

情感分析与认知

李维杰

(中国科学院自动化研究所复杂系统与智能科学实验室 北京 100190)

摘要 分析了情感分析的 3 个主要步骤,包括文本情感获取与表达、文本情感分类与计算以及文本情感分析的应用。情感分析得到的结论主要是对相关观点的摘要、对相关事件态度的预测或者统计等,但这些结论都没有发挥文本情感在认知中的作用。为了将情感分析应用于认知科学,提出了情感由情感信号和情感实体组成的观点。情感信号主要是指情感的一些形式载体,比如心跳加速、脸红等这些人体内外的某些表现,表达情感的文字、图片、声音等这类媒体。情感实体主要是指人类对情感形成的一种共识,比如爱、恨、憎恶、高兴、羞愧、嫉妒、内疚、恐惧、焦虑等与人的意识相关联的部分。同时提出了在人工智能中利用情感信息的设想。这对于模拟情感对认知的影响具有一定的意义。

关键词 情感分析,观点挖掘,认知科学,思维科学,人工智能

中图法分类号 TP181 **文献标识码** A

Introduce Sentiment Analysis to Cognitive Science

LI Wei-jie

(Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract We concluded the process of sentiment analysis, which has three main steps, acquisition and expression of sentiment, classification and estimation of sentiment, and application of sentiment analysis. The results of sentiment analysis can be classified into summarization of textual sentiment, estimation or analysis sentiment to persons or events in texts. But nobody gives any attention to the role of emotion in texts as it in cognitive. For replying to this problem, we tried to introduce sentiment analysis to cognitive science and get emotion in texts by utilizing sentiment analysis. We proposed that emotion should have two constituent parts: emotion signal and emotion element. Emotion signals are the carriers of emotion. The carriers include phenomena observed in body such as face color, heart rate, and media which expresses or pass emotion such as images, words, voices. Emotion elements are the ones which are accepted by the society, such as love, hate, delight, abashment, envy, guilt, dread, anxiety and so on, being related to consciousness. We also proposed some ideas on introducing emotion to artificial intelligence. It may be helpful for cognitive science to analyze and utilize emotion.

Keywords Sentiment analysis, Opinion mining, Cognitive science, Noetic science, Artificial intelligence

1 引言

情感分析(Sentiment Analysis),亦称作观点挖掘(Opinion Mining),两者可以互换^[1]。与研究文本主题的相关内容比较,观点挖掘偏向于文本中对相关主题所表达观点的发现与挖掘。文本有客观性与主观性之分:客观性文本偏向于一种事实的陈述和说明,主观性文本偏向于一种观点的表达和提及。观点挖掘的目的是将文本中的主观性文本识别出来,分析其中含有的观点以及它们之间的联系,将得到的结果应用到实际问题中^[1,31]。主体(说话人或者陈述人)在主观性文本中表达对某个实体的观点,里面涉及到大量的主观性情感。这些情感可以分为两类:正面的和负面的。这种定性分类称为“极性分析”。在对情感量化后得到的主观性文本中情感的强度称为“极性强度分析”。

认知科学(Cognitive Science)是 20 世纪 50 年代兴起的

一个研究方向,主要研究对象是人的智能和心理,涉及到许多学科的核心内容,例如语言学、教育学、神经科学、哲学、心理学、人类学以及计算机科学等^[26]。研究认知科学的目的是深入理解人的思维。我国的科学家在综合考虑多方面因素后,提出“思维科学”(Noetic Science)的概念^[15]。许多人都采用信息加工的方法来研究和分析人的认知^[11]。由于计算机的出现以及计算机与人脑的某些相似性的启发,研究者多利用认知心理学的理论做指导,使用计算机实现人工智能,进而达到深入研究认知科学的目的。

认知科学提出来后,亦受到部分学者的批评^[25],其中一条就是“现有的方法都忽略了人类的“情感”。人类的情感是一个复杂的过程,应该包括了生理、机理和心理因素以及其他一些更加复杂的社会人文过程。文本情感分析中的“情感”主要是指词义上概念性质的结论,只是人对其情感所做的表达,属于人类思维的产物。由此可以看出,文本的情感分析更多

只是一种对语言符号和概念相互关系的处理。从认知科学的角度看,文本的情感分析可能根本就不在“情感”的范畴,更多的只是语言学和心理学上的交叉。但是,如果考察这些概念性质的“情感”对人或者与人相关活动的影响,同样可以把文本的情感分析与认知科学联系起来。比如某人在购买电子产品之前,他/她会参考对该电子产品的评论,利用情感分析的方法得到相关结论,而这些结论可以辅助其做出购买决策。这些都涉及到认知科学中的相关内容。

本文的主要工作是在文本情感分析总结的基础上,试图将情感分析引入到认知科学中。本文第1节是相关内容简介和结构说明;第2节是本文的核心内容之一,主要是说明情感分析的过程和现状,同时包括对当前的一些工作的总结和概括;第3节是本文的另外一个核心内容,介绍认知科学中与情感相关的内容以及文本情感分析在认知科学中的地位和作用;最后是本文的总结。

2 情感分析

情感分析是语言学在信息检索中的一个实际应用方向。它的主要任务是识别主观性文本或者计算文本中主观性极性。早期的工作可以追溯到 Hearst 等人^[2]。他们认为智能系统在处理文本的时候,除了关心文本中的主题外,还应该注意文本中评论的立场。这里的“立场”就是情感分析中的情感极性。文本的情感分析可以分为3类^[31]:(1)文本的主观性和客观性识别;(2)主观性文本极性分析;(3)主观性文本极性强度分析。文本的主观性和客观性识别是一个二值分类问题,多将文本中的句子作为研究对象,识别出主观性和客观性文本。由于主观性文本偏向于观点的表达和提及,其中涉及到主体的情感,比如主体所持态度是正面的还是负面的,这对理解文本的内容有帮助。这个过程是分析对主观性文本的极性。在得到主观性文本的极性后,如果实现对极性的量化,就可以通过计算的方法估计文本中立场和分析相关态势。这是主观性文本极性强度分析的目的。

2.1 情感的获取与表达

情感的获取与表达是指从文本中抽取出主观性文本所表达的情感,再使用某种方式表示这些情感。目前主要有两种表达的方式:基于“特征-情感词”对的方法和基于向量空间的方法,两种方法适用于不同的背景。比如在做文本的观点摘要的时候,基于“特征-情感词”对的方法比较适用;在做文本极性分析的时候,基于向量空间的方法比较适用。在获取与表达文本情感的时候,另一个需要关注的问题就是不同粒度之间的关系,主要是指篇章、语句和词语之间关系对文本极性的影响。利用小粒度对象之间的关系帮助更大粒度情感的分类,同时更大粒度表达的情感应该是影响小粒度对象所表达的情感。

基于“特征-情感词”对的方法在事件人物关注度分析和观点摘要中应用比较多,其主要目标就是从文本中抽取最“佳”的“特征-情感词”对。有两个标准衡量“特征-情感词”对是否是最佳:特征是不是关注的对象,情感词是不是对特征的情感准确表达。如图1所示,图中候选特征用下划线标示,候选情感词用波浪线标示。分析语句 S_1 可知,前半句关于天气的说明只是为了与后半部分做对比,不是关注的重点;后半句中涉及到两个特征:“小明”和“(小明的)心情”,不难发现

“(小明的)心情”才是重点。这里对 S_1 的分析是人自然语言的理解和分析,可以很容易地找到关注对象。自动分析的时候采用机器学习的方法,准确率和效率都得考虑。如果能够实现人的自然语言理解功能,准确率会提高。这是目前文本情感分析最大的挑战之一^[10]。现有的自然语言处理工具还很难达到这样的目的。很多人就采用手动标注特征的方法,指定领域内人们关心的全部特征。或者采用机器学习的方法,手动标注部分特征,然后自动增长,得到更多的特征^[31,32]。

S_1	今天天气很好, 可是那个 <u>帅气</u> 的 <u>小明</u> 感觉 <u>很烦</u> , <u>心情很低</u> ;
S_2	封面 <u>华丽</u> , 可是 <u>毫无意义</u> , 就是颜色的组合
S_3	屏幕 <u>很大</u> , 可是 <u>很难受</u> 。

图1 语句中的“特征-情感词”对

情感词一般是形容词、副词,或者蕴含情感的动作(喜欢、爱、讨厌、恨等),通过分词工具得到词性后可以很容易判别。最后一步就是将特征与情感词进行匹配,组成“特征-情感词”对。有两种方法:将与特征距离最近的情感词与其进行匹配^[21];通过分析句子中词语的依存关系,选择修饰特征的或者说明特征行为的情感词进行匹配,或者按照某种规则推导出“特征-情感词”对^[20,34]。第一种方法简单,但是准确率不高。如图1中的 S_2 ,用第一种方法分析可以得到“封面-华丽”。分析可知,最佳的“特征-情感词”对应该是“封面-毫无意义”。第二种方法准确率相对比较,但是速度受到影响。这主要是因为需要完成对语句的结构分析,同时语句结构分析工具的准确率对最后的结果影响很大。如图1中的 S_3 ,分析可知:最佳的“特征-情感词”对应该是“屏幕-大”和“屏幕-难受”。图2是用哈尔滨工业大学信息检索实验室开发的汉语句法分析器得到的分析结果。如果只使用词语之间的结构依存关系,可以得到“屏幕-大”。为了得到“屏幕-难受”,需要定义相关的规则再进行推导。还有一个值得关心的问题就是对隐含情感的分析 and 挖掘^[21]。隐含情感主要是指情感词自己隐含的情感,或者是多个情感词形成的集合隐含的情感。比如, S_3 中的“难受”隐含的情感可能就是“由于屏幕太大,看起来特别费力,导致眼睛难受”或者别的。在得到全部的“特征-情感词”对后,需要分析特征和情感词内部的关系,才能得到这些隐含情感。

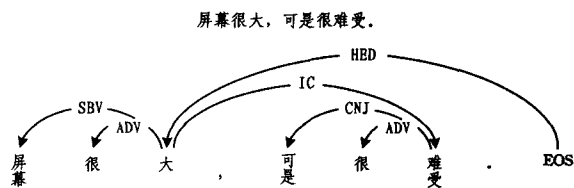


图2 对语句的结构分析

基于向量空间的方法是在文本的向量空间模型的基础上对文本的一种表示。向量空间模型(Vector Space Model, VSM)由 Gerard Salton 等人于 20 世纪 60 年代末提出,最早使用在著名的 Smart 系统中,它是自然语言处理中常用的模型之一^[3]。在向量空间模型中,文本被看成是含有的特征项组成的串,每一特征项都依据规则赋予一个权重,表示它们在文档中的重要程度。其中,“文本”通常指文章中具有一定规

模的片断,如句子、句群、段落、段落组直至整篇文章;“特征项”是最小的不可分的语言单元,可以是字、词、词组或短语等。

可以直接将文本分类中的向量空间模型应用到文本的情感分析中。基于向量空间的方法需要考虑两个问题:特征选择和权重设定。特征选择的主要目的就是消除向量中的噪声和避免“维数灾难”,比较常用的方法有基于文档频率和信息增益。特征项权重的设定与分类器的设计和学习算法有很大关系。文本的情感分析中,已有的研究工作都是在某个特定的特征提取方法下进行的实验。有人的工作表明,在相对比较少的特征项条件下还能得到相对比较高的准确率^[4]。也有人的工作说明,在某些情况下,与信息增益方法相比,传统的特征提取方法对系统的性能提高没有什么帮助^[5]。中国科学院自动化所的李寿山做了多种方法的比较实验;在对多种方法进行理论分析的基础上,作者在文中提出带权重的文档频率和比率方法,实验证明文中提出的方法在不同领域内均取得了较好的分类效果^[9]。

2.2 情感的分类与计算

情感的分类与计算的目的是完成文本情感极性的分类和情感极性强度的计算。情感的分类与计算是情感分析的核心内容,如通过计算文本中情感分值判断文本的主观性和客观性,通过计算文本中正极性或者负极性情感所占比例判断文本的极性。获取与表达情感的方法不同,对情感进行分类与计算的方法也不同。如果采用空间向量的方法,可以使应用贝叶斯、支持向量机、k-最近邻等实现文本情感分类。亦可以采用马尔科夫链的方法识别出文本中所有的情感与观点^[17]。另外,空间向量法与“特征-情感词”对的方法一样,均是在情感词典知识的基础上实现对文本情感的分类与计算。

情感词的识别与情感词的极性和极性强度等属性在文本情感分析中起着至关重要的作用。词或者短语所表达的情感是人类在长期使用语言的过程中积累和约定俗成的,其中涉及到很复杂的社会文化背景。情感词典是这样—个集合:集合中每一项是词语或者短语,称为“项”;每项都有属性,包括极性和极性强度等。由于自然语言的灵活性,每一项都不可能只有一个极性。例如 Esuil 和 Sebadtian 就认为每项有 3 个属性:客观性及其分值、正面性及其分值、负面性及其分值^[31]。情感词典的生成有两种方法:专家手动标注和自动生成。专家手动标注的方法就是邀请语言专家对情感词典中的项进行极性标注和强度量化。自动生成的方法就是使用统计的方法,对每项在语料库中的极性和强度进行估计。

在对文本情感进行分类与计算的时候,可以将文本表示为一个向量 u 。如果用“特征-情感词”对的方法表示文本情感, u 中每一项就是文本中出现的一个“特征-情感词”对;如果用向量空间的方法表示文本, u 中每一项均与特征项相对应。根据情感词典就可得到向量 u 中每一项的极性以及极性强度。再计算 u 中客观性与主观性项的比例以及正负极性项的比例,就可以判断文本的极性以及极性强度^[16]。这种方法是对极性的一种平均。实际使用这种方法的时候,需要考虑另一个问题:全局极性与局部极性的关系。比如在处理语句级别文本的时候,如果出现并列关系的句子,前后句子的极性应该是一致的;如果出现转折关系的句子,前后的极性不一致,后半句的极性应该更加突出。汉语表达以灵活著称,文本

里面可以全部是正面极性的词,可是文本本身却拥有负面立场。这些现象在网络文本,特别是在一些能够充分表达个人观点的文本(比如博客、论坛)中比较普遍。由于用户自由表达情感,形式不受约束,其中情感演化问题十分突出。

2.3 情感分析的应用

由于互联网的普及以及 Web2.0 技术的大量应用,人们可以更加自由地发表意见。这些资源之中含有大量的情感信息,准确获取与跟踪其中情感信息,不论是对商业活动还是社会安全和文化生活等都有很大的意义。这些都表明情感分析的应用会更加普及^[1]。

目前关于文本情感分析的应用可以总结为 3 类:情感摘要、情感预测、情感统计。情感摘要就是将文本中表达情感或者观点的实体抽取出来,组合在一起作为原始文本的摘要。对商品评论的分析、对文本中情感表达的识别等就是属于这一类型^[23,24]。情感摘要可以使别人的情感或者意见一目了然,从而更加容易获取人们对产品或者服务各个部分的态度。情感预测就是使用情感分析的技术预测相关人对将来可能出现的某种事件的情感或者立场。对辩论中每个参与者的立场分析属于这一类^[33],再比如与社会安全相关的“恐怖论坛”的识别等^[22]。情感统计主要是分析人们在多种载体上表达的对事件、对人物的情感或者态度,并使用某种方式将一些有用的结论信息呈现出来。对新闻和博客的分析、对经济类新闻的分析、对政治人物的分析等均属于这类应用^[7,16]。情感统计得到的结论可以让决策者更加全面地认识事物的发展态势。情感预测与情感统计的关系:情感预测是在情感统计的基础上做出的,偏重于预测技术和方法的研究;情感统计偏重于对情感信息的统计及统计结果表达的技术和方法。

文本情感分析近 10 年来发展迅速,国内外的研究者对这方面的工作都比较关注。文本情感分析在实际应用方面取得的巨大成绩推动了情感分析的发展,而文本情感分析目前面临的挑战也吸引了很多研究者参与其中。这些挑战不是来自文本情感分析本身,而是来自与文本情感分析相关的一些基础技术,比如语言学、自然语言处理。在语言学上首要挑战就是情感词典的建立。情感是一个复杂的社会概念。让专家构建情感词典可信度可能会高些,但是毕竟是部分人的意见,不能够完全代表社会这个复杂整体。机器学习的方法得到的情感词典覆盖面可能会比较广,但是单纯地使用词之间的相似关系或者位置关系得到的情感应该不准确^[31]。自然语言处理的挑战就是文本情感的获取。这个问题在汉语中可能会更加突出,主要是因为汉语表达的灵活性很强。如 2.2 节中提到的一种情况:文本里面可以全部是正面极性的词,可是文本本身却拥有负面立场,这就涉及到社会背景等复杂因素。虽然存在挑战,但是—些人的工作仍然具有很强的启发意义。Nikolay Archak 等人把经济学中的产品价格与需求量关系模型引入到文本情感分析中,将顾客对产品特征的评价作为其中的一个参数,在此基础上对产品的特征进行排序^[6]。Mikhail Bautin 等人对多语言文本进行情感分析^[7],这对处理网络文本很有启发意义。网络上存在一些特殊的表达方式,其中含有丰富的感情。借助互联网,这类特殊的表达方式传播速度快、范围广,很容易形成共识。在处理网络文本的时候,可以参考 Mikhail Bautin 等人的思路,将这类特殊的表达方式转换为普通的表达方式后再处理。Nitin Jindal 等人针对互联

网上产品评论中存在虚假消息的情况,通过情感分析方法实现信息过滤^[6]。除了在产品评论中外,其他形式的文本中一样存在这类虚假信息,如博客、论坛等。可以将对这些信息的处理作为对网络文本的一个预处理过程,以净化数据,提高数据的有效性。

3 认知与情感分析

本文所指的“认知”与“认知科学”中的“认知”有相同的意思。上世纪 50 年代出于对人工智能研究的需要,认知科学逐渐兴起。之后一部分人认识到:离开认知科学的支持,人工智能不可能取得成功^[27]。这些都推动着认知科学的发展。我国在上世纪八九十年代也多次组织关于认知的讨论班或者研讨会^[15]。1991 年,美国 San Diego 的 University of California 在世界上第一次授予认知科学方向的博士学位,这标志认知科学作为一门学科正式被认可和接受。研究认知科学的目的是弄清楚人的心智或者思维,并用这些知识从事有益的活动。我国科学家在综合考虑多方面因素后,提出“思维科学”的概念与之相对应^[13,15]。比较“认知科学”与“思维科学”就可以发现,“思维科学”这个提法的针对性和目的性更加明确,因为只有研究思维或者人的认知能力才是有意义的。目前一部分人提倡“第二代认知科学”的核心思想与“思维科学”的内容存在很多的重合^[35]。认知科学的研究对象主要有智能、注意、语言、学习、记忆、感知和行为等。

认知科学增长了人类的知识空间,但对认知科学的批评和质疑还是存在^[25]。这些批评主要就是:忽视情感在思维过程中的作用;忽视顿悟在思维过程中的作用;忽略环境在思维过程中的作用;将身体与人的心智严格分割;忽视人的社会性;人脑应该不只是一个简单的计算系统,更应该是一个动态系统;人脑的计算模式不应该是简单的数学运算,可能采用量子计算方式等等。这些批评和质疑还不能直接推翻或者颠覆认知科学的合理性,但可以作为认知科学发展的新方向。另外,这些早期的批评和质疑中的一部分已经被我国科学家所解决。比如,关于顿悟的作用,我国科学家认为思维应该包括逻辑思维、形象思维以及逻辑思维与形象思维相结合得到的创造思维,这里所说的“创造思维”就是西方学者提到的顿悟^[15,29]。我国学者对“创造思维”的认识早就完成。本文讨论的重点是认知科学中的“情感”。

可以从情感的脑机制来认识和了解“情感”。对情感的脑机制的解释和说明有两个著名学说:James-Lange 学说和 Cannon-Bard 学说^[12]。James-Lange 学说认为人体验情感是因为对身体中的生理变化有反应;但 Cannon-Bard 学说认为情感的体验能够独立于情感的表达之外而产生,同时情感的体验与躯体的生理之间没有必然的联系。这两个观点有些针锋相对,但还是存在共同的地方:情感包括情感体验和情感表达;情感体验、情感表达和人体生理结构之间存在联系。这里所说的情感体验就是日常所说的情感,指人对“情感”的某些感受,它可以对人产生影响;情感表达应该就是人体内外环境刺激生产的反射行为在体内外的某些表现。“情感表达”和“情感体验”是脑神经科学专家的提法,对情感认识限于脑神经科学领域。为了更好地理解,可以将情感分为情感信号和情感实体。情感信号与“情感表达”相对应,主要是指情感的一些形式载体:心跳加速、脸红等这些人体内外的一些表现,

表达情感的文字、图片、声音等这类媒体。可以认为情感信号就是情感的物质载体或者物质形态。情感实体与“情感体验”相对应,主要是指人类对情感形成的一种共识,比如爱、恨、憎恶、高兴、羞愧、嫉妒、内疚、恐惧、焦虑等,这与人的意识有关系。

情感是一个复杂的社会概念,它应该是由个人发起,经社会认可后产生。另外,情感与人体以及人的意识都有密切的联系。如果将情感作为一个整体进行研究,很有可能导致一些比较混乱的情况,比如“情绪”与“情感”的关系。一种情况是将“情绪”与“情感”严格地区分,认为情绪与人的自然性需要有关,同时带有明显的外部表现;而情感与人的社会性需要有关,是一种复杂的体验^[14]。另外一种将“情绪”与“情感”混用,认为“情绪”与“情感”本质上没有区别,只是适用场合的区别等等^[12]。这些观点都有合理之处,但是都不全面。将情感分为情感信号和情感实体,可以更好地认识情感。情感信号是情感实体与人脑一起产生的某些可以观察到的信息,这些信息很容易被归为情感。将情感分为情感信号和情感实体就可以避免这种情况的发生。公认的情感实体有爱、恨、憎恶、高兴、羞愧、嫉妒、内疚、恐惧、焦虑等,而且带有普遍性,在不同的文化中都存在。将情感分为情感信号和情感实体,就可以将关心的重点放在情感实体与情感信号的关系以及情感信号与人体的关系,而不考虑情感实体与人体的关系。

将情感分为情感信号和情感实体,使在人工智能中实现情感和分析情感在认知过程中的作用更加明确,这样可以应对“忽视情感在思维过程中作用”的批评。之前人们试图将情感引入到认知科学中,在某些方面还是存在不足。分析胡包钢等人在文献^[28]中提出的 3 个系统可知,智能、便携式个人身体保健与监护系统和司机安全行车的智能监控系统,属于对情感信号的分析和应用;计算机游戏与娱乐系统,属于将情感实体转换为情感信号的研究。胡包钢等人提到的“情感状态”就是情感信号,但是把对情感信号的处理归为情感计算是不全面的。随着脑神经科学的发展,当人们对脑神经信号和脑控制原理认识更加全面后,如果还把情感信号的处理归为情感计算,很有可能会限制人们对情感在认知科学中的认识,因为这类系统很容易被认为是神经科学方面的知识在医疗等领域的实际应用,这会使认知科学继续受到“现有的方法都忽略了人类的‘情感’”的批评。

将情感引入到人工智能中,可以发挥情感在认知方面的作用。通过情感分析得到文本中的态度,态度中含有个人的情感。这些情感是情感实体,它以文字的形式记录下来,文字就是一种情感信号。要做的工作就是将这些情感实体转为另外形式的情感信号,再使用这些情感信号。例如个性化搜索^[30]方面就可以考虑引入情感分析,模拟情感对人为的影响。用户在提交查询项的时候,同时说明对查询结果的“情感要求”,比如“需要迅速提交结果”、“需要准确提交结果”或者“提供大量的结果”等等。采用文本情感分析方法提取查询项中的情感,根据情感极性强度选择不同的排序算法对查询结果进行处理,将不同的结果提供给用户,以实现个性化搜索。要深入了解情感在认知中的作用,就必须研究情感信号与情感实体的转换关系以及情感信号与人体的相关关系,这对在人工智能中模拟人的情感同时发挥情感在认知科学中的作用同样有帮助。

结束语 本文有两个核心内容:文本情感分析和认知科学中的情感。文本的情感分析是当前的研究热点。与研究文本主题的相关内容比较,文本的情感分析更加偏向于文本中对相关主题所表达观点的发现与挖掘,以及这些观点之中所表达或者隐含的情感信息。它的主要目标是发现文本中含有情感的态度或者观点,根据这些情感信息进行相关预测,用某种更好的方式将这些情感提供给用户。比如对顾客购买行为进行估计、对文本的立场进行估计、对不同条件下产品的特征对顾客的影响进行排序等。

情感在人们智能活动中扮演着重要的不可缺少的角色。但是,情感又是一个复杂的社会概念,它应该是由个人发起,经社会认可后产生。本文提出将情感分为情感信号和情感实体,这可以使对情感的认识更加明确,能很好地分析情感在认知科学中的作用。情感信号主要是指情感的一些形式载体,比如心跳加速、脸红等这些人体内外的某些表现,表达情感的文字、图片、声音等这类媒体。情感信号就是情感的物质载体或者物质形态。情感实体主要是指人类对情感形成的一种共识,比如爱、恨、憎恶、高兴、羞愧、嫉妒、内疚、恐惧、焦虑等与人的意识相关联的部分。研究情感信号与情感实体的转换关系以及情感信号与人体的相关关系,对在人工智能中模拟人的情感同时发挥情感在认知科学中的作用都有帮助。在后期工作中,可以考虑通过文本情感分析的方法得到情感实体后,将情感实体转换为其他形式的情感信号,通过情感信号影响相关学习算法参数的方式,实现情感对认知影响的模拟和应用。

参 考 文 献

[1] Pang Bo, Lee Lillian. Opinion mining and sentiment analysis [J]. *Foundations and Trends in Information Retrieval*, 2008, 2 (1/2):1-135

[2] Hearst M A. Direction-based Text Interpretation as an Information Access Refinement [M]. Jacobs P, ed. *Text-Based Intelligent Systems*. Lawrence, Erlbaum, 1992

[3] Sebastiani F. Machine Learning in Automated Text Categorization [J]. *ACM Computing Surveys*, 2002, 34:1-47

[4] Pang Bo, Lee Lillian, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques [C]// *Proceedings of EMNLP, the Conference on Empirical Methods in Natural Language Processing*, 2002:79-86

[5] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions [C]// *Proceedings of EMNLP*, 2003:105-112

[6] Archak N, Ghose A, Ipeirotis P G. Show me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews [C]// *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Jose, CA(US), 2007:56-65

[7] Bautin M, Vijayarenu L, Skiena S. International sentiment analysis for News and Blogs [C]// *Proceedings of International Conference on Weblogs and Social Media*. Seattle, Washington, USA, 2008

[8] Jindal N, Liu Bing. Opinion Spam and Analysis [C]// *Proceedings of International Conference on Web Search and Data Mining*. Palo Alto, California, USA, 2008:219-229

[9] 李寿山. 情感文本分类方法研究[D]. 北京:中国科学院自动化

所, 2008

[10] Liu Bing. Sentiment Analysis [C]// *Keynote talk at the 5th Annual Text Analytics Summit*. Boston, USA, June 2009:1-2

[11] 司马贺. 人类的认知——思维的信息加工理论[M]. 荆其诚, 张厚燊, 译. 北京:科学出版社, 1986

[12] Bear M F, Connors B W, Paradiso M A. 神经科学——探索脑[M]. 王建军, 等译. 北京:高等教育出版社, 2004

[13] 赵南元. 认知科学揭秘[M]. 北京:清华大学出版社, 2002

[14] 楚明瑞. 人类大脑的工作模型[M]. 北京:中国科学技术出版社, 2006

[15] 赵光武. 思维科学研究[M]. 北京:中国人民大学出版社, 2001

[16] Godbole N, Srinivasaiah M, Skiena S. Large-scale Sentiment Analysis for News and Blogs [C]// *Proceedings of International Conference on Weblogs and Social Media*. Boulder, Colorado, U. S. A, 2007

[17] Breck E, Choi Yejin, Cardie C. Identifying expressions of opinion in context [C]// *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence*, 2007

[18] Attardi G, Simi M. Blog Mining Through Opinionated Words [C]// *Proceedings of the Fifteenth Text Retrieval Conference*, 2007

[19] Devitt A, Ahmad K. Sentiment Polarity Identification in Financial News: A Cohesion-based Approach [C]// *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Czech Republic, June 2007:984-991

[20] Kim Soo-Min, Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text [C]// *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. Sydney, July 2006:1-8

[21] Su Qi, Xu Xinying, Guo Honglei, et al. Hidden Sentiment Association in Chinese Web Opinion Mining [C]// *Proceedings of the International World Wide Web Conference Committee*. Beijing, China, April 2008:959-968

[22] Abbasi A, Chen H. Affect Intensity Analysis of Dark Web Forums [C]// *Proceedings of Intelligence and Security Informatics*, 2007:282-288

[23] Zhuang Li, Jing Feng, Zhu Xiao-yan. Movie Review Mining and Summarization [C]// *Proceedings of the 2006 ACM CIKM International Conference on Information and Knowledge Management*. Virginia, USA, 2006

[24] Cognitive Science (Stanford Encyclopedia of Philosophy)[EB/OL]. <http://plato.stanford.edu/entries/cognitive-science/>

[25] Cognitive science[EB/OL]. http://en.wikipedia.org/wiki/Cognitive_science

[26] Zhou Liang, Hovy E. On the Summarization of Dynamically Introduced Information: Online Discussions and Blogs [C]// *AAAI Symposium on Computational Approaches to Analysing Weblogs*, 2006:237-242

[27] Johnson-Laird P N. Mental models in cognitive science [J]. *Cognitive Science*, 1980, 4:71-115

[28] 胡包钢, 谭铁牛, 王珏. 情感计算——计算机科技发展的新课题 [N]. *科学时报*, 2000-03-24(3)

[29] 尹红凤, 戴汝为. 论思维及模拟智能[J]. *计算机研究与发展*, 1990, 4:1-15

好。表2中的CPU平均计算时间表明,PBIL算法是最快的优化算法,BBO在8个算法中列第5。不过,在大量现实的工程应用中,适应度函数估计消耗的计算是种群优化算法中代价最大的一部分。

Dan Simon也在文献[7]中表明,实验中没有刻意调整优化算法参数,因为不同的参数值可导致优化算法中性能的大幅变化;其二,验证结果是以基准函数为基础的,基准函数选取的是算法测试中常用的函数;现实优化问题与基准函数关系不大,但对于不同问题相应结果可能会发生改变,出现不同的结论;因此,该实验意在表明算法处理常规优化问题是十分有效的,也表明生物地理学优化算法是一种基于种群的、能够解决工程优化问题的有效算法。

结束语 生物地理学优化算法是一种新颖的优化算法。其独特的机制为智能优化算法研究领域注入了新的理念,而且生物地理学众多的研究成果也将为该算法的进一步研究提供深厚的理论基础。实验结果也表明,该算法在一般优化问题上其性能可靠、有效,在一些问题上其性能超过传统优化算法。这些结果也深层次地表明,生物地理学中的机制是大自然长期进化形成的规律,其在优化问题处理方面是独特、有效的。

此外,如图5所示,生物地理学优化算法的自适应机制不仅存在于生物栖息地迁移过程,在经济学的市场调节机制中的均衡价格形成机制就与BBO中的种群迁徙机制类似。可见这种自适应机制在很多领域中起到了重要作用。这些生物迁徙规律和人类社会中的经济运行机制及其研究成果也可作为优化算法理论有益的补充,为研究优化算法的内在机制提供更多支持。

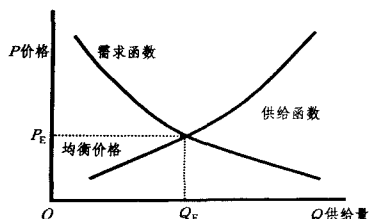


图5 市场中的均衡价格形成示意图

生物地理学优化算法的研究刚刚起步,其中很多问题没有解决,例如栖息地数量和拓扑结构、迁移模型等因素对算法性能的影响还需要进一步阐释。BBO算法在动态优化问题、多目标优化问题和组合优化问题中的性能如何?如何构建更为有效的机制提高算法效率?这些问题都有待进一步深入研究。

(上接第15页)

- [30] Ma Zhongming, Pant G, Liu Sheng, et al. Interest-based Personalized Search [J]. ACM Transaction on Information Systems, 2007, 25(1)
- [31] Esuli A, Sebastiani F. SentiWordNet: A Publicly Lexical Resource for Opinion Mining [C]// 5th Conference on Language Resources and Evaluation, 2006: 417-422
- [32] Al Masum Shaikh M, Prendinger H, Ishizuka M. SenseNet: A Linguistic Tool to Visualize Numerical-Valance Based Sentiment of Textual Data [C]// Proceedings of 5th International Conference

- [1] 汪定伟. 智能优化方法[M]. 北京: 高等教育出版社, 2007
- [2] Dorigo M, Stutzle T. 蚁群优化[M]. 张军, 译. 北京: 清华大学出版社, 2007
- [3] 段晓东, 王存睿, 刘向东. 粒子群算法及其应用[M]. 沈阳: 辽宁大学出版社, 2007
- [4] Wallace. The Geographical Distribution of Animals [M]. Boston, MA: Adamant Media Corporation, 2005
- [5] Darwin. The Origin of Species [M]. New York: Gramercy, 1995
- [6] MacArthur R, Wilson E. The Theory of Biogeography [M]. Princeton, NJ: Princeton Univ. Press, 1967
- [7] Simon D. Biogeography-based Optimization [J]. IEEE Trans. Evolutionary Computation, 2008, 12(6): 702-713
- [8] Gilpin H M. Meta Population Biology [M]. New York Academic, 1997
- [9] Wesche T, Goertler G, Hubert W. Modified habitat suitability index model for brown trout in southeastern Wyoming [J]. North Amer. J. Fisheries Manage, 1987, 7: 232-237
- [10] Onwubolu G, Babu B. New Optimization Techniques in Engineering [M]. Berlin; Germany: Springer-Verlag, 2004
- [11] Storn R. Differential evolution [J]. Dr. Dobb's Journal, 1997, 22: 18-20
- [12] Storn R. System design by constraint adaptation and differential evolution [J]. IEEE Trans. Evol. Comput, 1999, 3: 22-34
- [13] Michalewicz Z. Genetic Algorithms Data Structures Evolution Programs [M]. New York: Springer, 1992
- [14] Beyer H. The Theory of Evolution Strategies [M]. New York: Springer, 2001
- [15] Mezura-Montes E, Coello C. A simple multimembered evolution strategy to solve constrained optimization problems [J]. IEEE Trans. Evol. Comput, 2005, 9: 1-17
- [16] Parmee I. Evolutionary and Adaptive Computing in Engineering Design [M]. New York: Springer, 2001
- [17] Eberhart R, Shi Y, Kennedy J. Swarm Intelligence [M]. San Mateo, CA: Morgan Kaufmann, 2001
- [18] Eberhart R, Shi Y. Special issue on particle swarm optimization [J]. IEEE Trans. Evol. Comput, 2004, 8(3): 201-228
- [19] Clerc M. Particle Swarm Optimization [M]. Amsterdam, The Netherlands: ISTE Publishing, 2006
- [20] Khatib W, Fleming P. The stud GA: A mini revolution? in Parallel Problem Solving from Nature [M]. New York: Springer, 1998
- rence on Natural Language Processing, Hyderabad, India, 2007: 147-152
- [33] Thomas M, Pang Bo, Lee Lillian. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts [C]// Proceedings of EMNLP, 2006: 327-335
- [34] 姜德成, 姚天防. 汉语句子语义极性分析和观点抽取方法的研究 [J]. 计算机应用, 2006, 26(11): 2643-2625
- [35] 李其维. “认知革命”与“第二代认知科学”刍议 [J]. 心理学报, 2008, 40(12): 1306-1327