机器翻译研究现状与展望*)

戴新宇　尹存燕　陈家骏　郑国梁
(南京大学计算机软件新技术国家重点实验室　南京210093)
(南京大学计算机科学与技术系　南京210093)

摘　要　本文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，详细介绍和分析了基于混合策略的机器翻译方法，对统计以及机器学习方法在机器翻译中的应用进行了描述；论文还介绍了当前机器翻译评测技术；最后对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，混合策略，机器学习

Machine Translation: Past, Present, Future

DAI Xin-Yu　YIN Cun-Yan　CHEN Jia-Jun　ZHENG Guo-Liang
(State Key Laboratory for Novel Software Technology, Department of Computer Science & Technology, Nanjing University, Nanjing 210093)

Abstract　This paper firstly presents the history of machine translation, and introduces some classical paradigms of machine translation: RBMT, SBMT and EBMT. Secondly, we introduce the recent research on machine translation, and describe the hybrid strategies on machine translation in detail, and discuss the applications of machine learning for machine translation. We also analyze the current techniques about evaluation on machine translation. Finally, we draw a conclusion and prospect on the research of machine translation.

Keywords　Machine translation, RBMT, SBMT, EBMT, HSBMT, Machine learning

1　引言

近年来，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、目标等也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2部分介绍几种经典的机器翻译方法；第3部分给出近期机器翻译的研究现状，介绍基于混合策略的机器翻译方法，对统计机器学习在机器翻译中的应用进行总结；第4部分讨论当前机器翻译的评测系统；最后，对机器翻译研究进行总结和展望。

2　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地产生。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[1]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[2]，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。

*)本论文工作得到863课题资助(编号：2001AA114102，2002AA117010-04)。戴新宇　博士生，主要研究自然语言处理；尹存燕　助教，主要研究自然语言处理；陈家骏　教授，博士生导师，主要研究自然语言处理、软件工程；郑国梁　教授，博士生导师，主要研究软件工程。

2.1 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:规则可以很准确地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/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、如何更好地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[3]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地来自于标注语料库。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[4]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。

2.2 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(如 Isabelle 1987所做的天气预报翻译)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,这些问题我们将在本文的3.2部分讨论。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[1]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。解码过程用图1来表示。

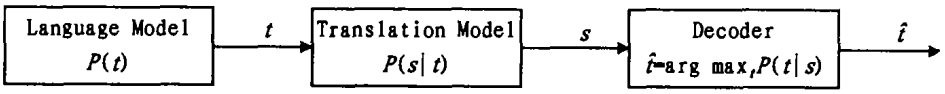

图1 基于统计的机器翻译示意图

基于统计的机器翻译也可以用下面这个公式来说明:

$$best\text{-}translation\ T = argmax_T\ fluency(T)\ faithfulness(T, S)$$

其中,T 表示目标语言句,S 表示源语言句。$fluency(T)$相当于语言模型,它反映目标语言句子的质量,$faithfulness(T,S)$相当于翻译模型,表示从源语言到目标语言的翻译质量。从上面的公式可以看出,翻译的过程其实也是一个寻求最优翻译结果的过程。

因此,基于统计的机器翻译的关键首先是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采用基于语法的语言模型[27],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[5]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[9,10],以及找到比较好的平滑算法[26]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[6]。

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[7]将一些浅层的词汇信息结合到统计模型当中去,Franz Josef[8]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，这从某种意义上说也是基于规则的机器翻译方法和基于统计机器翻译方法的融合，或者说是统计方法在处理语言学问题上的延伸，体现了下面我们要介绍的基于混合策略的机器翻译方法研究的必要性。

2.3 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[2]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[11]在2002年的一篇文章中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[9]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[10]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[12]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[13]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。

3 基于混合策略的机器翻译方法研究

根据上面对几种机器翻译方法的介绍，可以看出，不同的机器翻译方法有各自的优势和局限性，基于规则的方法优势在于可以很准确地描述语言特征规律，符合理性思维；而基于统计的方法可以缓解知识获取的瓶颈问题。因此如何发挥各种机器翻译方法的优势，把各种方法有效地结合起来，从而改善机器翻译系统的性能，成为我们研究的重点。在现阶段，把传统的基于规则的方法和基于语料库的方法相结合，已经成为机器翻译研究的主流。在结合策略上，一种策略是进行多引擎的机器翻译，结合各种机器方法，充分发挥各种机器翻译方法的优点，生成高质量的译文。还有一种策略是在基于规则的机器翻译方法中，利用语料库资源，采用统计学和机器学习的技术解决机器翻译中的局部问题，进而提高整体性能。

3.1 基于混合策略的机器翻译研究(Hybrid Strategies Machine Translation)

在基于单一方法的机器翻译中，不管采用哪种方法，总是不能取得理想的效果，究其原因，主要是因为各种方法固有的问题造成的，例如基于统计的机器翻译方法采用的 n 元语法模型无法解决长距离依赖问题，以及语料库的标注体系、语料库的数据稀疏等等问题，而基于规则的方法很难覆盖所有的语言现象，并且在对源语言和目标语言分析生成过程中的歧义问题解决得不够理想。

于是，基于混合策略的机器翻译方法成为研究的焦点，基于混合策略的方法充分利用各种机器翻译方法的优势，避免各种方法的不足，做到翻译结果的最优化，从而达到提高翻译系统性能的目的。

Frederking[14]首先提出了多引擎的机器翻译的思想，并且利用多引擎机器翻译的思想设计了 Pangloss Mark III 机器翻译系统，该系统结合了基于规则的机器翻译方法、基于实例的机器翻译方法和基于词汇转换的机器翻译方法，主要设计思想是：接收输入句，用多个翻译引擎并行翻译句子片断(短语和词)，将每个翻译单元存储在一个 chart 中，并根据某种评分标准给每个翻译单元打分，最后利用动态规划算法给出最优翻译结果。合适的评分标准关系到能否选择到最佳的翻译结果，Pangloss 系统采用人工评估和启发式评估方法相结合对翻译结果进行评分，其后 Brown[28]在多引擎的机器翻译系统中加入统计模型，利用 N-Gram 对候选结果进行选择，减少了 Pangloss 系统中的对翻译结果评估的人工参与。Nomoto[25]则用预测统计模型指导评估选择多引擎翻译结果。

Satoshi[12]的文章介绍了基于规则和基于实例相结合的日英机器翻译系统，首先分析了基于规则和基于实例方法的优缺点，提出了两种翻译方法结合的翻译方法，其具体算法是：首先查找与输入句相似的源语言候选句集，如果句集为空，则转向基于规则的翻译系统处理，生成目标语；如果不为空，则对候选句集中的句子按照和输入句的相似度进行排序；排序后按照基于词汇共现的方法聚集出目标语候选句，然后根据源语言句集和对应的目标语句集找出最优的双语句对；最后比较输入句和最优句对中源语言句进行比较，用规则对目标语句子进行替代、重排序等操作，最终生成目标语。以句对做指导，生成目标语这个阶段主要是规则指导的。

在基于混合策略的机器翻译系统中，基于规则的方法一

般用于对源语言进行语言分析，而统计和实例的方法则对语言资源进行自动获取以及如何利用语言资源处理方面起着重要的作用。

3.2 统计和机器学习方法在机器翻译中的应用

根据上文对基于混合策略机器翻译系统的介绍，机器翻译面临着词性标注、句法分析、消歧、目标语生成、语言知识自动获取、标注语料的构造、双语语料对齐、模型参数估计、平滑数据稀疏等等问题，对于中文，还有分词的问题。在语言分析生成以及语言知识库构造过程中，机器学习的技术正在被广泛运用，并且对机器翻译提供了巨大的帮助。

应用于词性标注的机器学习方法有：隐马尔可夫方法[15]、基于错误驱动的转换方法[16]、决策树方法[17]、最大熵方法[18]等。而K近邻法[19]、朴素的贝叶斯方法则被应用于词义消歧。概率上下文无关文法[4]、决策树模型[20]、最大熵模型[21]则被用于句法分析。J. Tenni[22]、Charniak[4]、Brown[5]等在统计机器翻译模型中，运用机器学习方法获取语言规则，进行参数估计。概率论、互信息等则在语料库对齐的研究中发挥着重要的作用。

经典的机器学习方法，如 Na ïve Bayes、Neural Network、KNN、Decision-Trees、Rule Learning based on error-driven transformation、HMM、Maximum Entropy，以及概率统计学方法，被应用于机器翻译处理的各个阶段，它们能够处理自然语言的分类，歧义消减，预测推理等等问题。如何更好地利用这些方法，以及寻求新的学习算法来解决机器翻译中的问题已经成为研究热点。

统计以及机器学习的方法实质是利用统计数据和机器学习算法在知识源的指导下解决机器翻译中遇到的问题。而知识源来自各种标注语料库，例如，词性标注需要标注词性的语料库，句法分析需要句库。构建大规模的标注语料库是统计机器学习方法研究的基础。当然，现在人们也尝试着在小规模的标注语料库的指导下，用某些机器学习方法解决问题，例如 Bootstrapping、Co-Training、Conditional Random Fields 等等。

4 机器翻译评测技术分析

机器翻译的评测技术对机器翻译的研究和发展具有重要的意义，有了评测，我们才可以评价各种机器翻译方法的优劣，并且为机器翻译方法的改进提高知道。我们也可以通过评测评价机器翻译系统的性能，因此机器翻译评测技术同样是机器翻译研究领域的一个重要课题。

机器翻译的评测方式现阶段有两种：人工评测和自动评测。其中人工评测包括：译文用词是否准确；是否忠于原文的语法语义；以及译全率、流利度等等。人工评测主要是以译文的用词、语法、语义等为标准来判断译文是否忠实于原文，评测比较精准，同时具有主观性。

目前机器翻译的自动评测标准主要有 IBM 的 BLEU 标准[23]和 NIST 提出的 NIST 标准[24]。

BLEU 方法是一种基于 N-Gram 的自动评测方法，它的主要思想是统计共现的 N 元词的个数，即比较在参考译文中和系统的翻译结果中共现的 N 元词的个数，一元词的共现代表了翻译的忠实度，表明原文中有多少词被翻译过来。二元以上的共现反映了译文的流利度。BLEU 方法还引入长度惩罚因子，考虑了译文长度对翻译质量的影响。IBM 的研究报告说明 BLEU 的评测效果和人工评测效果比较一致。

NIST 评测技术是 BLEU 评测技术的一个改进，每个 N 元词被赋予权重，一个 N 元词在参考译文中出现的次数越多，则这个 N 元词包含了更多的信息量，权重更高。同时 NIST 采用算术平均而不是几何平均，一元词的共现结果对评分结果影响更大，使得评价更能反映忠实度。NIST 还改进了长度惩罚因子，减少了译文句子长度对评分结果的影响。

目前的机器翻译评测基本上采取人工评测和自动评测相结合的方法。

总结与展望 本文系统的介绍了机器翻译的研究方法，对基于规则、基于统计、基于实例的机器翻译方法进行了分析和评价；同时介绍了基于混合策略的机器翻译方法，对统计机器学习方法在机器翻译中的应用进行了总结；分析了机器翻译评测技术。

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，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地认识语言特征规律等。

直到现在，ALPAC 提出的报告仍然有值得我们反思的地方。我们应该重新审视机器翻译的终结目标，即全自动获得高质量的翻译，由于语言之间的差别涉及到语言文化上、社会背景上的差异，完全意义上的全自动翻译并不是一个现实的目标，我们需要对目标进行重新的定位。现阶段的机器翻译，应该起着辅助人工的作用，在某些受限领域寻求突破，机器翻译应该作为一个工具，而不是一个独立的系统，去给自然语言处理的其它任务服务，比如数据挖掘、信息检索等，去寻求新的应用场景，比如移动电话的短信翻译、电视的字幕翻译、各种信息的多语发布等等。

机器翻译始终是一项有挑战性的工作，值得我们锲而不舍地去深入研究。

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: 121～142

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

（下转第184页）

Trees. Wadsforth International Group,1984
13 Cover T M. The best two independent measurements are not the two best. IEEE Trans. Syst. Man Cybern. ,1974,4(2):116～117
14 John G,Kohavi R,Pfleger K. Irrelevant features and the subset selection problem. The Eleventh International Conference on Machine Learning,1994. 121～129
15 Aha D W,Bankert R L. Feature selection for case-based classification of cloud types. In:working notes of the AAAI94 workshop on case-based reasoning,1994. 106～112
16 Provan G M,Singh M. Learning bayesian networks using feature selection. In: Proc. 5th Intern. Workshop on AI and Statistics, 1995. 450～456
17 Inza I,Larraaga P,Sierra B. Feature subset selection by Bayesian networks based on optimization. Artificial Intelligence,2001,123 (1-2):157～184
18 Caruana R A,Freitag D. Greedy attribute selection. The Eleventh intl. conf. on machine learning,1994. 28～36
19 Moore A W,Lee M S. Efficient algorithms for minimizing cross validation error. In: The Eleventh Intl. Conf. on Machine Learning,1994. 190～198
20 Davies S,Russl S. Np-completeness of searches for smallest possible feature sets. In:Proc. of the AAAI Fall 94 Symposium on Relevance,1994. 37～39
21 Liu H,Motoda H. Feature Selection for Knowledge Discovery and Data Mining. Boston: Kluwer Academic Publishers,1998
22 Natendra P M,Fakunaga K. A branch and bound algorithm for feature subset selection. IEEE Trans. Comput. ,1977. 917～922
23 Pudil P,Novovicova J,Kittler J. Floating search methods in feature selection. Pattern Recognition Letters, 1994,15(11):1119～1125
24 Somol P,Pudil P,Novovicova J,Pacik P. Adaptive floating search methods in feature selection. Pattern Recogniton Letters,1999,20 (11－13):1157～1163
25 Yang J,Vasant H. Feature subset selection using a genetic algorithm. IEEE Intelligent Systems,1998,13:44～49
26 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
27 Kudo M,Jack S. Comparison of algorithms that select features for pattern classifiers. Pattern Recogniton,2000,33:25～41
28 John G,Kohavi R,Pfleger K. Irrelevant features and the subset selection problem. In: The Eleventh Intl. Conf. on Machine Learning,1994. 121～129
29 Blum A L,Langley P. Selection of relevant features and examples in machine learning. Artificial Intelligence, 1997,97(2):245～271
30 Caruana R A,Freitag D. Ho v useful is relevance? Working notes of the AAAI Fall Symposium on Relevance. New Orleans,LA: AAAI Press,25～29
31 Pfahringer B. Compression-based feature subset selection. In:IJCAI-95 Workshop on data Engineering for Inductive Learning, 1995. 101～106
32 Liu H,Motoda H,Dash M. A monotonic measure for optimal feature selection. In: 10th European Conf. on Machine Learning, 1998. 101～106
33 Lee H M,Chen C M,et al. An efficient fuzzy classifier with feature selection based on fuzzy entropy. IEEE Trans. on systems and cybernetics-Part B: Cybernetics,2001,31(3):26～432
34 Weston J,Mukherjee S,et al. Feature selection for SVMS. In:Advances in Neural Information Processing Systems,2000,13:668～674
35 范劲松,方廷建.特征选择和提取要素的分析及其评价.计算机工程与应用,2001(13):95～99
36 Almullim H,Dietterich T G. Learning with many irrelevant features. In:Proc. Ninth National Conf. on Artificial Intelligence, 1991. 547～552
37 Hamming R W. Coding and information theory. Englewood Cliffs,NJ: Prentice-Hall,1986
38 Domingos P. The role of Occam's razor in knowledge discovery. Data Mining and knowledge Discovery,1999,3(4):409～425
39 Hall M A. Correlation-based feature selection for discrete and numeric class machine learning. In: The Seventeenth Intl. Conf. on Machine Learning,2000. 359～366
40 Kira K,Rendell L A. The feature selection problem: traditional methods and a new algorithm. In:Proc. of the Ninth National conf. on Artificial Intelligence,1992. 129～134
41 Kononenko I. Estimation attributes: analysis and extensions of RELIEF. In:Proc. of the 1994 European Conf. on Machine Learning,1994. 171～182
42 Marko R S,Kononenko I. Comprehensible interpretation of relief estimates. In:The Eighth Intl. Conf. on Machine Learning,2001. 443～440
43 Moore A W,Lee M S. Efficient algorithms for minimizing cross validation error. In:The Eleventh Intl. Conf. on Machine Learning,1994. 190～198
44 Dash M,Liu H. Feature selection for clustering. In:4th Pacific-Asia Conf on Knowledge Discovery and Data Mining,2000. 110～121
45 Sanmay D. Filters,Wrappers and a boosting-based hybrid for feature selection. In:Intl. conf. on Machine Learning,2001. 74～81

(上接第179页)

13 MaTait K. Translation Pattern Extraction and Recombination for EBMT:[PH. D thesis]. UMIST,2001
14 Frederking, Nirenburg. Three Heads are Better than One. In: proc. of ANLP-94,Germany,1994
15 Brant T. TNT - A Statistical Part-Of-Speech Tagger. In:proc. of the 6th Applied Natural Language Processing Conf. Seattle, Washington,USA,2000
16 Brill E. A Corpus-based Approach to Language Learning:[PH. D Thesis]. UPENN,1993
17 Daelemans W,Zavrel J,Berck P,et al. MBT: A Memory-Based Part of Speech Tagger Generator. In:Proc. of the Workshop on Very Large Corpora,Copenhagen,Denmark,1993
18 Ratnaparkhi A. A Maximum Entropy Model for Part-Of-Speech Tagging. In:proc. of EMNLP-96,Philadelphia,PA. 1996
19 Ng,Tou H,Lee H B. Integrating multiple knowledge sources to disambiguate word sense: An Example-Based Approach. In: the Proc. of the 34th ACL,1996. 40～47
20 Magerman D M. Statistical Decision-Tree Model s for Parsing. In: the Proc. of the 33rd ACL. Cambridge,MA,1995. 26～30
21 Ratnaparkhi A. A Linear Observed Time Statistical Parser Based on Maximum Entropy Models,In EMNLP2,Providence,RI. ,1997
22 Tenni J,Lehtola A,et al. Machine Learning of Language Translation Rules. In:IEEE Intl. Conf. on Systems,Man,and Cybernetics Tokyo International Forum,Tokyo,JAPAN,1999
23 Kishore,Salim,Todd,et al. Bleu: a Method for Automatic Evaluation of Machine Translation:[IBM Research Report]. 2001
24 utomatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics:[NIST Report]. 2001
25 Nomoto T. Predictive Models of Performance in Mulit-Engine Machine Translation,MT Summit IX,2003
26 Chen, Stanley F, Goodman J. An empirical study of smoothing techniques for language modeling:[Technical Report TR-10-98]. Center for Research in Computer Technology,Harvard University,1998
27 Charniak E,Knight K,Yamada K. Syntax-based Language Model for Statistical Machine Translation,MT Summit IX,2003
28 Brown R, Frederking. Applying statistical English Language Modelling to symbolic machine translation. TMI' 95, Leuven, Belgium,1995. 221～239