

用户评价中比较句的识别和倾向性分析

吴晨^{1,2} 韦向峰³

(清华同方光盘电子出版社 北京 100084)¹ (同方知网(北京)技术有限公司 北京 100084)²
(中国科学院声学研究所 北京 100190)³

摘要 在互联网上用户评价内容中很多比较句的比较结果反映了语句陈述者对比较对象的倾向性态度。根据已有的10类比较句句型总结了它们的常见概念搭配,在基于语义块的语句倾向性分析方法的基础上构建了比较句自动识别系统和比较句倾向性自动分析系统。采用第四届中文倾向性分析评测的语料进行实验,对语料中的比较句进行了识别,对比较句中的要素进行了抽取并且分析了比较句的倾向性,识别结果和倾向性分析结果均好于所有参评系统的平均值。

关键词 用户评价,比较句,倾向性分析,概念搭配

中图法分类号 TP391.43 文献标识码 A

Opinion Analysis and Recognition of Comparative Sentences in User Views

WU Chen^{1,2} WEI Xiang-feng³

(Tsinghua Tongfang Compact Disc Digital Press, Beijing 100084, China)¹

(Tsinghua Tongfang Knowledge Network Technology Co., Ltd(Beijing), Beijing 100084, China)²

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)³

Abstract There are many opinions of narrators in the comparative results of comparative sentences which come from the content of user views on the internet. This paper summarized some usual conceptual collocations in ten kinds of comparative sentences. We proposed a method of recognizing comparative sentences and analyzing the opinion of a comparative sentence based on the semantic chunks of a sentence. An experiment was done by analyzing the corpus in the 4th Chinese Opinion Analysis Evaluation, which requires recognizing comparative sentences and analyzing the opinion and elements in a comparative sentence. The experimental results are both better than the average of all systems in the evaluation.

Keywords User views, Comparative sentences, Opinion analysis, Conceptual collocation

1 引言

随着互联网的普及和深入发展,人们的生活和相关信息也越来越多地反映到网络上。博客、社区、论坛、微博和微信的发展,使得人们发表了大量信息、言论和意见于其上,如何从这些海量的网络文本中挖掘出用户对人物、产品和事件的评价,成为一个重要且应用广泛的研究热点。例如人们在买一辆汽车之前,经常会到网络上的汽车论坛了解某些品牌或车型的口碑、期望购买车型在网络上的评价信息等等,从中获取汽车用户的评价信息以为自己决定购买何种品牌和车型提供帮助。企业如出版社则希望在出版某一方向和类型的图书之前,先了解读者和出版商对该类图书的反馈意见和相关评价。在网络时代这些都可以通过网络文本的评价分析,获取用户的评价对象、评价结果和类型,来获得大量数据的倾向性结果,从而避免采用人力分析调研的费时费力、样本量少

等弊端。

倾向性分析技术就是通过分析网络文本语句中所表达的评价倾向,得到文本叙述者或者评价者对评价对象的评价结果和类型。目前倾向性评价的结果类型通常分为正面(褒)、负面(贬)和中立,影响倾向性分析的主要因素有句型、褒贬词、逻辑词、程度词等。通过自然语言处理的词法分析、句法分析、语义分析等技术,建立相应的分析模型和分析算法,可以获取文本中的评价者、评价对象和评价结果。然后,根据评价对象和评价结果结合具体的实际应用可以得到产品的用户评价倾向、多数用户的倾向性意见等。

本文从语句中一类较为特殊的句子比较句入手,研究比较句的句子类型、语义分析模型和倾向性分析方法。比较句表达了陈述者对比较对象、比较范围、比较结果的陈述和结论,有的比较句同时还表达了陈述者对比较对象的褒贬倾向,这对于语句的倾向性分析是十分重要的。针对比较句的特

本文受国家科技支撑计划课题:读者需求分析与按需出版(2013BAH47F02),中国科学院信息化专项:民族语言信息处理学科领域基础科学数据整合与集成应用(Y329251431)资助。

吴晨(1979—),男,博士,高级工程师,主要研究方向为自然语言处理、数据挖掘、数字出版,E-mail: wuchen@cnki.net; 韦向峰(1976—),男,博士,副研究员,主要研究方向为自然语言处理、语音识别、语音合成。

点,很多语言学家进行了总结和研究。有的语言学家把比较句分为平比、差比和极比^[1],有的把比较句分为类同、高下、倚变等类型^[1],还有的总结了比较句的 28 种句式^[2]。这些研究^[3-5]为比较句习得和解析提供了指导,也为比较句的计算机分析和处理提供了语言学基础和类型划分参考。自然语言处理研究者在此基础上进行了比较句的识别和比较关系挖掘的研究,提出了基于支持向量机 SVM 的比较句识别方法^[6]和基于条件随机场 CRF 的比较关系抽取方法^[7];有的研究者则通过基于信息熵和语义角色句法分析树的方法进行比较句的识别和关系抽取^[8],还有的研究者从关联特征词、特征融合的角度对中文比较句的识别进行了研究^[9-10]。在国外,针对英文比较句,Jindal 等人首先利用支持向量机和 CSR(Class Sequential Rules)算法对比较句进行识别^[14],之后又用 LSR(Label Sequential Rules)对比较元素进行抽取,取得了较好的效果^[15];Feldman 等人利用规则的方法从论坛上抽取产品的名称和属性,并对不同产品进行比较^[16];Sun 等人则通过 Web 搜索来寻找两个对象之间的关系^[17]。

本文的余下部分首先介绍比较句的语义句型划分,然后总结比较句句型中常出现的概念搭配,并根据比较句的结构和概念搭配特点进行比较句的识别,最后,在一般语句倾向性分析方法的基础上,根据 3 种不同类型的比较句倾向性的特点,得到比较句倾向性分析方法。在比较句识别、比较句对象属性提取和倾向性分析的实验中,采用了第 4 届中文倾向性分析评测(COAE2012)的评测语料,实验结果比参加评测所有系统的平均水平要高,但在比较句的要素抽取和倾向性分析方面仍需要进一步的深入研究。

2 比较句的语义句型与识别

依据文献^[18],比较句可分为相互比较、集内比较和标准比较 3 种类型。最基本的比较是两两相互比较,例如“A 优于 B”,在汉语里面可以形成不同的句式,如“A 比 B 优”、“与 B 相比 A 优”、“A 和 B 一样优”。集内比较是在一个集合内部的比较,一般会出现最高级的“最”字,或者用否定的表达句式,如“世界上没有哪个国家比美国更强大”。标准比较指比较双方有一个是标准,一般用“符合”、“相符”来表明,如“信封的尺寸要符合国家统一标准”。在相互比较中,可以出现平比和差比,表示两个比较对象一样或者差不多的是平比,表示两个比较对象不一样或者一个优于(强于、大于等)另一个的是差比。在差比中,如果比较的表现含有褒贬倾向,那么从中可以得到陈述者对比较对象的褒或贬的倾向性。而在集内比较中,如果比较表现中含有褒贬倾向,那么也可以得到对比较对象的一个褒贬倾向性。

2.1 比较句的语义句型

比较句最基本的要素包括,比较对象和比较结果。比较对象可以是人、物、事以及它们的属性,比较结果可以分为相同(一样、一致、差不多等)和相异(不同、不一样等),相异的比较结果通常会指明优劣(根据比较对象或属性的不同而使用不同的词语概念)。根据文献^[18],比较句的各种语义句型如下:

- 句型 1 $D00J = DB1 + jD0 + DB2$
- 句型 2 $jD000J = jDBC + jD0$
- 句型 3 $jD001J = jDB + jD0 + DC$

句型 4 $jD021J = DB1 + ljl v + DB2 + jDC$

句型 5 $jD022J = l02 + DB2 + ljl v + DB1 + jDC$

句型 6 $jD01J = Cn + DB0 + jDC$

句型 7 $jD011J = Cn + jlv116 + DB + ljl v + DB0 + jDC$

句型 8 $jD012J = Cn + jlv116 + ljl v + DB0 + jDBC$

句型 9 $jD02J = DBC + jD0 + DB02$

句型 10 $jD020J = DBC + l02 + DB02 + jD0$

其中,句型 1、句型 2 和句型 3 属于相互比较句。相互比较句中,DB1 为第一个比较对象,DB2 为第二个比较对象,jD0 为比较结果(语句的特征语义块)。比较对象可以是具体的人、物、事,也可以是它们的属性,第一个比较对象经常以对象加属性的形式出现,而第二个比较对象的属性经常可以省略。例如:

例 1 骏捷和 F3 的车内空间和外观毫不逊色于 A3。

($jD00J = DB1 + jD0 + DB2$)

例 1 中的“骏捷和 F3 的车内空间和外观”就是第一个比较对象及其属性 DB1,“毫不逊色于”是比较结果 jD0,“A3”是第二个比较对象 DB2,其属性“车内空间和外观”在语句中被省略了。

例 2 速腾 1.4T 手动与宝来最高配的车价差不多。

($jD000J = jDBC + jD0$)

例 2 中,两个比较对象及其属性合起来成为 jDBC“速腾 1.4T 手动与宝来最高配的车价”,而后面的“差不多”则是比较结果 jD0。

例 3 闰土和我仿佛年纪。($jD001J = jDB + jD0 + DC$)

例 3 中,比较对象 jDB 是“闰土和我”,“仿佛”是比较结果 jD0,“年纪”是比较对象的属性,即比较对象和比较对象的属性分离了。

比较句句型 4 和句型 5 属于参照比较句。参照比较句是相互比较句中非常常见的句型,是以某一个比较对象为参照进行比较的比较句。汉语中参照比较句的表达一般会用“比”、“与……相比”等词语,例如:

例 4 这车子比 q7 和卡宴还要炫。($jD021J = DB1 + ljl v + DB2 + jDC$)

例 4 中,“这车子”是第一个比较对象 DB1,“比”是专门用于比较的语言逻辑概念 ljl v(还可以是“相比”、“比较”),“q7 和卡宴”是作为参照的比较对象 DB2,“还要炫”是比较结果 jDC。

例 5 与前代的产品相比,它的操控感明显更好。

($jD022J = l02 + DB2 + ljl v + DB1 + jDC$)

例 5 中,“与”是比较对象的指示标志概念 l02(还可以是“跟”、“和”、“同”),“前代的产品”是作为参照的比较对象 DB2,“相比”是专门用于比较的语言逻辑概念 ljl v,“它的操控感”是另一个比较对象 DB1,“明显更好”是比较结果 jDC。当然,比较句句型在语义块的出现顺序上是可以灵活调整的,如例 5 可以表达为“它的操控感与前代的产品相比明显更好”,其语义块顺序就从“ $l02 + DB2 + ljl v + DB1 + jDC$ ”调整为“ $DB1 + l02 + DB2 + ljl v + jDC$ ”。

比较句句型 6、句型 7 和句型 8 属于集内比较句。集内比较句是在一个集合内进行比较,某个对象与集合内的其它对象都进行比较后,比较的结果一般表现为最高级。其基本语义句型是句型 6($jD01J = Cn + DB0 + jDC$)。其中,Cn 是比较的集合,DB0 是比较的某个对象,jDC 是比较结果。例如:

例 6 这个价位的 Andriod 手机里 v880 最好。(jD01J=Cn+DB0+jDC)

例 6 中, 比较的集合 Cn 是“这个价位的 Andriod 手机里”, 比较的主要对象 DB0 是“v880”, 比较的结果 jDC 则是“最好”。

集内比较句可以加入否定概念 jlv116 来表示某个比较对象与集合内其它对象的比较, 比较结果往往也是最高级。例如“世界上没有哪个国家比美国更强大”属于比较句句型 7 (jD011J=Cn+jlv116+DB+ljl+DB0+jDC), 其中“世界上”是比较集合 Cn, “没有”是否定概念 jlv116, “哪个国家”是集合内的其它对象 DB, “比”是专门用于比较的语言逻辑概念 ljl, “美国”是用于比较的主要对象 DB0, “更强大”是比较结果 jDC。这种句型也可以表示为“世界上没有比美国更强大的国家”, 这时语从句属于比较句句型 8 (jD012J=Cn+jlv116+ljl+DB0+jDBC), 其中“更强大的国家”是比较对象和结果融合后的语义块 jDBC。

比较句句型 9 和句型 10 属于标准比较句。标准比较句的比较对象中有一个比较对象是标准, 例如“信封的尺寸要符合国家标准”属于比较句句型 9 (jD02J=DBC+jD0+DB02), 其中“信封的尺寸”是 DBC, “要符合”是 jD0, “国家标准”是 DB02 即标准; 又如“信封的尺寸要与国家标准相符”属于比较句句型 10 (jD020J=DBC+l02+DB02+jD0), 即“信封的尺寸”是 DBC, “与”是 l02, “国家标准”是 DB02 即标准, “相符”是 jD0, 而“要”是 jD0 的一部分, 分离到了前面。

2.2 比较句的识别方法

从上述比较句的语义句型及句型中的语义块分析可以看出, 比较句中通常会出现两个要比较的对象, 因此需要用对象指示符“l02”或者表示比较的概念“ljl”进行区分, 同时比较句的比较结果语义块“jD0”也可以区分两个比较对象。所以, 根据比较句的语义句型中构成语义块的概念特征, 可以识别判断语句是否是比句。首先, 对语句进行分词处理并得到词语对应的概念符号; 其次, 判断语句中是否出现了符合比较句语义块概念特征的概念或概念搭配; 最后, 根据构成语句的实际语义块是否满足比较句的语义句型来确定语句是否是比句。表 1 总结了实际语料的比较句中常见的概念搭配和词语。

表 1 比较句的概念搭配

比较句句型	概念搭配或概念特征	
	ljl	j
句型 4	比	大/小、好/坏、多/少、快/慢、新/旧、长/短、便宜/贵、逊/胜、高/低、矮/漂亮/好看/丑、强/弱、舒服/先进、轻松/重/沉、宽/窄、简单/复杂、增加/减少、优秀、大气、方便
句型 5	l02 与、和、同、跟	ljl 比、相比、比较、比起来、比较起来
	l02	ljl 其它概念
句型 2	与、和、同、跟	一样、媲美、相同、无异、一模一样、相似、相近、接近、差不多、不同 区别、差距、竞争
句型 6、7、8	l1 或集合 中、里、内、以内、地区、全国	j60d01 最、最好、最佳
句型 1	ljl 好于、大于、高于、好过、赛过、胜于、不亚于、不如、不及、不低于	

实际语料中经常出现的是参照比较句的第一种形式, 即比较句句型 4 (jD021J=DB1+ljl+DB2+jDC), 其中 ljl 的概念常常是“比”字, 而后面的 jDC 常常是表示数量、质量、量度、新旧、美丑等 j 类基本概念。

对于参照比较句的第二种形式, 即比较句句型 5 (jD022J=l02+DB2+ljl+DB1+jDC), 其中的 l02 经常是“与、和、同、跟”, 而后面的 ljl 经常是“比、相比、比较”等专门用于比较的概念。

对于相互比较句的比较句句型 2 (jD000J=jDBC+jD0), jDBC 中两个比较对象之间经常用“和”等概念相连接, 而 jD0 是“一样、媲美、相同、差不多”等概念, 还可以是“区别、差距、竞争”等概念(这时前面一般要加“有”、“没有”等概念)。

对于集内比较句, 即比较句句型 6、句型 7、句型 8, 一般会出现表示比较结果的“最”字。但反过来, 出现“最”字的语句并不都是比句, 要成为比句还需要出现集合, 集合一般是辅块。辅块有一些标志概念 l1, 这些标志概念与最高级概念 j60d01 可以形成集内比较句的识别特征。

对于相互比较句, 即比较句句型 1 (jD00J=DB1+jD0+DB2), 其中 jD0 通常表示比较结果的概念加上“于”、“过”, 或者前面加上“不”等形成表示比较的概念, 如表 1 所列。

2.3 比较句识别的实验结果

为了验证上述对于比较句的语义类型分析和概念特征, 本文使用了第 4 届中文倾向性分析评测(COAE2012)的比较句识别任务即评测任务 2.1 的语料。评测语料分为汽车和电子两个领域, 每个领域各有 3600 个句子, 有的是比句有的不是比句。

本文的比较句识别系统主要是对语句进行分词处理, 然后根据词语的语义概念符号和比较句的语义句型判定语句是否是比句, 主要依据是语句中是否出现比较句的概念搭配和概念特征, 并且是否符合比较句语义句型的语义块要求。根据评测方给出的标准答案, 本文的比较句识别系统的识别结果正确率为 78.1%, 召回率为 84.0%, 二者的平均 F 值 81.1%。这个结果比所有参评系统的平均值(正确率 77%、召回率 81%和 F 值 78%)要高。

本文的比较句识别系统出现识别错误的主要原因如下: 1) 分词错误导致的识别错误, 例如“比”字在“性价比、百分比、高宽比”中就不是专门用于比较句的比较概念, 因而也就不能构成比句; 2) 使用了否定特殊表达的比较句, 例如“没有+DB2+jDC”, 这里 jDC 概念除了指基本概念还可以是很多其它概念, 甚至可以是一个句子, 而根据否定概念和表示比较结果的基本概念只能识别出一小部分这种类型的比较句; 3) 用“是字句”表达的比较句, 例如“跟 STMP3770 还是有点小差别的”, 这里“跟...比”变成了“跟...是”, 而“是字句”的功能更加复杂多变, 比较句只是其中很少的一部分; 4) 概念符号缺乏导致的识别错误, 由于概念是通过词语表达和获取的, 因此会出现系统没有收录的词语, 从而得不到跟比较相关的概念, 特别是表示比较结果的概念, 表示人、事、物属性的量质量的词语和概念非常之多, 不可能形成统一的知识库, 希望能够通过自动学习或者统计训练的算法加以解决。

3 比较句的倾向性分析

比较句虽然具有自己独特的概念特征和语义句型, 但是

从语句的倾向性分析角度来说,通用的倾向性分析模型和分析方法也适用于比较句。下面先介绍一般语句的倾向性分析方法,然后再结合比较句的自身特点说明比较句的倾向性分析方法,最后给出比较句倾向性分析的实验结果。

3.1 一般语句的倾向性分析方法

语句倾向性分析主要是通过词语的倾向性分析得到的。本文首先把影响语句倾向性态度的主要词语分为4类:褒贬词、对象词、逻辑词和程度词,然后根据语句的语义块分析结果转化为简单的二元模型或三元模型,通过褒贬词的倾向性计算得到语句对某个被评价对象的倾向性态度。

褒贬词是指本身具有某种或多种倾向性态度的词语,例如“好”、“坏”、“赞扬”、“谴责”、“光荣”、“露骨”等等。对象词是表示被评价对象或态度持有者的词语,可以是表示人、物品、事件、属性等的各种各样的词语。逻辑词主要是表示肯定或否定的词语,如“是”、“不是”、“不”、“未必”、“没有”等。褒贬词经逻辑词修饰后倾向性可能会发生反转,例如“好”加上否定修饰“不”后变成了“不好”,“好”与“不好”二者的倾向性态度完全相反。程度词是指一些描述程度的修饰词语,如“最”、“很大”、“较大”、“一定程度”等。程度词在运算中不影响倾向性的极性(褒贬性),但可改变句子倾向性(褒或贬)的强弱程度。

对象词和褒贬词是构成语句倾向性分析模型的最基本要素,在具有倾向性态度的语句中一般会同时出现被评价对象词和褒贬词。因此,最简单的二元模型评价形式为“A B”,其中A是褒贬词,B是对象词,反之亦可。例如,语句经简化处理后为“该方法好”,那么语句陈述者对“该方法”的倾向性态度是正面(“好”),语句倾向性态度的取值为“+1”。

倾向性评价分析的三元模型的基本形式为“C X D”,其中C是评价者(对象词),X是褒贬词,D是被评价对象(对象词)。如“他鄙视这种做法”,则评价者“他”对被评价对象“这种做法”的倾向性态度是反面的。从本质上看,三元模型省略评价者后就是二元模型,此时评价者默认为语句文本的陈述者。当不关心评价者只关心被评价对象和对其的倾向性态度时,三元模型可以转化为二元模型处理。而四元模型或更多元的模型,如“A 陈述 B X”(其中B是对象词,X是褒贬词),也可以转化为二元或三元模型处理。

根据二元模型或三元模型,语句倾向性分析的计算结果要么为+1(褒),要么为-1(贬),其结果由褒贬词的属性值确定。如果语句中没有出现褒贬词,则语句倾向性计算结果为0(不具有倾向性)。当语句中出现逻辑词时,在逻辑词的辖域范围内如果有褒贬词,则应乘上逻辑词的属性值(否定为-1,肯定为+1)。当语句中出现程度词时,在程度词的辖域范围内如果有褒贬词,那么应乘上程度词的属性值。

3.2 比较句的倾向性分析

对于相互比较句,句中一般会出现两个比较对象。如果相互比较句具有倾向性,其倾向性要么是对两个比较对象的褒(或者贬),要么是对一个对象褒而对另外一个对象贬。例如“我的技术和他一样好”,这时对两个对象的倾向性都是褒;“我的技术比他好”,这时对“我”是褒,对“他”是贬。对于“DB1+比+DB2+褒贬词”类型的参照比较句,当褒贬词为褒义时,对DB1褒,对DB2贬;反之,当褒贬词为贬义时,对DB1贬,对DB2褒。如果相互比较句中没有出现褒贬词,如“我的

技术和他一样”,那么比较句不具有倾向性。

对于集内比较句,其比较结果一般为含有“最”的最高级,一般都具有倾向性。此时褒贬词的褒贬倾向就是对主要比较对象DB0的褒贬倾向。对于用否定形式表达的集内比较句(参见2.1小节的比较句语义句型),虽然句中出现了否定概念,但是语句对主要比较对象DB0的褒贬倾向依然与褒贬词相同,这一点是需要格外注意的。

对于标准比较句,如果比较结果是“相符”、“符合”之类的,那么比较句不具有褒贬性。如果比较结果是“超过”、“优于”、“强于”之类的,那么对于比较句中除比较标准DB02之外的比较对象(DBC中的DB),倾向性为褒;反之,如果比较结果是“达不到”、“低于”之类的,那么对于比较句中除比较标准DB02之外的比较对象(DBC中的DB),倾向性为贬。

因此,根据一般语句的倾向性分析方法(参见3.1小节)和比较句自身的特点,可以得到比较句对某一个比较对象的褒贬倾向性。本文选取的实验语料仍然是第4届中文倾向性分析评测(COAE2012)的比较句评测语料,评测任务要求抽取出比较句的比较对象及其属性,以及对比较对象及其属性的倾向性。评测语料分为汽车和电子两个领域,每个领域各有3600个句子,有的是比较句有的不是比较句,要求从比较句中抽取出比较对象及其属性,并给出对它们的倾向性结果(1代表褒义,0代表中性,-1代表贬义)。

本文的比较句倾向性分析系统首先对比较句进行语义块切分,然后根据语义块的类型对应得到比较句中的比较对象,再根据语句中褒贬词的褒贬极性和比较句的自身特点确定比较句对于比较对象的褒贬倾向性。根据评测方给出的标准答案,本文的比较句倾向性分析系统的准确率为25.3%,召回率为19.5%,二者平均的F值为22.4%,比所有参评系统的平均值(准确率24.4%、召回率19.2%和F值21.8%)要高。从比较句的比较对象及其属性抽取的抽取结果来看,还是有很大难度的,因为所有参评系统都没有取得高于30%的准确率或召回率。这主要是因为自然语言中描述物品或者物品属性的词语可以说是无限的,而目前还没有能够准确获取这些词语的规则或模型,所以比较句的比较对象及其属性的抽取仍然是需要进一步研究的难题。

结束语 用户评价的倾向性分析有着重要和广泛的用途,从中可以获取用户的偏好和倾向性需求,辅助企业的产品决策或者政府部门的社会公共政策。比较句是一类比较有特点的语句类型,在比较结果中也经常会表达出语句陈述者的倾向性态度。本文从比较句的语义句型以及其中的概念搭配入手,分析总结了比较句中常见的一些概念搭配现象,并据此提出了识别比较句的方法。该方法也存在缺点和不足,需要引入统计学习的方法进行比较句概念搭配特征的总结和学习,尽可能适应特殊的比较句句型以及多变的词语和概念搭配。还尝试了抽取比较句中的对象及其属性等要素,并根据一般语句的倾向性分析模型给出了比较句倾向性的分析方法,但是这方面的系统分析正确率和召回率普遍不高。这说明在比较句的对象及其属性抽取方面,还有很多工作要做,需要进一步的深入研究。此外,比较句并不都完全遵守这些语句类型和搭配,还可能出现比较对象的省略和指代、特殊比较句表达等等,比较对象和比较结果的词语概念多样化也是比

较句分析的难点。统计模型与规则分析方法相融合有可能提高比较句及其倾向性分析的结果和性能。

参考文献

- [1] 宋锐,林鸿飞,常富洋.中文比较句识别及比较关系抽取[J].中文信息学报,2009,23(2):102-107,122
- [2] 陈璐,周小兵.比较句语法项目的选取和排序[J].语言教学与研究,2005(2):22-33
- [3] 许国萍.现代汉语差比范畴研究[M].上海:学林出版社,2007
- [4] 车竞.现代汉语比较句论略[J].湖北师范学院学报,2005(3):60-63
- [5] 刘焱.现代汉语比较范畴的语义认知基础[M].上海:学林出版社,2004
- [6] 黄小江,万小军,杨建武,等.汉语比较句识别研究[J].中文信息学报,2008,22(5):30-37
- [7] 黄高辉,姚天昉,刘全升.基于CRF算法的汉语比较句识别和关系抽取[J].计算机应用研究,2010,27(6):2061-2064
- [8] 李建军.比较句与比较关系识别研究及其应用[D].重庆:重庆大学,2011
- [9] 杜文韬,刘培玉,费绍栋,等.基于关联特征词表的中文比较句识别[J].计算机应用,2013,33(6):1591-1594
- [10] 张辰,冯冲,刘全超,等.基于多特征融合的中文比较句识别算法[J].中文信息学报,2013,27(6):110-116

(上接第424页)

频繁,并且靠海的优越条件使得其成为游客集中的区域,故交通状况相对较差;李沧区、崂山区是新兴城区,其道路依据现代城市规格设计,相对宽阔,商业中心处于建设完善期,规划相对合理,故交通状况普遍良好。

结束语 公共交通是保障大多数市民交通出行的最佳方式,行程速度作为评价公共交通的重要指标,是实现公共交通调度的关键。本文采用公交车GPS数据,首先从聚类类数的确定与初始聚类中心的选择两方面对K-means聚类算法进行改进,然后提出了基于K-means聚类算法计算平均行程速度的数据融合模型,并以青岛市4个城区的实例进行了验证。结果表明应用该模型分析公共交通状况可以达到很好的效果,能够为人们的出行和交管部门的决策提供帮助。

参考文献

- [1] 赖云波.公交浮动车到达时间实时预测研究[D].重庆:重庆大学,2011
- [2] 沙云飞,曹瑾鑫,史其信.基于GPS的路段旅行时间和速度估计算法研究[C]//第一届中国智能交通年会论文集,2005
- [3] 姜桂艳,常安德,李琦,等.基于出租车GPS数据的路段平均速度估计模型[J].西南交通大学学报,2011,46(4):638-645
- [4] 杨兆升.关于智能运输系统的关键理论——综合路段行程时间预测的研究[J].交通运输工程学报,2001,1(1):65-67
- [5] Rice J, van Zwet E. A. Simple and effective method for predicting travel times on freeways [J]. IEEE Transactions on Intelligent Transportation System, 2004, 5(3): 200-207
- [6] Kerner B, Demir C, Herrtwich, et al. Traffic state detection with floating car data in road networks [C] // Proceedings of IEEE Conference on ITS, Vienna, Austria: IEEE, 2005: 700-705
- [7] 聂庆慧,夏井新,张韦华.基于多源ITS数据的行程时间预测体

- [11] 王素格,王凤霞,宋雅.基于序列模式的汉语比较句识别方法[J].山西大学学报(自然科学版),2013,36(2):172-179
- [12] 王凤霞.比较句识别及观点要素抽取方法研究[D].太原:山西大学,2013
- [13] 周红照,侯明午,侯敏,等.基于语义分类的比较句识别与比较要素抽取研究[J].中文信息学报,2014,28(3):136-141,149
- [14] Jindal N, Liu Bing. Identifying comparative sentences in text documents [C] // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006. New York: ACM Press, 2006: 244-251
- [15] Jindal N, Liu Bing. Mining comparative sentences and relations [C] // Proceedings of the 21st National Conference on Artificial Intelligence, 2006. Boston: AAAI Press, 2006: 1331-1336
- [16] Feldman R, Fresko M, Goldenberg J. Extracting product comparisons from discussion boards [C] // Proceedings of the 7th IEEE International Conference on Data Mining, 2007. Washington DC: IEEE Computer Society, 2007: 469-474
- [17] Sun Jian-tao, Wang Xuan-hui, Shen Dou, et al. CWS: a comparative web search system [C] // Proceedings of the 15th International Conference on World Wide Web, 2006. New York: ACM Press, 2006: 467-476
- [18] 苗传江. HNC(概念层次网络理论)导论[M].北京:清华大学出版社,2005

系框架及核心技术[J].东南大学学报(自然科学版),2011,41

(1):199-205

- [8] 姚丽亚,关宏志,魏连雨,等.基于实时交通信息的行程时间估算及路径选择分析[J].公路交通科技,2006,23(11):86-90
- [9] 张和生,张毅,胡东成.路段平均行程时间估计方法[J].交通运输工程学报,2008,8(1):89-97
- [10] 张和生,张毅,温慧敏.利用GPS数据估计路段的平均行程时间[J].吉林大学学报(工学版),2007,37(3):533-537
- [11] 柴华骏,李瑞敏,郭敏.基于车牌识别数据的城市道路旅行时间分布规律及估计方法研究[J].交通运输系统工程与信息,2012,12(6):41-48
- [12] 徐巍,黄浩斌,林建华,等.基于智能卡口系统的道路行程速度计算与实际应用[C]//第八届中国智能交通年会论文集,2013
- [13] 翁建成,荣建,任福田,等.基于非参数回归的快速路行程速度短期预测算法[J].公路交通科技,2007,24(3):93-98
- [14] 姜桂艳,李继伟,张春勤.城市主干路拥挤路段基于地点交通参数的行程速度估计[J].吉林大学学报(工学版),2010,40(5):1203-1209
- [15] 朱健梅.竞争性运输通道选择的博弈模型研究[J].西南交通大学学报,2003,38(3):336-340
- [16] 王殿海,曲昭伟.对交通流理论的再认识[J].交通运输工程学报,2001,1(4):55-59
- [17] 李明涛.快速路短时间尺度地点交通参数多部预测方法研究[D].吉林:吉林大学,2009
- [18] 孙吉贵,刘杰,赵连宇.聚类算法研究[J].软件学报,2008,19(1):48-61
- [19] 周世兵,徐振源,唐旭清.新的K-均值算法最佳聚类数确定方法[J].计算机工程与应用,2010,46(16):27-31
- [20] 张原.公交路网旅行速度估计方法[D].北京:北京交通大学,2012