 机器翻译走向沉寂。最近的二十年, 随着语言学理论的发展、计算机技术的进步以及统计学和机器学习方法在自然语言处理领域中的广泛应用, 人们对机器翻译本身的应用背景、目标等有了更加准确的认识, 机器翻译在此背景下取得了长足的发展。基于统计、基于实例等新的机器翻译方法也都是在这一时期出现, 一些机器翻译系统也从实验室走向了市场。在中国, 机器翻译的研究从 50 年代开始, 多家大学和研究机构先后开发出俄汉、英汉、汉英、日汉、汉日、蒙汉、藏汉等机器翻译系统, 同时在汉语的自然语言理解方面做了大量的研究。在看到机器翻译研究取得进展的同时, 我们也知道, 由于对语言规律本身认识仍然不足, 以及计算机对语言理解的局限性, 再加上不同语言之间存在着语法结构、构造方式、语言习惯、社会背景等等的不同, 机器翻译的效果与大家所期待的仍有非常大的差距。本文第二部分介绍几种经典的机器翻译方法; 第三部分给出近期专利文本机器翻译的研究现状, 介绍几种专利文本的机器翻译方法; 最后, 对机器翻译研究进行总结和展望。

一、典型的机器翻译方法和技术

机器翻译是建立在语言学、数学、信息学、计算机科学等学科基础上的多边缘学科(它的发展是建立在语言学、数学、计算机科学的发展基础之上)。现代理论语言学的发展、计算机科学的进步以及概率统计学的引入, 对机器翻译的理论和方法都产生了深刻的影响。Weaver 机器翻译思想的提出, 开始了机器翻译的研究热潮。Chomsky 在 50 年代后期提出的短语结构语法, 给出了“从规则生成句子”的原则。由于短语结构语法采用单一标记的短语结构来描述句子的构成, 描述粒度过粗, 因此存在约束能力弱、生成能力过强问题, 人们逐渐意识到仅依靠单一的短语结构信息, 不能充分判别短语类型和确认短语边界, 于是, 复杂特征集和词汇主义被引入自然语言语法系统, 广义短语结构语法、词汇功能语法、中心语驱动的短语结构语法等语言学理论也相应的产生。在这个发展过程中, 基于规则方法一直是机器翻译研究的主流。在基于规则的方法中, 语言规则的产生需要大量的人力, 而且大量的语言规则之间往往存在着不可避免的冲突。另外, 规则方法在保证规则的完备性和适应性方面也存在着不足。而此时, 统计学方法在自然语言处理的语音识别领域

取得了比较好的效果, 于是, 基于统计的机器翻译应运而生。随着双语语料的大量增加、计算机性能的提高, 基于实例的机器翻译方法被提出, 并由此泛化产生了基于模板的机器翻译方法。下面我们分别介绍几种典型的机器翻译方法。

(一) 基于规则的机器翻译方法 (Rule-Based Machine Translation)

从 Chomsky 提出转换生成文法之后, 基于规则的方法一直是机器翻译研究的主流, Chomsky 认为一种语言无限的句子可以由有限的规则推导出来。早期的机器翻译系统, 从体系结构上可以分为直译式、转换式和中间语言式, 它们的不同之处在于对源语言分析的深度, 它们的相同点是都需要大规模的双语词典、大量的源语言推导规则、语言转换规则和目标语言生成规则。其中, 转换式的基于规则方法对源语言分析得比较深, 它涉及到词汇结构分析、语法分析、语义分析, 并完成词汇、语法、语义三层结构从源语言到目标语言的转换, 而且转换式的方法又充分考虑了源语言和目标语言之间的特征联系, 它比中间语言方法更容易获得高质量的翻译结果。因此, 转换式的方法更多地被应用在早期的机器翻译系统中, 整个翻译过程被分为: 源语言分析部分, 转换部分和目标语生成部分。而早期的系统, 如德国西门子的 METAL 系统、美国的 SYSTRAN 系统、日本日立公司的 ATHENE 系统以及中国中软公司的 HY-1 汉英系统, 都是基于转换的机器翻译系统。基于规则的机器翻译的优点在于: 规则可以很准确的描述出一种语言的语法构成, 并且可以很直观地表示出来。机器可以按照一组规则来理解它面对的自然语言, 这组规则包含了不同语言层次的规则, 包括用以对源语言进行描述的源语言分析规则、用以对源语言和目标语言之间的转换规则以及用于生成目标语的生成规则。由此可见, 基于规则的机器翻译的核心问题是构造完备的或适应性较强的规则系统。但是, 规则库的建立需要花费大量的人力和物力, 即使如此, 规则的完备性仍然不能得到保证, 规则库很难覆盖所有的语言现象。随着规则数量的增加, 规则之间的冲突很难避免; 很难用系统化的规则分类体系、恰当的规则粒度去刻画语言特征。而且早期的规则系统采用的都是确定性规则, 即: 非此即彼的规则, 系统的适应性很差。基于上述问题, 如何自动地获取语言规则、如何更好的表示规则以及如何更好地增强系统的适应能力成为研究人员关注的焦点。随着大量语料库的产生, 统计方法为我们提供了很好的从已有的语言资源中自动得到我们所需要的语言信息的工具。复杂特征集和合一运算的提出也使得我们能以更细的粒度、更加准确的知识表示形式来描述规则, 而词汇化的信息也更多的来自于标注语料库。针对确定性规则降低了系统的鲁棒性的弱点, 概率上下文无关文法从全局最优的角度考虑, 产生最优的翻

译结果,为机器翻译系统的实用化奠定了基础。随着这些方法的引入,传统的基于规则的机器翻译方法研究逐步发展成为对以规则为基础、语料库方法为辅助的高性能机器翻译方法的研究。

(二) 基于统计的机器翻译方法 (Statistical-Based Machine translation)

除了在某些特定的受限领域,基于规则的机器翻译,取得了比较好的效果之外,在大部分的实验中,基于规则的机器翻译远远没有达到人们的要求。而随着语料库语言学的发展和统计学、信息论在自然语言处理领域的应用,人们尝试着用统计的方法进行机器翻译的研究。对于机器翻译来说,基于统计的方法可以从两个层面上来理解,一种是指某些概率统计的方法在具体的机器翻译过程中的应用,比如用概率统计的方法解决词性标注的问题、词义消歧的问题等。另一种较狭义的理解是指纯粹的基于统计的机器翻译,翻译所需的所有知识都来源于语料库本身。这一节我们主要介绍这种纯统计的机器翻译方法。IBM 的 Brown 在 1990 年首先将最初应用于语音识别领域的统计模型用于法英机器翻译。基本思想是:用信道模型把机器翻译看作一种解码的过程。

基于统计的机器翻译的关键是,首先定义最适合的语言概率模型和翻译概率模型,其次,需要从已经存在的语言资源中,对语言模型和参数模型的概率参数进行估计。早期的语言模型基本上采用二元、三元模型,语言模型的参数估计需要大量的单语语料库,二元模型参数通过两个词的搭配频率来计算,三元模型参数则通过计算相邻三元词的出现频率进行估计。近年来,基于统计的机器翻译翻译采用基于语法的语言模型,利用树库训练语言模型参数。翻译模型需要计算源语和目标语对之间的转换概率,因此翻译模型参数估计需要大量对齐的双语语料库。Brown 详细介绍了五种翻译模型,并且用 EM 算法对双语语料进行训练,估计翻译模型参数,尽管统计机器翻译在语音识别领域取得了一定的成功,但是它需要大量的双语语料库,而且存在着数据稀疏问题。因此,如何构建大规模的对齐双语语料库,以及找到比较好的平滑算法进行准确的参数估计,成了基于统计机器翻译系统实现中的关键问题。除此之外,要找到最优的译文,也需要好的搜索算法。统计机器翻译的质量很大程度上取决于语言模型和翻译模型,而最初的统计模型很少考虑语言的特征信息,对一些特定语言所特有的语言特征分析得不够。例如,在英语的词汇形态分析中,对单词“see”和“saw”,原型和过去型完全按照不同的词汇进行参数估计,这就造成了对语言模型和翻译模型构建的不准确性。IBM 将一些浅层的词汇信息结合到统计模型当中去,Franz Josef 在 ACL' 02 上发表的文章介绍的基于最大熵的统计机器翻译方法中,训练模型中也充分考虑了源语言和目标语言的语言特征。语言特征的引入,提高了基于统计的机器翻译的性能,同时也对语料库的标注提出了更高的要求。

(三) 基于实例的机器翻译方法 (Example-Based Machine Translation)

基于实例的机器翻译思想最早由 Nagao 提出,其基本思想是,在已有的源语言实例句库中,待翻译句子按照类比原理匹配出最相似的实例句,取出实例句对应的目标语句子,进行适当的改造,最终得出待翻译句子所相应的目标语句子。整个翻译过程实际上是一个匹配过程。它的特点是不需

要对源语言进行任何的分析,仅仅是通过类比进行翻译。从翻译过程来看,句子一级对齐的双语语料库是基于实例的机器翻译系统的知识源,在基于实例的机器翻译系统中,双语对齐语料库被称为翻译记忆库(Translation Memory)。对于基于实例的系统,首先,待翻译句子需要从翻译记忆库中找出最相似的源语言句子,一般根据词典或者语言的本体知识(Ontology),根据句子中词汇或者词类之间的语义距离来计算句子的相似度,Federica 在 2002 年的一篇文章概述了基于实例机器翻译相似度的计算方法以及句子匹配算法。基于实例的机器翻译系统的翻译质量取决于翻译记忆库的规模和覆盖率。因此如何构建大规模翻译记忆库成为基于实例的机器翻译研究的关键问题。对于双语语料对齐研究,Gale 等描述了基于长度和基于偏移量的语料库的句子和段落对齐方法,Kay 提出了基于词汇特征的句子对齐方法。现阶段,由于缺少大规模的双语对齐语料库,基于实例的机器翻译方法匹配率并不是很高,而基于实例的机器翻译如果匹配成功,可以获得高质量的译文,因此基于实例的机器翻译一般和基于规则的机器翻译结合使用,Satoshi 就提出了基于规则和基于实例相结合的机器翻译方法,产生了比较好的翻译结果。对于匹配命中率过低的问题,我们试着做到短语级的双语对齐,以提高匹配命中率,通过短语结构的局部匹配,组合相应的目标语句子框架,完成句子的翻译,这种方法进而泛化为基于模板(Template-based)的机器翻译,通过大规模的双语语料,自动抽取翻译模板,翻译过程匹配模板库。这种方法增加了匹配的命中率,模板库规模比实例库要小,因此也提高了系统的效率。而模板的自动获取仍然是翻译的关键。基于实例的机器翻译方法依然面临着很多的问题,对于相似度计算,如果计算词类或者短语级的相似度,则需要首先对我们的翻译记忆库本身进行标注。而且很难定义一个相似度标准选出最合适的相似句,此外随着翻译记忆库规模的扩大,需要一个高速的查询匹配算法,同时需要在增加翻译记忆库的规模、提高匹配率的同时,保证翻译记忆库的冗余度。在机器翻译研究的过程中,各种机器翻译方法层出不穷,其它的还有基于模式的机器翻译、基于神经网络的机器翻译、基于对话的机器翻译、基于原则的机器翻译等等,由于这些方法不是主流,本文就不再一一介绍。现有的各种机器翻译方法在现阶段的机器翻译研究中被广泛采用,它们之间已经没有严格的界限。基于规则的机器翻译方法结合语料库的方法,大量使用统计方法获取语言信息,而基于统计的机器翻译和基于实例的机器翻译更是相互渗透,这两种方法统称为基于语料库的方法,因为它们同样依靠双语语料库。

二、专利文本的机器翻译方法

近年来,随着信息交流量的急速增加,专利文献的翻译的需求也与日俱增,因此专利机器翻译受到越来越多的关注与重视。与其他翻译有所不同,专利文本翻译有其独特的特点和难点。特别是汉英专利文献机器翻译具有广泛的实际需求,同时也存在需要特殊考虑的问题。针对汉英专利文献机器翻译中的特殊问题,本部分将列出当前各研究单位所研究探索的方法:基于词典的句对齐方法、基于概念层次网络(HNC)的方法、微粒群优化改进 Champollion 的专利文本机器翻译。

(一) 基于词典的句对齐方法的专利文本机器翻译

和其他语料相比,专利语料的句对齐更加困难。之前的

研究表明,在专利语料中,按照字面规范翻译的文本大约仅占文本总数的四分之一,并且普遍存在的松弛翻译问题,同时专利语料中存在大量的多领域的专业术语,这些术语分布在日常生活,运输,化学,纺织,建筑,机械,物理,电气等领域。

句对齐算法分为两类:一类是无指导的方法,包括基于长度的方法和基于统计的方法。基于长度的方法把句子长度作为对齐依据;基于统计的方法把句子间的对齐关系分解为词之间的对齐关系,词之间的对齐关系使用统计模型获得。这两种方法的优点是不需要词典或训练语料,只要提供对齐的双语语料就可以。缺点是在噪声语料上的性能较差。第二类是有指导的对齐算法,主要是基于词典的方法。和基于统计方法一样,基于词典的方法也把句子间的对齐关系分解为词之间的对齐关系。不同之处在于,基于词典的方法通过词典获得词与词之间对齐关系,而不是通过统计模型获得。这种方法的优点是在噪声语料上的性能较好。Ma Xiaoyi 在 2006 年的论文中实现了基于词典的句对齐的工具 Champollion。

从句对齐的角度看,专利语料的句对齐不适合使用无指导的方法。因此文献以开源的句对齐工具 Champollion 为基础,使用专利领域词典,有效的提高句对齐的正确率。改进 Champollion 的对齐运算,使其能够使用多词表达式(multi-word expression)来扩展其词典。改进算法中,句子间的对齐关系被分解成 n-gram 间的对齐关系,提高词对齐的效率。

因此,在使用基于词典的句对齐方法时,和常规语料相比,需要更多词典来达到满意的对齐效果。专利文本句对齐是专利研究与应用的基础任务,被广泛应用于专利自动翻译、跨语言检索等领域。

(二) 使用微粒群优化改进 Champollion 的专利文本机器翻译

统计机器翻译效果依赖于平行语料的对齐效果,在语料充分的情况下,需要减少对齐中的噪声,对齐精度显得尤为重要。

针对专利文本统计机器翻译,提出基于群智的微粒群优化(PSO,particle swarm optimization)改进 Champollion 对齐精度。用三个优化权重改写 Champollion 相似度函数,以较小代价搜索最优对齐惩罚、长度惩罚参数。

PSO 算法中包含多个个体,每个个体是 D 维搜索空间中的一个代理,以一定速度在空间中飞行。

个体飞行时参考本身经验及邻居经验

第 i 个微粒的位置为: $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$

第 i 个微粒的速度为: $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$

第 i 个微粒的最好位置为: $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})$

当前群体的最好位置为: $P_g = (p_{g1}, p_{g2}, \dots, p_{gD})$

每次迭代的第 d 维 ($1 \leq d \leq D$) 按如下公式变化:

$$v_{id} = w \cdot v_{id} + c_1 \cdot \text{rand}() \cdot (p_{id} - x_{id}) +$$

$$c_2 \cdot \text{Rand}() \cdot (p_{gd} - x_{id})$$

$$X_{id} = X_{id} + V_{id}$$

其中, w 称为惯性权重, c_1 和 c_2 为加速常数, rand() 为 [0,1] 范围随机数。

对每个微粒,将它的适应值和全局所经历最好位置的值

作比较,如果较好,修正微粒的值。

Champollion 借用信息检索中 TF-IDF 计算片段的相似度,表达为:

$$\text{sim}_1(E, C) = \sum_{i=1}^k \lg(\text{stf}(e_i', c_i') \cdot \text{idtf}(e_i'))$$

$$\text{s.t. idtf} = T / \# \text{tf_doc}$$

$$\text{stf} = \# \text{tf_segm} + 1.$$

其中: stf 是片段范围内的词频, idtf 是文档范围内的逆向词频, T 是文档中的总词数, # tf_doc 是片段中的词语在文档范围内的词频。

Champollion 引入对齐惩罚和长度惩罚,表达为:

$$\text{sim}(E, C) = \text{sim}_1(E, C) \cdot \text{align_pena}_{ij} \cdot \text{leng_pena}(E, C).$$

Stf 与 idtf 被包装成 $\text{sim}_1(E, C)$, 给予一个权重,参与优化改写相似度函数如下:

$$\text{sim}(E, C) = r_1 \cdot \text{sim}_1(E, C) \cdot r_{ij} \cdot \text{align_pena}_{ij} \cdot r_2 \cdot \text{leng_pena}(E, C).$$

其中, r_1 是 $\text{sim}_1(E, C)$ 的权重, r_2 是长度惩罚参数的权重, r_{ij} 是非 1-1 惩罚参数的权重,包括: $r_{12}, r_{21}, r_{22}, r_{23}, r_{32}, r_{13}, r_{31}, r_{33}, r_{14}, r_{41}$

在训练集上用 PSO 生成参数和权重,调用 Champollion; 用 F1 值评估句子对齐结果;将评估值作为微粒适应度反馈到 PSO;通过 PSO 发现最优参数和权重。

(三) 基于概念层次网络(HNC)理论的专利文本机器翻译

规则系统的缺点:规则书写过于简单,或互相矛盾,或不足以处理复杂的语言现象,规则精确描述语言现象,规则描写过于复杂,概括性差,不可行。

基于概念层次网络(HNC)理论的纯规则汉英专利机器翻译系统,该系统将整个翻译过程分为分析、转换、生成三个模块。首先利用 HNC 理论提供的语义特征对源语言进行语义分析,得到源语言的语义分析树;其次在此基础上对语义分析树进行过渡转换,将其结构调整为目标语的合法表达形式;最后在转换树上进行目标语生成。

语义分析过程:根据语义特征,分为 L 类概念和 v 类概念。在句间关系层面,将句子切分为若干小句,判断小句间的语义关系和共享关系。在小句层面,切分小句结构。语块内部构成层面,判断构成形式,切分结构。串联结构(一种 TD-SCDMA 系统中上行多小区联合检测的方法)串联语块 1+ 串联标记“的”+……+ 串联标记“的”+ 串联语块 n; 并联结构,并联语块 1+ 并联标记“、/ 和 / 或”+……+ 并联标记“、/ 和 / 或”+ 并联语块 n; 句蜕,句蜕修饰成分+ 句蜕标记“的”+ 句蜕中心成分。由数据排序装置所排序的数据中的有效数据,得到语义分析树。

过渡转换:调整顺序,给出 L 类概念的翻译,给出 V 类概念的形态特征,分层次进行,为转换保留分析现场,利用父子关系进行传递。句间关系过渡转换,小句结构过渡转换:包括格式转换,样式转换,特征语块的转换。语块内部结构过渡转换:串联结构的过渡转换,是一种 TD-SCDMA 系统中上行多小区联合检测的方法;句蜕的过渡转换,由 A/D 转换器转换的电压和电流被转换为功耗。

目标语生成:对概念类别进行判定和选择,对译词进行比较和选择,目标词语变形。

基于概念层次网络(HNC)理论的机器翻译发展探索经历了十几个年头,与传统的机器翻译系统相比有较强的适应

性, 期盼有较为成熟的翻译系统产生。

三、总结与展望

本文主要介绍了机器翻译的发展历史, 目前主流的机器翻译方法及其分析, 专利文本机器翻译的现状及探索。在机器翻译发展的几十年里, 机器翻译取得了很大的进展, 特别是最近的十五年, 各种机器翻译技术不断出现。网络的兴起, 也给机器翻译提供了新的应用背景, 使得一些实用化的机器翻译系统出现在市场上, 如在线翻译, 网页、电子邮件的翻译等等。同时, 机器翻译依然面临很多的问题, 如: 知识获取问题、歧义问题以及如何更好的认识语言特征规律等。目前的翻译系统的翻译级为句子, 尤其是专利文本, 在切分句子成分缺失较为严重。有人提议考虑篇章级的文献翻译, 利用源语言, 实现自动生成目标语言。已有多家研究机构在朝着这个方向尝试, 我们期待着有更大的突破和技术成果带给社会。

参考文献

- [1] Brown P F, Cocke J, Della S A, et al. A Statistical Approach to Machine Translation. Computational Linguistics, 1990, 16(2): 79-85.
- [2] Nagao M. A Framework of a mechanical translation between Japanese and English by analogy principle. Artificial and Human Intelligence, 1984: 173-180.
- [3] Kay M. Unification grammar: [Technical Report]. Xerox Palo Alto Research Center, 1983.
- [4] Charniak E. Statistical Language Learning. Cambridge, MA: MIT Press, 1993.
- [5] Brown P F, Della S A, Robert L M, et al. The Mathematics of Statistical Machine Translation: Parameters Estimation. Computational Linguistics, 1993, 19: 263-311.
- [6] Nieben, Vogel S, Ney H, et al. A DP Based Search Algorithm for Statistical Machine Translation, ACL36/ COLING17, 1998. 960-967.
- [7] Wilks Y. Corpora and Machine Translation. In: Proc. of Machine Translation Summit IV. Kobe, Japan, 1993. 137-146.
- [8] Josef F, Ney H. Discriminative Training And Maximum Entropy Models for Statistical Machine Translation. In: proc. of the 40th ACL, Philadelphia, 2002. 295-302.
- [9] William A G, Church K W. A Program For Aligning Sentences in Bilingual Corpora. In: proc. of the 29th ACL, 1991. 177-184.
- [10] Martin K. Text Translation Alignment Computational Linguistics, 1993, 19.
- [11] Federica, Riccardo, Paolo. Searching Similar Sentences for EBMT, SEBD' 02, Italia, 2002.
- [12] Satoshi, Francis, Yamato. A Hybrid Rule and Example-based method for Machine Translation. NLPRS-97, 1997.
- [13] MaTait K. Translation Pattern Extraction and Recombination for EBMT: [PH.D thesis]. UMIST, 2001.
- [14] Casillas J, Cordon O, et al. Genetic feature selection in a fuzzy rule-based classification system learning process for high dimensional problems. Information Sciences, 2001, 136(1-4): 135-157.
- [15] Kudo M, Jack S. Comparison of algorithms that select features for pattern classifiers. Pattern Recognition, 2000, 33: 25-41.
- [16] 王永生. 英汉机器翻译系统中的词性标注研究[J]. 计算机工程与应用, 2010, 46(20).
- [17] Xiaoyi Ma. Champollion: A Robust Parallel Text Sentence Aligner. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC). Genova, Italy, 2006.
- [18] Xin Yu, Jian Wu, et al. Research and realization of Dictionary-based Chinese-Tibetan Sentence Alignment. Journal of Chinese Information Processing. 2011, 9, 25(4): 57-62.
- [19] 李鹏. 高性能的中英文句子对齐算法及其应用[D]. 清华大学出版社, 2009.
- [20] 黄曾阳. HNC(概念层次网络)理论[M]. 北京: 清华大学出版社, 1998.
- [21] 晋耀红. 基于概念层次网络(HNC)理论的语言理解技术及其应用[M]. 北京: 科学出版社, 2006.
- [22] 王凌著. 智能优化算法及其应用[M]. 清华大学出版社, 2001.
- [23] 黄曾阳. HNC的发展和未来[J]. 汉语学报, 2001(3): 46-64.
- [24] 李颖, 池毓焕. 基于机器翻译的原型句蜕及其包装研究[J]. 装甲兵工程学院学报, 2003(3).
- [25] 池毓焕, 李颖. 汉语原型句蜕向英语多元逻辑组合的变换[J]. 装甲兵工程学院学报, 2004, 3, 18(1).
- [26] 李颖. 面向汉英机器翻译的包装句蜕处理[J]. 计算机应用, 2004(06).
- [27] 覃张华, 王景中. 基于HNC理论的短文本语境框架提取实现[J]. 北京工商大学学报(自然科学版), 2007(05).
- [28] 张克亮. 基于HNC理论的汉英机器翻译策略研究[J]. 解放军外国语学院学报, 2003(05).
- [29] 张运良, 张全. 基于HNC理论的语义相关度计算方法[J]. 计算机工程与应用, 2005, 34.
- [30] 杜玲莉, 纪再祥等. 基于HNC理论和自动文摘的计算机辅助英汉科技翻译模式[J]. 湖北工程学院学报, 2012(06).
- [31] 李文, 洪亲等. 基于n-gram的字符串分割技术的算法实现[J]. 计算机与现代化, 2010(9).
- [32] 刘波, 王凌等. 微粒群优化算法研究进展[J]. 化工自动化及仪表, 2005(23).
- [33] 武朝华, 汪镭. 微粒群优化算法综述[J]. 软件设计开发, 2009: 1270-1274.
- [34] 肖健, 徐建等. 英中可比语料库中多次表达自动提取与对齐[J]. 计算机工程与应用, 2010(31).
- [35] 刘群. 机器翻译研究新进展[J]. 当代语言学报, 2009(2).
- [36] 戴新宇, 尹存燕等. 机器翻译研究现状与展望[J]. 计算机科学, 2004(11).
- [37] 熊文, 蒋宏飞等. 专利文本统计机器翻译中使用微粒群优化改进Champollion. 第九届机器翻译大会, 2013, 10.