

基于判别式的藏语依存句法分析

华却才让^{1,2}, 赵海兴¹

(1. 青海师范大学藏文信息研究中心, 西宁 810008; 2. 陕西师范大学计算机科学学院, 西安 710062)

摘 要: 现有藏语句法体系复杂, 不利于藏文自然语言处理的应用。为此, 提出基于判别式的藏语依存句法分析方法, 采用感知机方法训练句法分析模型, CYK 自底向上算法解码生成最大生成树。实验结果表明, 在人工标注的测试集上, 句法分析正确率达到 81.2%, 可实际应用到藏语依存树库的构建和其他自然语言处理中。

关键词: 藏语依存句法; 句法标注规范; 最大生成树; 特征模板; 依存句法; 感知机

Tibetan Text Dependency Syntactic Analysis Based on Discriminant

Hua-que-cai-rang^{1,2}, ZHAO Hai-xing¹

(1. Tibetan Information Research Center, Qinghai Normal University, Xining 810008, China;

2. School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

【Abstract】 The existing Tibetan syntax system is complex, which is not conducive to the application of Tibetan natural language processing. So this paper describes an approach based on discriminant for analysis of Tibetan text dependency structure, where perceptron training method is used to training parsing model. And it also proposes a maximum spanning tree with CYK from the bottom-up algorithm for decoding. Experimental results show that, the method obtains acceptable score of 81.2% on manual test set. And it is applicable to Tibetan dependency library and other natural language processing.

【Key words】 Tibetan dependency syntax; syntax tagging specification; maximum-spanning tree; feature template; dependency syntax; perceptron

DOI: 10.3969/j.issn.1000-3428.2013.04.069

1 概述

藏语文法从土弥桑布札开始就已经建立一套完整的传统语法体系, 在 1 300 多年漫长的语言演变过程中, 虽然发生了不少变化, 但藏语语法的 3 项常用的语法手段: 语序、形态和作格特征基本没有变^[1], 并仍然符合现代藏语语法, 这使现代藏语形成了完整而独特的语法体系。

藏语句法体系复杂, 语序为主宾谓, 动词形态丰富, 格助词接续规范较多且复杂, 其用法受语法、语义语用等多种因素的制约, 而且藏语词性标注规范制订时间较晚^[2], 影响了藏语依存句法树库的构建和分析。作为藏文信息处理的一项基础性工作和自然语言处理的一项关键技术, 藏语句法分析问题的解决直接对信息检索、信息抽取、语义分析以及机器翻译等藏文自然语言处理应用产生推动作用。

在句法分析的研究中, 依存语法以其形式简洁、易于

标注、便于应用等优点, 已经受到研究人员的重视, 而且符合藏语格接续特征。目前, 主流语言英语、汉语等的句法分析已经有了比较系统和成熟的研究方法: (1)有监督的句法分析; (2)无监督的句法分析; (3)基于双语映射的句法分析^[3]。其中, 有监督的句法分析方法比较成熟, 有利于藏语依存句法分析。

以目前藏语言的研究现状来看, 除了项目组制订的依存规范^[4]和人工建立的 1 万句藏语句法树库外, 没有见到较成熟的句法理论研究和标准树库。本文利用人工制订藏语依存句法规范、构建依存树库, 分析和总结适合藏语依存句法的特征, 在训练树库上训练并建立句法模型, 最后实现一个基于判别式最大生成树藏语依存句法分析器。

2 依存句法标注规范

藏语绝大部分句子的中心词在句子尾部, 主语、宾语

基金项目: 国家自然科学基金资助项目(61063033, 61163018); 国家“973”计划前期研究专项基金资助项目(2010CB334708); 青海省科技基金资助项目(2011-Z-752)

作者简介: 华却才让(1976—), 男, 副教授、博士研究生, 主研方向: 计算语言学, 藏文信息处理; 赵海兴, 教授、博士生导师

收稿日期: 2012-04-16 **修回日期:** 2012-06-27 **E-mail:** cairanghuaque@yahoo.com.cn

和补语等都在中心词之前, 这使成分内部的依存具有复杂性, 很难区分成分间依存关系, 成了藏语依存句法标注要解决的首要问题, 为此研究了实用藏文文法之三十颂解析^[5]、藏语的格助词搭配规则、意义及依存关系。例如句子中出现主格时, 其前面一般为使动成分而后面部分为被动成分, 如图 1 所示, 即基本可确定主语和宾语。同时借鉴

了其他语言的句法依存研究经验^[6], 补充了部分必要的句法依存关系, 如方位结构的依存关系等。尽可能使制定的句法依存关系满足藏语的各种语法现象, 也考虑了依存关系过于细致和庞大, 会导致交互式句法分析器的鲁棒性、可操作性下降和统计数据稀疏等问题^[7]。结合藏语实际切分标注语料为例, 制定 33 类藏语句法依存关系, 如表 1 所示。

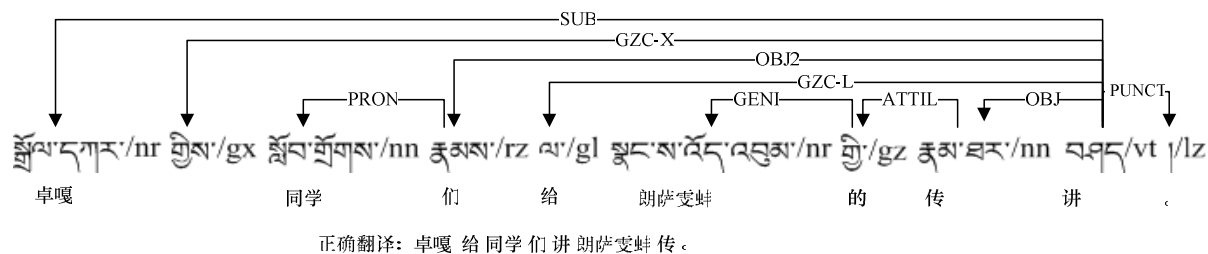


图 1 句子 S1 的句法依存分析树

表 1 藏语依存句法标注体系

序号	依存关系	序号	依存关系	序号	依存关系
1	谓语结构(TOP)	12	属格结构(GENI)	23	代词结构(PRON)
2	主谓结构(SUB)	13	标点符号依存(PUNCT)	24	从句依存结构(CSUB)
3	宾语结构(OBJ)	14	限量、时依存结构(QT)	25	关联词结构(GLC)
4	定语结构(ATTIL R)	15	连数结构(NUM)	26	词的重叠结构(DUP)
5	间接宾语结构(OBJ2)	16	方位宾语结构(FOBJ)	27	词缀结构(AFFIX)
6	状语机构(ZY)	17	连词结构(LNL R)	28	“=” 词结构(ANG)
7	结果补语结构(SCOM)	17	陈述词结构(YY)	29	量 数词附加语结构(QDIG)
8	状态补语结构(ZHB)	19	离合词结构(SUWLR)	30	叹 状态词结构(EXCL)
9	并列结构(COOR)	20	待述词结构(SCL R)	31	宾语提前结构(AOBJ)
10	助动词结构(AUXIW)	21	终结和疑问结构(FASQP)	32	存在结构(EXIU)
11	主 于 从格结构(GZC)	22	饰集词结构(OSL R)	33	特殊疑问词结构(SQW)

3 依存句法分析模型

现有的许多可以应用于依存句法分析的模型以解决藏文依存分析问题, 例如概率依存、最大生成树和状态转移等模型。由于最大生成树模型巧妙地将特征与权重评价模型相关联, 能获取更丰富的藏语文本特征, 以及词之间上下文接续特征, 因此选择最大生成树模型分析藏语依存句法。

若给定包含 N 个词的一个句子, 任意 2 个词之间都可能存在依存关系, 寻找最可能的依存树的任务是从 $N \times (N - 1)$ 种可能的依存边(无自环)完全图中寻找分数最大的树, 于是, 若 y 为句子 x 的依存树, 则 $(i, j) \in y$, $1 \leq i, j \leq |x|$ 且 $i \neq j$, 其中, (i, j) 表示句子 x 中的词 x_i 和 x_j 之间存在有向边, x_i 为词 x_j 的父节点; 根据 Eisner(1996)分解法可将依存树 y 的分数 $s(x, y)$ 表示为:

$$s(x, y) = \sum_{(i, j) \in y} s(i, j) = \sum_{(i, j) \in y} (\sum_k w_k \times f_k(i, j)) \quad (1)$$

其中, w_k 为特征 $f_k(i, j)$ 在训练样本 $T = \{(x_n, y_n)\}_{n=1}^L$ 上做了第 m 次迭代后得到的平均权重, 表示每个特征对依存边的贡献, 决定依存边的强弱。

由此依存最大生成树模型可以定义为:

$$\tilde{y} = \arg \max_y s(x, y) = \arg \max_y \sum_{(i, j) \in y} s(i, j) = \arg \max_y \sum_{(i, j) \in y} \exp(\sum_k w_k \times f_k(i, j)) \quad (2)$$

其中, y 为训练集中的标准依存树。下面将确定如何选择合适藏语依存句法的特征集, 学习最佳的权重, 从词性标注好的句子上找到权重分值最大的依存生成树。

4 特征选择及训练

4.1 特征选择

针对藏语依存句法分析问题, 选择有效的词法和句法结构特征是使用最大生成树模型需要解决的一个关键问题。在最大生成树模型^[8]中提出了每个特征是由词 i 和 j 及前后的词语和词性构成。为丰富句法特征信息, 文献[9]提出了词 i 和 j 之间的距离句法信息。这种方法解决了 2 个词之间顺序位置、相邻关系、是否动词居中以及 2 个词中间或左右是否存在标点符号等问题。

本文考察了影响藏语句法和词法结构的多种因素, 归纳为:

(1)藏语的主宾谓语序中, 动居句尾、物主词居前、位置词居后、形居名后。

(2)藏语动词心态分“三时一式”、使动与非使动、自主与不自主。

(3)作格中主格、属格、于格和从格等主要格词类得固定接续特征，且与形态动词相关。

为更好地依据藏语句子上下文特点确定模型的上下文激发环境，在藏语的词对依存的训练和解码中合成了以上 2 种句法特征生成模板，定义了 62 个藏语依存句法分析特征模板，具体的特征模板如表 2 所示。

表 2 藏语依存词对分类特征模板

特征分类	特征		
一元特征	$p\text{-word}, p\text{-pos}$	$p\text{-word}$	$p\text{-pos}$
	$c\text{-word}, c\text{-pos}$	$c\text{-word}$	$c\text{-pos}$
二元特征	$p\text{-word}, p\text{-pos}, c\text{-word}, c\text{-pos}$	$p\text{-word}, p\text{-pos}, c\text{-word}$	$p\text{-word}, p\text{-pos}, c\text{-pos}$
	$p\text{-word}, c\text{-word}, c\text{-pos}$	$p\text{-pos}, c\text{-word}, c\text{-pos}$	$p\text{-word}, c\text{-word}$
	$p\text{-pos}, c\text{-pos}$	$p\text{-pos}, c\text{-word}$	$p\text{-word}, c\text{-pos}$
词对及 左右词性特征	$p\text{-pos-1}, p\text{-pos}, c\text{-pos-1}, c\text{-pos}$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos}, c\text{-pos+1}$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos-1}, c\text{-pos}$
	$p\text{-pos-1}, p\text{-pos}, c\text{-pos}, c\text{-pos+1}$	$p\text{-pos-1}, p\text{-pos}, c\text{-pos-1}$	$p\text{-pos-1}, p\text{-pos}, c\text{-pos+1}$
	$p\text{-pos}, p\text{-pos+1}, c\text{-pos-1}$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos+1}$	$p\text{-pos-1}, c\text{-pos-1}, c\text{-pos}$
	$p\text{-pos-1}, c\text{-pos+1}, c\text{-pos}$	$p\text{-pos+1}, c\text{-pos}, c\text{-pos-1}$	$p\text{-pos+1}, c\text{-pos}, c\text{-pos+1}$
	$p\text{-pos}, c\text{-pos}, c\text{-pos-1}$	$p\text{-pos}, c\text{-pos}, c\text{-pos+1}$	$p\text{-pos}, p\text{-pos-1}, c\text{-pos}$
	$p\text{-pos}, p\text{-pos+1}, c\text{-pos}$		
距离特征	$p\text{-word}, d$	$p\text{-pos}, d$	$p\text{-word}, p\text{-pos}, d$
	$c\text{-word}, d$	$c\text{-pos}, d$	$c\text{-word}, c\text{-pos}, d$
	$p\text{-word}, p\text{-pos}, c\text{-word}, c\text{-pos}, d$	$p\text{-word}, p\text{-pos}, c\text{-word}, d$	$p\text{-word}, p\text{-pos}, c\text{-pos}, d$
	$p\text{-word}, c\text{-word}, c\text{-pos}, d$	$p\text{-pos}, c\text{-word}, c\text{-pos}, d$	$p\text{-word}, c\text{-word}, d$
	$p\text{-pos}, c\text{-pos}, d$	$p\text{-word}, c\text{-pos}, d$	$p\text{-pos}, c\text{-word}, d$
	$p\text{-pos-1}, p\text{-pos}, c\text{-pos-1}, c\text{-pos}, d$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos}, c\text{-pos+1}, d$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos-1}, c\text{-pos}, d$
	$p\text{-pos-1}, p\text{-pos}, c\text{-pos}, c\text{-pos+1}, d$	$p\text{-pos-1}, p\text{-pos}, c\text{-pos-1}, d$	$p\text{-pos-1}, p\text{-pos}, c\text{-pos+1}, d$
	$p\text{-pos}, p\text{-pos+1}, c\text{-pos-1}, d$	$p\text{-pos}, p\text{-pos+1}, c\text{-pos+1}, d$	$p\text{-pos-1}, c\text{-pos-1}, c\text{-pos}, d$
	$p\text{-pos-1}, c\text{-pos}, c\text{-pos+1}, d$	$p\text{-pos+1}, c\text{-pos}, c\text{-pos-1}, d$	$p\text{-pos+1}, c\text{-pos}, c\text{-pos+1}, d$
	$p\text{-pos}, c\text{-pos}, c\text{-pos-1}, d$	$p\text{-pos}, c\text{-pos}, c\text{-pos+1}, d$	$p\text{-pos}, p\text{-pos-1}, c\text{-pos}, d$
	$p\text{-pos}, p\text{-pos+1}, c\text{-pos}, d$		

藏语依存词对分类特征模板内容分 4 类：

(1)一元特征。定义为父结点或子结点(单个词)的特征信息构成。

(2)二元特征。由父子结点共同的特征信息构成。

(3)词对左右词性特征。考虑到更好地抽取到藏语格助词的搭配规律而补充了此特征信息。

(4)距离特征。词对间包括其他词(结点)时的依存关系的特征信息。

在表 2 中， $p\text{-word}$ 表示依存树中父结点词； $p\text{-pos}$ 表示父结点的词性； $c\text{-word}$ 表示依存树中子结点词； $c\text{-pos}$ 表示子结点的词性； $p\text{-pos-1}$ 表示父结点左边的词性； $c\text{-pos+1}$ 表示结点右边的词性； d 表示词对间所包含其他词(依存结点)个数。当 D 的值为负数时，表示句法树中抽出词对的父结点在子结点的左侧；而当 D 的值为整数时，表示句法树中抽出词对的父结点在子结点的右侧。

4.2 特征训练

对于有监督的判别式依存模型，一般在英汉等系统中运用在线的训练方法，如感知机、MIRA^[10]和最大熵等。为有效学习藏语依存关系，本文采用在线感知机方法训练权重：

$$\begin{aligned} &\min \|w\| \\ &\text{s.t. } s(x, y) - s(s, y') \geq C(y, y') \\ &\forall (x, y) \in T, y' \in dt(x) \end{aligned}$$

其中， $C(y, y')$ 表示在线模型训练时解码最佳生成依存树 y' 的损失值；而 y 为当前标准依存树，则损失值 $C(y, y')$ 计算为最佳生成依存树 y' 中依存对头结点的错误个数， y' 的最大损失值为句子 x 的长度。目标是通过特征训练解码出最佳依存生成树，使得损失值最小，趋于 0。训练算法见算法 1，其中， $FITINST(inst[n])$ 函数对当前实例在在线特征模型基础上解码生成对最佳依存树 y' ，拟合过程将利用生成 y' 的依存特征集，与此实例的参考依存(标准实例)特征集进行拟合操作，操作时给每个参考特征分值加 1.0，最好(best)依存特征集的每个特征在前者(参考依存)生成的句子的特征集中存在，则相应分值要减 1.0；不存在，则加该特征并赋分值-1.0(等于减)。最后将实例的每个拟合好的扩展特征集并入总特征模型。

算法 1 特征抽取及权重训练

Training data $T = \{(x_n, y_n)\}_{n=1}^L$
 $w=0; v=0;$
for $r \leftarrow 1..R$ do

```

for n ← 1..L do
    GENFEATS(inst[n]) ▷ extracted features according
template features
    FITINST(inst[n]) ▷ fitting features w(i) according
bestDeps and refDeps
    w(i+1) = update w(i)
    v = v + w(i+1)
Output: w = v / (R * L)

```

5 句法解码

从式(2)可以看出, 解码算法是寻找权重最大的依存树, 本文利用 Chu-Liu/Edmonds 的最大生成树寻找最佳依存树^[11]。为有效分析依存关系, 采用了 CYK 自底向上的动态规划算法, 见算法 2。其中, $V[i,j]$ 包含跨度 $[i,j]$ 的句法分析片段, 当跨度为 0 时推导出词本身分值, 而当跨度在 1 和 $|x|$ 之间时, 将整个跨度依次序分解为不同子跨度的组合, 抽取子跨度按左右依存组合生成推导分支, 期间利用词图分数计算新分支的分数并且新分支入堆。

实际实现时采用了立方体剪枝策略^[12], 通过循环依次出堆时, 将判断当前跨度中此分支的存在与否及分数等确定是否存入, 若分数小于等于已存在的分支, 则被丢弃称为剪枝; 此步之后还要判断子跨度中可能存在多个平行子推导分支, 若存在且没有访问过的平行子推导分支, 则要加入堆。通过循环出堆, 同样要做立方体剪枝操作, 最终使堆空, 生成子跨度 $[i,j]$ 最大分值推导。用函数 $EVAL(d)$ 计算词图中某个距离为 d 的依存词对的所有特征的权重之和。

算法 2 依存分析算法

```

Input: sentence x to be parsed
for  $\langle i,j \rangle \subseteq \langle 1, |X| \rangle$  in topological order do
    buf ←  $\emptyset$ 
    for  $k \leftarrow i..j-1$  do ▷ all partitions
        for  $l \in V[i,k]$  and  $r \in V[k+1,j]$  do
            insert DERIV(l,r) into buf
            insert DERIV(r,l) into buf
     $V[i,j] \leftarrow$  top K derivations of buf
Output: the best derivation of  $V[1, |X|]$ 
function DERIV(p,c)
     $d \leftarrow p \cup c \cup \{(p\text{-root}, c\text{-root})\}$  ▷ new derivation
     $d\text{.evl} \leftarrow EVAL(d)$  ▷ evaluation function
    return d

```

6 实验与分析

实验采用依据藏语依存规范人工标注的藏语依存树库, 训练语料规模为 1 万句, 测试语料规模为 300 句, 测试语料平均长度小于 17 个藏语词, 对应的参考依存语料规模为 300 句。使用青海师范大学藏文智能信息处理研究中

心研制的班智达藏语词法分析器, 对藏语语料进行词法分析。采用 MIRA 拓展拟合处理方法开发了判别式藏语依存模型训练工具, 在训练过程中, 训练轮数为 10 次, 在线训练和解码时立方体剪枝的堆空间小于 20。从训练语料训练出的特征模型规模为 2 159 562 条。

为了客观评价本文提出的判别式藏文最大生成树分析方法的性能, 依据本文提出的方法, 实现了一个藏语依存句法分析系统, 以依存关系正确率($depP$)、中心词正确率($headP$)和整句完全依存正确率($allP$)为指标, 对系统的藏语依存分析结果进行评价, 如图 2 所示。相关计算公式如下:

$$depP = \frac{depOk}{depCnt} \quad (3)$$

$$headP = \frac{headOk}{senCnt} \quad (4)$$

$$allP = \frac{allOk}{senCnt} \quad (5)$$

其中, $depOk$ 为系统正确识别的依存关系个数; $enCnt$ 为测试语料的藏语句数; $depCnt$ 为对应测试语料的依存参考语料中的依存关系对个数; $headOk$ 为系统正确识别的中心词个数; $allOk$ 为整句完全识别正确的依存树数目。经测试语料测试集做了实验, 依存关系正确率($depP$)为 81.2%, 中心词识别正确率($headP$)为 87.49%, 整句完全依存正确率($allP$)达 34.58%。实验结果相对于当前公布的英语和汉语等语言的依存分析正确率偏低。经人工分析: (1)复杂句型的分析正确率低, 包括从句结构, 长度大于 12 个词的句子等; (2)训练语料存在部分标注不一致。(3)由于训练集侧重于日常用语和政府文献, 测试语料中特殊文献类的分析正确率也偏低, 如医学、生物等。

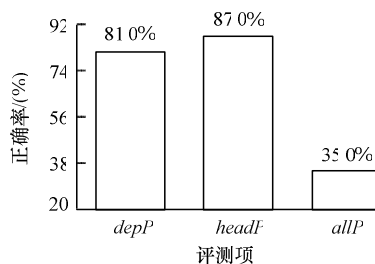


图2 系统对测试语料的各项评价指标

7 结束语

本文提出基于感知机的藏语依存句法分析方法, 通过分析藏文语料和藏语语法规则, 制订藏语依存句法标注规范, 构建规模为 1 万多句的依存树库。在此基础上选择适合藏语句法分析的特征模板, 用感知机方法训练的特征权重计算句子中词对间的依存关系, 最终选择分值最大的依存树。实验结果表明, 本文提出的藏语依存句法分析方法能够较好地解决目前藏语句法分析的需求和依存句法树库的构建问题。

下一步工作将藏语依存句法分析扩充为依存边的标注,这对依存分析质量的提高以及藏语文本语义分析具有重要的意义。此外,将尝试基于机器学习的半监督藏语依存句法分析方法,比较不同方法的优劣。

参考文献

- [1] 胡书津. 简明藏文文法[M]. 昆明: 云南民族出版社, 1988.
- [2] 才让加. 藏语语料库词语分类体系及标记集研究[J]. 中文信息学报, 2009, 23(4): 146-148.
- [3] 刘 凯, 乌日力嘎. 基于双语约束的蒙古语无监督依存分析[J]. 中文信息学报, 2011, 25(4): 25-29.
- [4] 华却才让, 赵海兴. 现代藏语依存句法标注初探[C]//第十二届全国少数民族语言文字信息处理学术研讨会论文集. 呼和浩特: [出版者不详], 2011.
- [5] 格桑居冕. 实用藏文文法[M]. 成都: 四川民族出版社, 1987.
- [6] 周 明, 黄昌宁. 面向语料库标注的汉语依存体系的探讨[J]. 中文信息学报, 1994, 8(3): 35-51.
- [7] 周 强. 汉语句法树库标注体系[J]. 中文信息学报, 2004, 18(4): 1-8.
- [8] McDonald R, Crammer K, Pereira F. Online Large-margin Training of Dependency Parsers[C]//Proc. of the 43rd Annual Meeting on Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2005: 91-98.
- [9] Collins M. A New Statistical Parser Based on Bigram Lexical Dependencies[C]//Proc. of the 34th Annual Meeting on Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 1996: 184-191.
- [10] McDonald R, Pereira F, Ribarov K, et al. Non-projective Dependency Parsing Using Spanning Tree Algorithms[C]//Proc. of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2005: 523-530.
- [11] Jiang Wenbin, Liu Qun. Dependency Parsing and Projection Based on Word Pair Classification[C]//Proc. of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden: [s. n.], 2010: 12-20.
- [12] Huang Liang, Chiang D. Better k-best Parsing[C]//Proc. of the 9th International Workshop on Parsing Technologies. Vancouver, Canada: [s. n.], 2005: 53-64.

编辑 顾逸斐

(上接第 299 页)

参考文献

- [1] 吴恩华. 图形处理器用于通用计算的技术、现状及其挑战[J]. 软件学报, 2004, 15(10): 1493-1504.
- [2] Barbay J, López-Ortiz A, Salinger A, et al. An Experimental Investigation of Set Intersection Algorithms for Text Searching[J]. Journal of Experimental Algorithmics, 2009, 14(7): 7-24.
- [3] Wu Di, Zhang Fan, Ao Naiyong, et al. Efficient Lists Intersection by CPU-GPU Cooperative Computing[C]//Proc. of International Symposium on Parallel & Distributed Processing. Atlanta, USA: IEEE Press, 2010.
- [4] Ao Naiyong, Zhang Fan, Wu Di, et al. Efficient Parallel Lists Intersection and Index Compression Algorithms Using Graphics Processing Units[J]. Proceedings of the VLDB Endowment, 2011, 4(8): 470-481.
- [5] Bentley J L, Yao A C C. An Almost Optimal Algorithm for Unbounded Searching[J]. Information Processing Letters, 1976, 5(3): 82-87.
- [6] Krauthgamer R, Mehta A, Raman V, et al. Greedy List Intersection[C]//Proc. of the 24th International Conference on Data Engineering. Washington D. C., USA: IEEE Computer Society, 2008: 1033-1042.
- [7] Wu Di, Zhang Fan, Ao Naiyong, et al. A Batched GPU Algorithm for Set Intersection[C]//Proc. of the 10th International Symposium on Pervasive Systems, Algorithms, and Networks. [S. l.]: IEEE Press, 2009: 752-756.
- [8] 陈 伟, 杜凌霄, 陈 红. 多核架构下的数据处理算法优化策略综述[J]. 计算机科学与探索, 2011, 5(12): 1057-1075.
- [9] Yang Canqun, Wang Feng, Du Yunfei, et al. Adaptive Optimization for Petascale Heterogeneous CPU/GPU Computing[C]//Proc. of International Conference on Cluster Computing. [S. l.]: IEEE Press, 2010.
- [10] 邹 岩, 杨志义, 张凯龙. CUDA 并行程序的内存访问优化技术研究[J]. 计算机测量与控制, 2009, 17(12): 2504-2506.

编辑 顾逸斐

