

# 基于概念的词汇情感倾向识别方法

陈岳峰<sup>1,2</sup>, 苗夺谦<sup>1,2</sup>, 李文<sup>1,2</sup>, 张志飞<sup>1,2</sup>

(1. 同济大学 计算机科学与技术系, 上海 201804; 2. 同济大学 嵌入式系统与服务计算教育部重点实验室, 上海 200092)

**摘要:** 词汇的语义倾向是文本倾向性分析的基础课题. 现有的词汇语义倾向计算通常是以词汇为基准, 而词是包括了多种不同情感倾向概念的粒度范畴, 影响分析的精度和效率. 据此, 提出在更细的粒度下, 利用 HowNet 工具中的“概念”进行倾向性分析, 设计了基于概念的语义倾向计算方法. 该方法使用聚类的概念, 利用 K-MEDOIDS 算法寻找基准概念. 实验结果表明, 基于概念的方法较传统基于词汇的方法准确率更高.

**关键词:** 文本倾向性分析; HowNet; 概念; 聚类; K-MEDOIDS

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2011)06-0489-06

## Semantic orientation computing based on concepts

CHEN Yuefeng<sup>1,2</sup>, MIAO Duoqian<sup>1,2</sup>, LI Wen<sup>1,2</sup>, ZHANG Zhifei<sup>1,2</sup>

(1. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2. The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China)

**Abstract:** The semantic orientation of words is the foundation of sentiment analysis. Current methods to compute semantic orientation of words are mostly based on reference words, while words belonging to the granularity category, including various sentiment orientation concepts, affect the analytical precision and efficiency. In this paper, a new method of semantic orientation computing was proposed based on the reference concepts using the HowNet tool to analyze the tendency. The clustering algorithm K-Medoids was used to search for the reference concepts. The experimental results show that the concept-based method outperforms the word-based method.

**Keywords:** sentiment analysis; HowNet; concept; clustering; K-Medoids

近年来, 文本的倾向性分析愈发受到人们的关注. 文本倾向性分析是指通过挖掘和分析文本中的立场、观点、看法、情绪、好恶等主观信息, 对文本的情感倾向做出类别判断. 文本倾向性分析可包含 3 个粒度: 词汇级别、句子级别以及文档级别. 词汇的倾向性分析是后 2 种粒度的基础. 一般的词汇语义倾向计算都是基于词汇的. 国外学者 Hatzivassiloglou 和 McKeown<sup>[1]</sup>、Turney<sup>[2,3]</sup> 以及 Jaap Kamps<sup>[4]</sup> 等的研究具有很大的启发意义. 文献[1]根据连词的起承转合关系, 判断 2 个词是同义词或是反义词, 从而得到形容词的极性, 但此研究并没有涉及倾向度的度量; 文献[2]利用词汇与程度强烈的褒义词(如 ex-

cellent) 的互信息, 减去它与程度强烈的贬义词(如 bad) 的互信息, 来计算词汇的倾向度; 文献[3]利用搜索引擎的 NEAR 关键字进行类似的研究. 在国内, 刘挺<sup>[5]</sup>、王素格<sup>[6]</sup> 对文本倾向性分析做了全面性的研究. 此外, 朱嫣岚<sup>[7]</sup>、杨昱曷<sup>[8]</sup> 以及熊德平<sup>[9]</sup> 等利用 HowNet 进行了倾向性分析的研究, 这些研究都是基于词汇与词汇之间的某种关联. 但是文献[7]采用目标词与基准词之间的相似度差值的方法, 实验结果的准确率并不是特别高. 同时, HowNet 中的概念是可以脱离词汇而独立存在的, 进而就能剥离出更纯粹的褒贬义概念(而不是使用混合有多种概念的褒贬义词)来进行实验.

为了提高倾向性分析的精度和效率, 在前人的成果和 HowNet 工具的基础上, 提出了一种基于概念的词汇语义倾向度分析方法. 该方法将 HowNet 当中存在的褒贬义概念进行聚类分析, 将聚类中心作为基准概念进行词汇的语义倾向计算.

收稿日期: 2011-03-15.

基金项目: 国家自然科学基金资助项目(60970061, 61075056, 61103067); 上海市重点学科建设资助项目(B004); 中央高校基本科研业务费专项资金资助项目.

通信作者: 陈岳峰. E-mail: dennislyve@gmail.com.

## 1 相关背景知识

### 1.1 HowNet 简介

HowNet 创始人董振东提到, HowNet 是一个以汉语和英语的词语所代表的概念为描述对象, 以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库. 关系是词汇语义的灵魂, 只有通过“关系”才可能教会计算机懂得或在某种程度上懂得“语义”, 也只有通过“关系”才可能教会计算机对语义进行运算.

在此要特别提到 HowNet 中 2 个主要的概念: “概念”与“义原”. “概念”是对词汇语义的一种描述, 又称为义项. 每一个词可以表达为几个概念. “概念”是用一种“知识表示语言”来描述的, 这种“知识表示语言”所用的“词汇”叫做“义原”. “义原”是用于描述一个“概念”的最小意义单位. 除了义原, HowNet 中还用了一些符号(如!、#、%等)来对概念的语义进行描述.

现今国内利用 HowNet 的语义倾向度识别方法通常都是基于相似度的, 因此先介绍基于 HowNet 的相似度度量方法. 刘群<sup>[10]</sup>提出了 2 个层面的相似度度量——概念与概念间的相似度和词汇与词汇之间的相似度.

对概念  $S_1$ 、 $S_2$ , 它们的相似度可表示为

$$\text{Sim}(S_1, S_2) = \sum_{i=1}^4 \beta_i \text{Sim}_i(S_1, S_2).$$

式中:  $\beta_i (1 \leq i \leq 4)$  是可调节的参数, 且有:  $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$ ,  $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$ ,  $\text{Sim}_1(S_1, S_2)$ 、 $\text{Sim}_2(S_1, S_2)$ 、 $\text{Sim}_3(S_1, S_2)$ 、 $\text{Sim}_4(S_1, S_2)$  分别表示第一独立义原描述式、其他独立义原描述式、关系义原描述式和符号义原描述式.

而对于 2 个汉语词语  $W_1$  和  $W_2$ , 如果  $W_1$  有  $n$  个义项(概念):  $S_{11}, S_{12}, \dots, S_{1n}$ ,  $W_2$  有  $m$  个义项(概念):  $S_{21}, S_{22}, \dots, S_{2m}$ , 则  $W_1$  和  $W_2$  的相似度为各个概念的相似度之最大值:

$$\text{Sim}(W_1, W_2) = \max_{i=1,2,\dots,n, j=1,2,\dots,m} \text{Sim}(S_{1i}, S_{2j}). \quad (1)$$

### 1.2 基于词汇的词汇语义倾向度识别方法

文献[7]根据若干对褒贬义基准词, 利用 HowNet 的相似度分析进行词汇的倾向计算, 在思路上是沿用了 Turney 的方法: 设共有  $k$  对基准词, 每对基准词包括一个褒义词和一个贬义词. 褒义基准词表示为  $\text{key}_p$ , 贬义基准词表示为  $\text{key}_n$ , 单词  $w$  的语义倾向值用  $\text{orientation}(w)$  表示, 则

$$\text{orientation}(w) = \sum_{i=1}^k \text{Similarity}(\text{key}_p, w) - \sum_{j=1}^k \text{Similarity}(\text{key}_n, w).$$

式中:  $\text{Similarity}(\text{key}, w)$  即是利用式(1)来计算. 式中的倾向度以 0 作为默认的阈值, 大于 0 即为褒义, 小于 0 即为贬义.

## 2 基于概念的词汇语义倾向度识别方法

### 2.1 问题的提出

文献[7]提出的方法的实验效果并不十分理想, 因此本文作了如下思考: 在 HowNet 的定义中, 词汇包含了若干概念, 对于一个基准褒贬义词, 它可能包含了具有褒贬义倾向的概念, 也可能包含不具褒贬义倾向的中性概念, 举例如表 1 所示. 基于词汇的语义倾向计算, 实际是希望利用基准词具有褒贬义倾向的概念来进行计算的, 更确切地说, 是希望利用褒义词包含的褒义概念, 以及贬义词包含的贬义概念来进行计算, 如图 1 所示. 图 1 中, 左右两边各是褒义基准词和贬义基准词, 求候选词的语义倾向正如虚线箭头所示, 是 2 组词与候选词的相似度之差, 但当褒/贬义基准词中包含了贬/褒义概念和中性概念时, 就会造成效率的损失(不必要的求取相似度)以及对实验效果的负面作用.

表 1 带有不同倾向概念的褒贬义词举例

Table 1 Examples of sentimental words containing concepts of different orientation

	褒义概念数	贬义概念数	中性概念数
老实	1	1	1
简单	1	1	1
随便	1	2	2
天真	1	1	0
贵	2	1	2

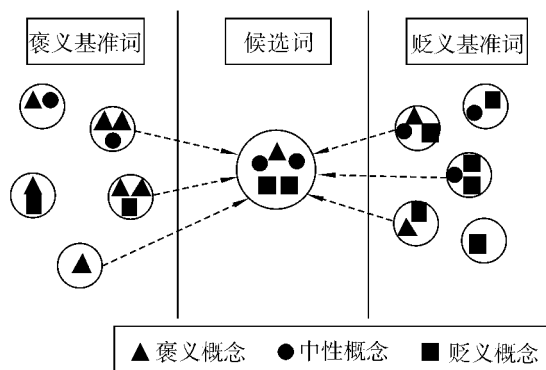


图 1 基于词汇的词汇语义倾向计算

Fig. 1 Orientation computing based on words

于是设想: 当基准词与候选词进行相似度计算时, 这样的中性概念或者反义概念(即褒/贬义词中的贬/褒义概念), 会不会对语义倾向的计算产生负面的效果, 并使得实验不得不进行许多无意义的、冗余的相似度计算, 能否有一种更纯粹的使用褒义和

贬义概念,避开无意义甚至对实验结果有反作用的概念的方法呢?

在 HowNet 这种基于世界知识的工具出现之前,是不能做到的,因为概念的出现必须以词汇为载体.但在 HowNet 出现后,概念可以脱离词汇而独立存在,使这样的方法变得切实可行.

此外还需要解决一个问题,即如何沿袭前人的研究思路,寻找若干对基准概念.概念之间的相似度给定之后,可以将相似度看作是距离的反比,利用聚类的方法寻找出若干个聚合,再从每个聚合中找出聚类中心的方法来获取基准概念.

因此,问题转化为下2个子问题:1)如何利用聚类算法寻找基准概念;2)如何利用基准概念进行词汇语义倾向度分析.

对以上2个问题的解决方案分别在2.2和2.3章节进行详细的介绍.大体思路如图2所示,先使用聚类算法在褒义概念空间和贬义概念空间中各找出 $n$ 个聚类中心(如白色图标所示),再通过这些聚类中心来对候选词的语义倾向进行计算(如虚线箭头所示).

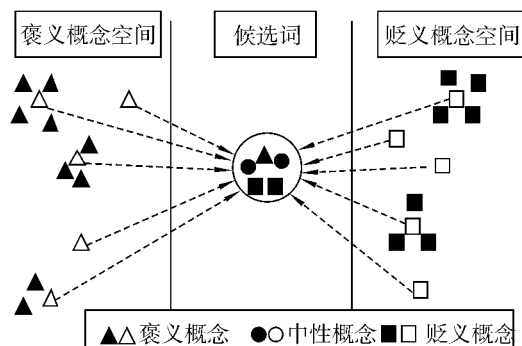


图2 基于概念的词汇语义倾向计算

Fig.2 Orientation computing based on concepts

## 2.2 利用聚类算法寻找基准概念

聚类分析指的是将一种模式的集合(通常表示为向量或者多维空间中的点),基于相似性分成多个组别的过程<sup>[11]</sup>.常用的聚类算法如 K-MEANS 算法、K-MEDOIDs 算法、CURE 算法、DBSCAN 算法等.

概念是一个分布在未知高维度空间中的点,无法用一系列的属性来表征一个概念.由于 K-MEANS 算法在每次迭代中都需要构造新的聚类中心,这个聚类中心是严格意义上类内各样本距离最小的点,有可能是之前未出现过的点,然而在概念空间中是无法构造出之前未出现过的点,因此类似 K-MEANS 的算法不适用.相反, K-MEDOIDs 算法的聚类中心是聚类中与每一个类内样本点的相似度总和最高的点,是从已有样本点中选取出来的,因此是适用的.

另一方面,虽然类似于 DBSCAN 算法这样的基于密度的聚类算法也能够使用,但存在一个明显的

缺陷就是很难控制聚类的数量.

综上,借鉴 K-MEDOIDs 的思路,设计一个基于 K-MEDOIDs 算法的基准概念获取方法.

一般的 K-MEDOIDs 算法过程如下(算法1):

输入:原始数据集以及所求的聚类个数 $k$ ;

输出: $k$ 个聚类;

1)初始化:随机选定 $n$ 数据点中的 $k$ 个作为中心点(medoids);

2)将每一个数据点聚合到最近的中心点;

3)For each 中心点 $m$

For each 非中心点 $o$

交换 $m$ 和 $o$ ,计算类内总的距离消耗;

4)选择消耗最低的点作为该聚类的新中心点;

5)重复2)~4)直至中心点不再改变.

根据 HowNet 的特点,利用算法1提出了基于 K-MEDOIDs 算法的基准概念获取方法(算法2):

输入:

concept\_pos:全体褒义概念;

concept\_neg:全体贬义概念;

$n$ :所求的基准概念数;

输出:

ref\_concept\_pos:基准褒义概念;

ref\_concept\_neg:基准贬义概念;

1)初始化:随机选定 concept\_pos 中的 $n$ 个作为中心点(medoids);

2)根据相似度,将每一个概念聚合到与其相似度最大的中心点;

3)For each 中心点 $m$

For each 非中心点 $o$

交换 $m$ 和 $o$ ,计算中心点与类内非中心点的相似度总和;

4)选择相似度总和最高的那个点作为该聚类的新中心点;

5)重复2)~4)直至中心点不再改变;

6)输出各聚类中心点作为基准褒义概念 ref\_concept\_pos;

7)将 concept\_pos 替换为 concept\_neg(即全体贬义概念),重复1)~6),输出基准贬义概念 ref\_concept\_neg.

## 2.3 利用基准概念进行词汇语义倾向度分析

在获得基准概念对之后,接下来的工作是如何利用它们进行词汇的语义倾向度分析.本文沿袭了前人的研究思路,提出了2个公式:

对于一个词汇 $W$ 和一个概念 $S$ ,如果 $W$ 有 $n$ 个义项(概念): $S_1, S_2, \dots, S_n$ ,它们之间的相似度是

$$\text{Sim}(W, S) = \max_{i=1,2,\dots,n} \text{Sim}(S_i, S).$$

当  $\text{concept}_{p_1}, \text{concept}_{p_2}, \dots, \text{concept}_{p_n}$  为褒义基准概念,  $\text{concept}_{n_1}, \text{concept}_{n_2}, \dots, \text{concept}_{n_m}$  为贬义基准概念时, 对于一个词汇  $W$ , 它的语义倾向度计算公式为

$$\text{orientation}(W) = \sum_{i=1}^n \text{Similarity}(\text{concept}_{p_i}, W) - \sum_{j=1}^m \text{Similarity}(\text{concept}_{n_j}, W). \quad (2)$$

### 3 实验结果与分析

#### 3.1 实验数据

为实验方便且易于比较, 本文仅考虑中文, 不考虑其他语言. 且需要两大类数据源, 一是褒贬义词表, 二是褒贬义概念表.

实验中使用的褒贬义词表是 HowNet 免费对外提供的 4 份褒贬义词表, 如表 2 所示. 其中前 2 份表组成贬义词组, 共计 4 559 个词; 后 2 份表组成褒义词组, 共计 4 739 个词.

表 2 褒贬义词表

Table 2 List of sentimental words

文件名	词汇数量	极性
HowNet_Sentiment_MinusFeeling_C	1 297	贬义
HowNet_Sentiment_MinusSentiment_C	3 261	贬义
HowNet_Sentiment_PlusFeeling_C	873	褒义
HowNet_Sentiment_PlusSentiment_C	3 866	褒义

在 HowNet 概念中有一栏专门的属性 S\_C, 指明该概念的中文语义倾向(相应还有属性 S\_E, 指明英文的语义倾向, 在此先不作考虑). 它共有 4 种值: MinusFeeling、MinusSentiment、PlusFeeling、PlusSentiment. 此外对中性的概念该属性为空. 故将 S\_C 值为前 2 个值的概念全部作为贬义概念, 共计 355 个概念, 将 S\_C 值为后 2 个值的概念全部作为褒义概念, 共计 305 个概念.

#### 3.2 实验方法及评价指标

实验利用算法 2 和式(2)进行, 根据不同的基准概念对数进行实验并作比较.

需要注意的是, 由于 K-MEDOIDS 的初始中心点是随机的, 对于不同的基准概念对数, 本文采用 10 次实验求取平均数作为最后的结果.

在评价部分, 实验不仅根据总体准确率来评价实验效果, 还将比较褒义词和贬义词的准确率, 如果两者自身的准确率越高, 彼此的差距越小, 就说明实验效果越好.

#### 3.3 实验结果与讨论

##### 3.3.1 使用基准词方法的性能

为与基于词汇的语义倾向度分析方法进行比较, 先对文献[7]中提到的 40 对基准词进行统计,

发现褒义基准词中包含 99 组概念, 贬义基准词中包含 127 组概念, 故实验 1 将基准概念对数设置在 90 ~ 140 (这样的话比较次数相当, 时间消耗差不多), 并与基于词汇的语义倾向度分析方法的准确度进行比较, 结果如表 3 所示.

表 3 实验结果

Table 3 Results of experiments

	总体 准确率	褒义词 准确率	贬义词 准确率
40 对基准词	73.9	89.4	57.7
90 对基准概念	76.2	91.1	60.6
100 对基准概念	78.3	90.8	65.1
110 对基准概念	78.5	95.0	61.3
120 对基准概念	80.5	91.8	68.8
130 对基准概念	81.4	95.6	66.7
140 对基准概念	81.9	95.5	67.5

对于总体准确率, 运用基准词的方式获得了 73.9% 的准确率, 而使用基准概念的方式下最好结果达到了 81.9%, 比传统方式高出了 8%. 基于概念的语义倾向度分析方法效果明显好于基于词汇的方法.

出现上述结果有 2 个主要的原因:

1) 基于概念的语义倾向度分析方法使用的都是带有褒贬含义的概念, 针对性更强, 对倾向度分析的作用更大、更直接.

2) 在 HowNet 知识库中, 一个褒义词可能不仅包含褒义概念, 还包含贬义概念和无褒贬含义的概念, 在进行倾向性分析的时候, 仅它的褒义概念会起正面作用, 而贬义概念和无褒贬含义的概念则可能会有反作用或是没有作用; 同样的问题也存在于一些贬义词中. 使用基于概念的语义倾向度分析方法, 可以消除反作用, 同时避免不必要的时间损耗. 对于褒义词和贬义词各自的准确率, 不管何种方式褒义词准确率明显高于相应的贬义词准确率, 基于相似度的方法似乎很难避免褒、贬义词准确率偏斜现象的出现. 但是相对基于词汇的方法贬义词仅有 57.7% 准确率, 基于概念的方法在准确率上有明显的改进, 在一定程度上纠正了准确率偏斜现象.

此外, 与基于词汇的语义倾向度分析方法相比, 基于概念的语义倾向度分析方法还体现出 2 个优点: 其一是自动化程度高, 在利用聚类算法寻找基准概念的过程中, 基准概念是从 HowNet 所提供的所有褒贬义概念中自动地选取, 而非人为指定基准词, 在认同 HowNet 是一种通用工具的情况下, 寻找基准概念的过程可认为是一种只需指定若干参数即可自动化的过程; 其二是分析速度更快, 在所使用的概念数大致相同, 且认为每次 HowNet 计算 2 个词相似度的时间复杂度相同的情况下, 基于概念的语义倾向度分析方法减少了分析词汇、提取概念的过程,

因此分析的速度更快。

### 3.3.2 基准概念数变化趋势分析

根据表3,基准概念方法的准确率随着基准概念数的增加而逐渐提高。下面对产生这一现象的原因进行深入的分析。

对使用改进的 K-MEDOIDS 聚类方法得到的整个概念空间的聚类结果进行统计,得到概念空间的稀疏向量,记为  $\text{centroids}(n)$ , 其中  $n$  表示基准概念数,即聚类中心个数。为了减少 K-MEDOIDS 方法中随机初始化带来的影响,每次聚类重复  $m$  次,在本实验中  $m$  设为 10。 $\text{centroids}(n)$  中每一维向量,即候选概念的权值按如下方式确定:首先将向量  $\text{centroids}(n)$  每一维初始化为 0,对于每次的聚类结果,在每个聚类中心所代表候选概念上的权值增加  $1/m$ ,重复  $m$  次实验。通过上述方式得到的向量  $\text{centroids}(n)$  中每一维的权值区间为  $[0, 1]$ ,该向量表征了整个空间中每个概念选为基准的概率。

按上述方法统计之后,对相邻 2 个聚类个数的聚类中心结果分别进行如下分析:首先计算两者的相似度,在此使用余弦相似度的计算方法;其次统计在  $\text{centroids}(n)$  中出现频数较  $\text{centroids}(n-10)$  增长最多的 10 个概念,以及它们在  $\text{centroids}(n)$  之前的出现次数。统计结果如表 4 所示,其中  $\Delta P_{\text{cur}}$  表示当前出现频数增长最多的 10 个概念的平均增加值,  $P_{\text{pre}}$  为这 10 个概念在之前的几个维度出现的平均次数。例如,  $\Delta P_{\text{cur}} = 0.46$ ,即表示出现频数增长最多的 10 个概念在该维度平均多出现了 0.46 次,而  $P_{\text{pre}} = 0.16$ ,即表示这 10 个概念在之前几个维度中分别只出现 0.16 次。

从表 4 中可以看出,相邻基准概念数的相似度非常高,都达到了 90%。这说明随着基准概念数的增加,每组基准概念除了保持与前一组基准概念的大致相似之外,都会稳定地引入一些之前出现不多的概念,扩展了基准概念的空间,从而一定程度上提高了准确率。

表 4 聚类中心统计

Table 4 Statistics on centroids

基准概念	相似度	$\Delta P_{\text{cur}}$	$P_{\text{pre}}$
90	0.90	0.46	0.16
100	0.90	0.47	0.18
110	0.90	0.48	0.16
120	0.91	0.45	0.20
130	0.92	0.46	0.16
140	0.92	0.50	0.18

## 4 结束语

本文提出了一种概念粒度层次下的语义倾向度

分析方法,一定程度上克服了现有方法中使用词汇作为基准影响性能的弊端。所提方法从 HowNet 知识库中抽取出概念,并使用 K-MEDOIDS 聚类算法寻找基准概念,与基于词汇的语义倾向分析方法相比,其自动化程度更高、分类的速度更快、准确度更高。但针对 K-MEDOIDS 聚类算法中初始点的选定存在随机性的问题,下一步的研究尝试使用人工干预的方式,基于可获取的先验知识进行初始基准概念的选取。

## 参考文献:

- [1] HATZIVASSILOPOULOS V, MCKEOWN K. Predicting the semantic orientation of adjectives[C]//Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL. New Brunswick, Canada, 1997: 174-181.
- [2] TURNEY P. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA, 2002: 417-424.
- [3] TURNEY P, LITTMAN M. Measuring praise and criticism; inference of semantic orientation from association[J]. ACM Transactions on Information Systems, 2003, 21(4): 315-346.
- [4] JAAP K, MAARTEN M, ROBERT J M, De RIJKE M. Using WordNet to measure semantic orientations of adjectives[C]//Proceedings of the Fourth International Conference on Language Resources and Evaluation. Lisbon, Portugal, 2004, IV: 1115-1118.
- [5] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.  
ZHAO Yanyan, QIN Bing, LIU Ting. Sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834-1848.
- [6] 王素格. 基于 Web 的评论文本情感分类问题研究[D]. 上海:上海大学, 2008:21-24.  
WANG Suge. Research on problems for sentiment classification of review texts based on web[D]. Shanghai: Shanghai University, 2008: 21-24.
- [7] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20.  
ZHU Yanlan, MIN Jin, ZHOU Yaqian, et al. Semantic orientation computing based on HowNet[J]. Journal of Chinese Information Processing, 2006, 20(1): 14-20.
- [8] 杨昱曷, 吴贤伟. 改进的基于知网词汇语义褒贬倾向性计算[J]. 计算机工程与应用, 2009, 45(21): 91-93.  
YANG Yubing, WU Xianwei. Improved lexical semantic tendentiousness recognition computing[J]. Computer Engineering and Applications, 2009, 45(21): 91-93.
- [9] 熊德平, 程菊明, 田胜利. 基于 HowNet 的句子褒贬倾向性计算[C]//中国人工智能学会第 12 届全国学术年会, 哈尔滨, 2007: 910-913.  
XIONG Deping, CHENG Juming, TIAN Shengli. Sentence

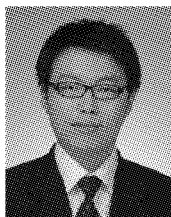
orientation research based on HowNet[C]// Proceedings of the 12th Annual Meeting of Chinese Association for Artificial Intelligence. Harbin, 2007: 910-913.

- [10] 刘群, 李素建. 基于《知网》的词汇语义相似度的计算 [C]//第3届汉语词汇语义学研讨会. 台北, 中国, 2002: 59-76.

LIU Qun, LI Sujian. Word similarity computing based on How-net[C]//The 3rd Chinese Lexical Semantics Workshop. Taipei, China, 2002: 59-76.

- [11] JAIN K, MURTY M N, FLYNN P J. Data clustering: a review[J]. ACM Computing Surveys, 1999, 31(3): 264-323.

作者简介:



陈岳峰, 男, 1986年生, 硕士研究生, 主要研究方向为文本倾向性分析、文本信息处理、数据挖掘。



苗夺谦, 男, 1964年生, 教授, 博士生导师. 中国计算机学会理事, 中国人工智能学会理事, 上海市计算机学会理事等. 已主持完成多项国家级、省部级自然科学基金与科技攻关项目, 并参与完成“973”计划项目1项, “863”计划项目2项等, 曾获国家教委科技进步三等奖、山西省科技进步二等奖、教育部科技进步一等奖、上海市技术发明一等奖、重庆市自然科学一等奖等. 主要研究方向为智能信息处理、粗糙集、粒计算、网络智能、数据挖掘等. 发表学术论文140余篇, 其中被SCI和EI检索70余篇, 出版教材及学术著作6部, 授权专利9项.



李文, 女, 1980年生, 博士研究生, 主要研究方向为文本信息处理、粗糙集、粒计算.

## 第7届智能信息处理国际会议

### 7th International Conference on Intelligent Information Processing

The IIP conference series provides a forum for engineers and scientists in academia, university and industry to present their latest research findings in any aspects of intelligent information processing. IIP 2012 attempts to meet the needs of a large and diverse community.

Topics of IIP 2012 include, but are not limited to:

Autonomic computing  
Automatic reasoning  
Automatic translation  
Bayesian networks  
Belief revision  
Bioinformatics  
Biomedicine  
Business intelligence  
Case-based reasoning  
Cloud computing  
Cognitive modeling  
Computational intelligence  
Computer vision  
Constraint programming  
Data mining e-learning  
E-services  
Evolutionary computation  
Fuzzy and rough set  
Games & Serious games  
Image processing  
Information retrieval

Intelligence science  
Intelligent agents  
Internet of things  
Knowledge acquisition  
Knowledge-based systems  
Knowledge discovery  
Knowledge engineering and management knowledge grid  
Knowledge representation  
Logics and logic programming  
Machine learning  
Machine translation  
Mobile agents  
Natural language processing  
Neural computing  
Ontology engineering  
Reinforcement learning  
Speech understanding and interaction  
Text understanding  
Web intelligence  
Very large knowledge systems

Website: <http://www.intsci.ac.cn/iip2012/>.