

基于正则表达式的藏文虚词“ཞིང་”类识别算法研究

更藏卓玛，安见才让

(青海民族大学计算机学院, 西宁 810007)

摘要:藏文虚词“ཨ”类是藏文语法中的非依他关联词,共有14个虚词。是藏文语法中的基础,并且是藏文信息处理的基础研究内容之一。本文结合藏文虚词添接规则和正则表达式对虚词“ཨ”类进行识别算法研究。以6本初中藏文教科书作为实验语料,通过实验得出藏文虚词“ཨ”类识别准确率达到99.94%。

关键词:识别;正则表达式;虚词“ᄃᆞ”类

doi : 10.3969/J.ISSN.1672-7274.2019.09.177

中图分类号: J292.14, TP31

文献标示码：A

文章编码: 1672-7274 (2019) 09-0214-02

藏文中共有 85 个虚词，它比实词少很多，但藏文虚词在藏文中的应用及其广泛，而且它在不同语境中扮演者不同的角色，因此出现了虚词的歧义问题，而因它的歧义性问题识别起来非常复杂，尤其是非依他关联词的识别。但对其的研究对藏文词和句子的处理起着重要的作用。因此，文章通过藏文虚词添加规则和正则表达式对藏文虚词“ཞེས་”类进行识别研究。

表1 虚词“ཞིང་”类添接表

后加字	ཤ་དབང་རྒྱལ།	ད་ན་མེད་དཔལ་དང་མཐུད་མེད།	ཤ
虚词“ཞིང”类	ཅིང་ཅིས་ཅིའོ་ཅིང་ཅིས།	ཞིང་ཞིས་ཞིའོ་ཞིང་ཞིས།	ཞིས་ཅིང་ཅིས་ཀྱིས་ཀྱི་ཅིང་ཅིས།

1 虚词“ཞིང་”类的正则表达式

藏文文本校对过程中校对藏文虚词的添接是一项很重要的工作,要完成这个工作首先要识别藏文虚词。

[illegible]

检测添接规则，并统计其次数。如图1所示。



图 1

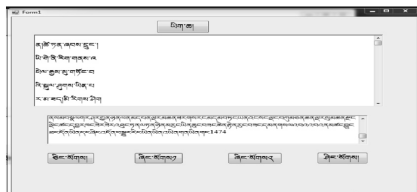
[illegible][illegible]

图2

(3) མཐའ་མེད་རྗེས་ས་ཁིང་ཞེས་ཞེའོ་ཞེ་ན་ཞེ་ག 的添接文法检测正则表达式

为:

[illegible]

图3

[illegible]

检测添接规则，并统计其次数。如图4所示。



图4

(5) 虚词“ཞིང་”类的识别算法流程图, 图5。

2 实验结果

六本初中藏文教科书做为实验语料,经试验得出在语料中“ཞེ”等14个字总共有9902,其中已识别出作为虚词的个数为3996,见表2。识别准确率为 $A = \frac{\text{正确结果数}}{\text{结果总数}} \times 100\% = \frac{3994}{3996} \times 100\% = 99.94\%$ 。

表2 “ཁོང་”等14个词语料中的数量统计

名称	数量	名称	数量	名称	数量	名称	数量	名称	数量
ཁོང་	1271	ཁོས་	774	ཅིག	1734	ཅོད་	0	ཅོན་	4
ཅིང་	521	ཅིས་	280	ཞིག	479	ཞོད་	0	ཞོན་	8
ཞིང་	863	ཞིག	3928	ཞོད་	0	ཞོན་	40		

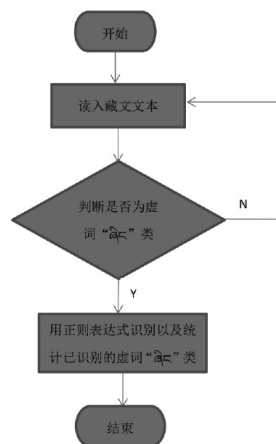


图5

表3 虚词“ཁོང་”类的歧义统计

名称	总数	为虚词总数	为虚词百分比
“ཁོང་”类	9902	3996	40.3%

3 结束语

本文结合藏文虚词添接规则和正则表达式对虚词“ཁོང་”类进行识别算法研究，并实现了藏文虚词“ཁོང་”类自动识别系统。根据实验结果统计，本系统中虚词“ཁོང་”类识别准确率达到99.94%。但在识别过程中难免会遇到具有歧义性的虚词，比如“སྒྲིག་ཅིག་ཅོད་”中的“ཅིག”是具有实意的词，而不是虚词。但它完全符合虚词“ཅིག”的添接规则。所以这种词在识别过程中也会识别出来，导致识别准确率下降。此类问题本人在今后的学习中继续探讨和研究。

参考文献

- [1] 拉毛措, 安见才让. 基于正则表达式的藏文“ཁོང་”格助词的识别及其检错算法研究[J]. 高原科学研究, 2018, 6, 15.
- [2] 卓玛吉. 藏文虚词自动识别研究[D]. 西宁: 青海民族大学, 2014.
- [3] 高定国, 扎西加, 赵栋材. 计算机识别藏语虚词的方法研究[J]. 中文信息学报, 2014 (1): 114-117.

(上接第189页)的资金使用率和周转率,有效降低了企业的运营成本,为企业构建以一个全新的生态服务系统,为用户在网络中匹配更适合的网络服务,并且在这个过程中按照价格最优化的原则进行运营商带宽选择,运营商能够充分依托SDN技术对用户的资源视图进行自动生成,再借助大数据技术对服务进行定价,再将定价发送到带宽交换点,云服务的消费者和企业同样能够在价格最优的原则下对云服务的运营商进行自主选择。

3 SDN技术在运营商IP承载网中的应用实践

除了谷歌公司平台率先在跨数据中心网络中融合SDN技术的实践外,2015年,中国联通公司提出了一套基于SDN和NFV技术下的CUBE-NET2.0网络构架,这套新兴的网络构架逐渐成为联通公司的网络重构以及技术转型升级的顶层构架。近两年来,联通公司还致力于CUBE-NET2.0业务的网络创新转型,在2018年3月面向OTT云网和政企客户发布了云联网的协同新产品,标志着联通企业的网络业务创新,同时也率先在国内开启了运营商向用户提供一体化服务的先河^[3]。

4 SDN技术在运营商IP承载网应用中的注意事项

根据上面的分析不难看出,将SDN技术运用到运营商IP承载网中十分必要,但是由于当前相关技术标准正在逐渐完善,在网络构架演进的过程中还是有很多地方需要我们注意。由于承载网络具有一定的动态性,因此应当立足于网络构架的可靠性和安全性等方面的因素来注意,在此基础上按照阶段进行演进。

首先,利用现有的技术手段对SDN技术进行改进和完善,通过运营商IP承载网下的核心和骨干层面对网络设备进行科学

部署,不断优化BGP路由的调度方案,进而实现链路流量实时、灵活的动态化疏导。其次,实现IP协议和SDN技术的共存,当逐步对SDN技术的优越性进行了验证后,将BGP、MPLS等TCP/IP协议下传统的技术在IP承载网中进行融合运用,保留路由传统的计算方式,另外注意不断提升路由性能的综合提高^[4]。最后,构建单纯的SDN构架,在保障了该技术的可靠性和安全性等因素的基础上,将TCP/IP协议下传统的BGP、MPLS技术剥离出来,得到单纯的SDN构架。

5 结束语

综上所述,当前,随着我国信息技术领域的飞速发展,SDN技术能够在运营商的IP承载网中构建起一道更加高效、灵活、开放的屏障,让网络服务体系化发展成为可能,有利于全面提高网络服务的使用效率,提升网络运营的整体服务质量,促进我国网络产品更好的发展,为切实满足用户的个性化网络服务需求提供了有利条件。

参考文献

- [1] 傅小宁, 张振文, 刘亚, 等. SDN技术在数据中心网络中的应用研究[J]. 邮电设计技术, 2016 (10): 56-61.
- [2] 鞠婧, 庄皓. IP承载网可靠性技术现网应用[J]. 科技与企业, 2015 (15): 102.
- [3] 朱永庆, 黄晓莹, 张文强. 云网协同时代运营商IP承载网发展[J]. 电信科学, 2017, 33 (11): 162-168.
- [4] 郭宏, 耿燕. IP承载网稳定性技术研究[J]. 中国新通信, 2017, 19 (06): 85.

(上接第213页)相应加权密度,傅里叶变换的模 $|\hat{f}(\omega)|$ 表示振幅。

傅里叶逆变换是用复指数信号的和表示各种形式的信号,复指数信号频率不同,他们的频率是原信号频率的整数倍。这些信号是高频信号,他们修饰与原信号频率相同的信号,使得傅里叶逆变换就等于原信号。频谱上的最低频率就是原信号频率,对应的幅值最大的。

傅里叶逆变换是在频率上求和,不是在时域上。那么,我们可以隔一段时间采集一次信号,然后作傅里叶变换,观察信号在频域上随时间的变化。

5 结束语

傅里叶变换在许多领域有着广泛的应用,比方说物理学、电子类学科、信号处理、海洋学。本文有助于我们深刻理解并掌握傅里叶变换的内涵。

参考文献

- [1] 白艳萍, 雷英杰, 杨明. 复变函数与积分变换[M]. 北京: 国防工业出版社, 2016: 166-180.
- [2] 同济大学数学系高等数学[M]. 北京: 高等教育出版社, 2007.6: 302-321.
- [3] 周顺. 物理中的数学[M]. 武汉: 武汉大学出版社, 2017.12: 97-110.