

## 基于情感关键句抽取的情感分类研究

林 政<sup>1,2</sup> 谭松波<sup>1</sup> 程学旗<sup>1</sup>

<sup>1</sup>(中国科学院计算技术研究所 北京 100190)

<sup>2</sup>(中国科学院大学 北京 100049)

(linzheng@ict.ac.cn)

## Sentiment Classification Analysis Based on Extraction of Sentiment Key Sentence

Lin Zheng<sup>1,2</sup>, Tan Songbo<sup>1</sup>, and Cheng Xueqi<sup>1</sup>

<sup>1</sup>(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100049)

**Abstract** A key problem of sentiment analysis is to determine the polarity of a review is positive (thumbs up) or negative (thumbs down). Unlike topic-based text classification, where a high accuracy can be achieved, the sentiment classification is a hard and complicated task. One of the main challenges for document-level sentiment classification is that not every part of the document is equally informative for inferring the polarity of the whole document. Thus, making a distinction between key sentences and trivial sentences will be helpful to improve the sentiment classification performance. We divide a document into key sentences and detailed sentences. Key sentence is usually brief but discriminative while detailed sentences are diverse and ambiguous. For key sentence extraction, our approach takes three attributes into account: sentiment attribute, position attribute and special words attribute. To make use of the discrepancy and complementarity of key sentences and detailed sentences, we incorporate key sentences and detailed sentences in supervised and semi-supervised learning. In supervised sentiment classification, a classifier combination approach is adopted because the original document is divided into two different and complementary parts; in semi-supervised sentiment classification, a co-training algorithm is proposed to incorporate unlabeled data for sentiment classification. Experimental results across eight domains show that our approach performs better than the baseline method and the key sentence extraction is effective.

**Key words** sentiment classification; key sentence; classifier combination; co-training; supervised learning; semi-supervised learning

**摘 要** 情感分析需要解决的一个重要问题是判断一篇文档的极性是正面的还是负面的。情感分类的正确率很难达到普通文本分类的水平,因为情感分类更难更复杂。在判断文档的情感极性时,不同的句子具有不同的情感贡献度,所以,对整篇文档的关键句和细节句进行区分将有助于提高情感分类的性能。关键句通常简短且具有判别性,而细节描述句通常复杂多样且容易引入歧义。在关键句抽取算法中,考虑3类属性:情感属性、位置属性和关键词属性。为了更好地利用关键句和细节句之间的差异性和互补性,将抽取的关键句分别用于有监督的和半监督的情感分类。在有监督情感分类中,采用的是分类器融合的方法;在半监督情感分类中,采用的是Co-training算法。在8个领域上进行实验,结果表明所提方法性能明显优于Baseline,从而证明情感关键句抽取算法是有效的。

收稿日期:2011-04-27;修回日期:2011-10-27

基金项目:国家自然科学基金重点项目(60933005);国家“八六三”高技术研究发展计划基金项目(2010AA012500);国家自然科学基金项目(60803085)

**关键词** 情感分类;关键词;分类器融合;联合训练;有监督学习;半监督学习

**中图法分类号** TP391.1

在自然语言处理领域,情感分析受到了很多研究者的关注<sup>[1-5]</sup>.情感分类和普通文本分类有些类似,但比普通文本分类更复杂.在基于主题的文本文本分类中,因为主题不同的文本之间词语运用不同,词语的领域相关性使得不同主题的文本可以很好地进行区分<sup>[6]</sup>.然而,情感分类的正确率比基于主题的文本文本分类低很多,这主要是由情感文本中复杂的情感表达和大量的情感歧义造成的.此外,在一篇文章中,客观句子与主观句子可能相互交错,或者一个主观句子同时具有两种以上情感,因此文本情感分类是一项非常复杂的任务<sup>[7]</sup>.这里,以一篇 Web 上的图书评论为例,如图 1 所示:

很多人说这是一个充满悲伤,流溢无奈的故事,或许正是这种评论让我一直没有勇气去认真阅读。我承认自己是个沦落俗套的人,虽然悲剧让人震撼而且记忆深刻,但从感情上更愿意看到美好的大团圆结局,虽然这样的童话在现实中是如此脆弱而不堪一击。  
.....  
这本书,我是一口气看完的,很喜欢。

Fig. 1 An example of book review.

图 1 图书评论示例

文中作者用了大量消极的词汇来描述阅读前的感受,比如“悲伤”和“脆弱”,但是在文章结尾,作者又用很积极的态度表达了他是喜欢这本书的.在这个例子中,整篇文本的极性是正面的,但由于出现大量负面词汇所以很容易被判别成负面的.在判定整篇文章的极性时,文章中所有句子的情感贡献度是不同的,如果对情感表达关键词和描述细节的句子进行区分,将有助于提高文本情感分类的性能.

在关键词抽取算法中,本文主要考虑了 3 个特征,分别是情感特征、位置特征和关键词特征.情感特征用于判断一个句子是否包含主观感受,同时衡量这个句子所包含的情感权重;位置特征用于确保位于文章开头或结尾的句子具有较高的得分;关键词特征考虑了经常出现在关键词中的一些特殊词或短语,比如“总之”等等.

在有监督的情感分类中,本文采用分类器融合的方法.分类器融合方法的主要特点是每个基础分类器越独立融合后效果越好,而原始文档恰被划分成两个不同且互补的视角,所以非常适用于分类器

融合的方法.

在半监督的情感分类中,本文采用 Co-training 的方法将未标注数据加入训练集.由于关键词和细节句遵循不同的词汇分布,所以一个可以被关键词分类器轻松标注的样本,不一定能被细节句分类器很好的标注,而细节句分类器却可以从这个不易标注的样本上学到新的知识以提高自身性能,反之亦然.

## 1 相关工作

按照学习方法的不同,情感分类可以分成有监督式、无监督式和半监督式的研究方法.

在有监督学习的方法中,文本的情感倾向性分析可以看成文本分类过程. Pang 等人<sup>[4]</sup>借助机器学习的方法对文本倾向性进行判别,取得了不错的效果. Gamon<sup>[8]</sup>证实先进行大量特征融合然后再进行特征约简,可以在有噪音的用户反馈数据集上取得较高的精度. Melville 等人<sup>[9]</sup>在文本分类的基础上融合了词典知识来对博客进行情感分析. Li 等人<sup>[10]</sup>将文本划分成主观句和客观句两个视角,分别训练分类器然后进行分类器融合,也取得了不错的效果.

在无监督学习的方法中,情感分析在没有任何标注数据的条件下进行. Turney<sup>[5]</sup>提出的方法具有较大影响,具体做法是:首先对文本进行词性标注,按照预定义的规则选取形容词和副词的某些搭配,然后计算每一个搭配与“excellent”和“poor”的互信息之差,最后对一个文本所有搭配的互信息差进行求和以判断其情感类别. Zagibalov 等人<sup>[11]</sup>针对中文的商品评论提出了一种自动的种子词选取技术,用于无监督的情感分类.

在半监督学习的方法中,大量无标注数据与少量有标注数据相结合. Li 等人<sup>[12]</sup>采用基于词汇先验知识和非负矩阵分解的方法解决半监督的情感分类问题,其中,有标注的文本、无标注的领域相关文本和领域无关的先验知识一并纳入分类. Dasgupta 等人<sup>[6]</sup>首先用谱分析技术挖掘出态度明确无歧义的评论,然后在这些无歧义评论的基础上,通过一种主动学习、直推式学习和集成学习相结合的方法对有歧义的评论进行分类. Wan<sup>[13]</sup>提出了一种 Co-training

的方法对有标注的英语语料和无标注的中文语料进行半监督学习,从而得到中文的情感分类器。

## 2 情感关键句抽取算法

情感关键句抽取算法主要考量了3个属性,分别是情感属性、位置属性和关键词属性。给定一篇文章,对每一个句子分别计算3个属性的得分,然后进行加权求和,得分最高的句子则被选为情感关键句。

已知任意文本  $d$  由一系列句子组成:  $d = \{s_1, s_2, \dots, s_m\}$ , 其中  $m$  代表句子数目,而每个句子  $s_i$  由一系列词组成:  $s_i = \{w_{i1}, w_{i2}, \dots, w_{in}\}$ , 其中  $n$  代表词的数目。每一个句子的最终得分可以表示成3个属性的加权求和形式:

$$f(s_i) = \lambda_1 \times f\_sentiment(s_i) + \lambda_2 \times f\_position(s_i) + \lambda_3 \times f\_keyword(s_i), \quad (1)$$

其中,  $\lambda_1, \lambda_2, \lambda_3$  是每一个属性的权值,通过最大化分类器的精度来得到。

### 2.1 情感特征

情感关键句主要表达作者的整体观点或偏好,而观点和偏好通常都由情感词体现。情感属性用于考察一个句子是否具有情感色彩,并且衡量其情感重要程度,打分函数如下:

$$f\_sentiment(s_i) = \frac{\left| \sum_{j=1}^n positive(w_{ij}) - \sum_{j=1}^n negative(w_{ij}) \right|}{n}, \quad (2)$$

其中  $positive(w_{ij})$  和  $negative(w_{ij})$  分别定义为:

$$positive(w_{ij}) = \begin{cases} 1, & w_{ij} \in positive; \\ 0, & w_{ij} \notin positive; \end{cases} \quad (3)$$

$$negative(w_{ij}) = \begin{cases} 1, & w_{ij} \in negative; \\ 0, & w_{ij} \notin negative. \end{cases} \quad (4)$$

本文采用情感词典 ([http://www.keenage.com/html/c\\_index.html](http://www.keenage.com/html/c_index.html)) 来判定一个词的情感极性,式(2)中分母  $n$  的引入是为了避免抽取过长的句子,分子采用绝对值的形式是为了抽取具有单极性情感色彩的句子。比如句子“这家酒店位置挺好但是服务很差。”,其中“挺好”和“很差”出现在同一个句子中,但是具有相反的极性,这会为整个句子的情感判定带来混淆,因此,本文只抽取具有单极性情感色彩的句子作为关键句。

### 2.2 位置特征

为了有效从互联网用户评论中抽取主要观点,需要格外重视文章的结尾部分<sup>[14]</sup>。本文认为文章开

头和结尾的句子同样重要。位置属性确保文章开头和结尾的句子成为关键句的得分大于文章中间的句子,打分函数定义如下:

$$f\_position(s_i) = \alpha \times pos(s_i)^2 + \beta \times pos(s_i) + c, \quad (5)$$

服从:

$$\begin{aligned} \frac{-\beta}{2\alpha} &= \frac{m}{2}; \\ pos(s_i) &= i; \\ \alpha &> 0; \beta < 0; \end{aligned}$$

其中  $m$  代表文档  $d$  中的句子数目;  $pos(s_i)$  是一个整数,指示着句子  $s_i$  在  $d$  中的位置;  $\alpha, \beta$  和  $c$  分别是多项式系数。事实上,函数  $f\_position(s_i)$  是一条开口向上的抛物线,横坐标代表句子的位置,取值区间 1 到  $m$ ,纵坐标代表该位置句子成为情感关键句的得分。由抛物线性质不难看出,文章正中间的句子位于抛物线的最低点,中间句子成为情感关键句的得分较小,而位于开头和结尾两端的句子成为情感关键句的得分较高。

### 2.3 关键词特征

情感关键句中经常包含一些总结性的词汇或短语,比如中文的“总之”和英文的“overall”,这种总结性关键词为情感关键句的抽取提供了很好的启发式信息。在本文工作中,我们在训练集上通过对每篇文本的最后一句话进行词频统计,整理得到 10 个比较常见的总结性关键词,如图 2 所示:

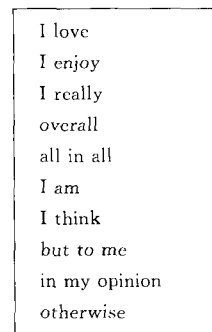


Fig. 2 Top 10 key words.

图2 前10个关键词

如果这些关键词出现在某一句子中,则该句成为关键句的可能性就比较大,因此关键词属性的打分函数定义如下:

$$f\_keyword(s_i) = \sum_{j=1}^n keyword(w_{ij}), \quad (6)$$

$$keyword(w_{ij}) = \begin{cases} 1, & w_{ij} \in keywords; \\ 0, & w_{ij} \notin keywords. \end{cases} \quad (7)$$

### 3 基于情感关键句的分类器融合

#### 3.1 动机

分类模型有很多,如果对于一个新问题缺乏足够多的先验知识,人们会倾向于尝试很多不同的分类方法,然后将不同分类器的预测结果融合起来以获得更好的结果。

在经过关键句抽取后,训练集自然而然被划分成两个子集:关键句集合和细节句集合。关键句集合和细节句集合的词汇分布不同,关键句通常是总结性的,比较简短,而细节句是描述性的丰富多样。

#### 3.2 本文方法

首先,训练 3 个基础分类器:用关键句集合训练得到分类器  $f_1$ ,用细节句集合训练得到分类器  $f_2$ ,用整个语料集训练得到分类器  $f_3$ 。其中,每一个基础分类器不仅输出类别标签,还输出测试样本属于每一个类别的概率。然后,用 3 个基础分类器的融合结果来为每一个测试样本指定类别。融合的方法有很多,不失一般性,本文采用简单求和的融合策略。

简单求和策略是将每个基础分类器输出的后验概率进行累加,用累加后的结果来进行类别判定。对于任一样本  $d$ ,其类别  $c_j$  判定公式如下:

$$j = \arg \max \sum_{i=1}^3 p(c_i | d). \quad (8)$$

### 4 基于情感关键句的 Co-training 算法

#### 4.1 动机

Co-training 算法最早由 Blum 和 Mitchell 提出<sup>[15]</sup>,是一种 Bootstrap 模式的学习方法,要求数据有两组相互独立而又能充分表示数据的特征集,并在每个特征集下训练一个分类器。

在关键句集合和细节句集合上训练的分类器可以为彼此提供不同的知识。关键句集合的词汇分布相对集中,有些词出现频率比较高,比如“推荐”和“失望”等等,所以关键句集合上训练的分类器对这些词的敏感度较高。如果一篇文本存在关键句,那么关键句上训练的分类器将会对其很好的分类。换言之,关键句分类器更适用于处理歧义存在的情况,因为它可以忽略正文中的复杂特征而去集中处理关键句中的有限特征。然而,由于评论撰写者有着不同的写作习惯,并不是所有的评论都包含情感关键句,所以细节句分类器依然是必要的。如果评论中不存在

歧义,细节句分类器会很好地发挥性能,因为其训练语料规模更大,特征空间更丰富。

#### 4.2 本文方法

在半监督情感分类中,本文采用一种 Co-training 的算法。在本文的 Co-training 算法中,一共采用 3 个视角,分别是关键句视角、细节句视角和全文视角。算法 1 显示了本文采用的 Co-training 算法流程。

**算法 1.** 基于关键句的 Co-training 算法。

已知:

特征集  $F_{key}, F_{detail}, F_{full}$ ;

有标注的训练集  $L$ ;

无标注集合  $U$ ;

迭代  $I$  轮:

- ① 从  $L$  中基于  $F_{key}$  学习第 1 个分类器  $f_{key}$ ;
- ② 用  $f_{key}$  对  $U$  进行标注;
- ③ 从  $U$  中选出  $n_1$  个最确信的正类样本和负类样本组成集合  $S_{key}$ ;
- ④ 从  $L$  中基于  $F_{detail}$  学习第 2 个分类器  $f_{detail}$ ;
- ⑤ 用  $f_{detail}$  对  $U$  进行标注;
- ⑥ 从  $U$  中选出  $n_2$  个最确信的正类样本和负类样本组成集合  $S_{detail}$ ;
- ⑦ 从  $L$  中基于  $F_{full}$  学习第 3 个分类器  $f_{full}$ ;
- ⑧ 用  $f_{full}$  对  $U$  进行标注;
- ⑨ 从  $U$  中选出  $n_3$  个最确信的正类样本和负类样本组成集合  $S_{full}$ ;
- ⑩ 从  $U$  中去除集合  $S_{key} \cup S_{detail} \cup S_{full}$ ;
- ⑪ 把集合  $S_{key} \cup S_{detail} \cup S_{full}$  和相应的标注加入集合  $L$  中。

### 5 实 验

#### 5.1 实验设置

我们在 8 个领域的商品评论上进行了多组实验。前 4 个领域(book, DVD, electronic, kitchen)的商品评论是由 Blitzer 等人<sup>[16]</sup>从 <http://www.amazon.com/> 上采集的,后 4 个领域(network, software, pet, health)的商品评论是由 Li 等人<sup>[12]</sup>从 <http://www.amazon.com/> 上采集的。这 8 个领域的语料各含 1 000 篇正面评论和 1 000 篇负面评论。在实验中,本文选择贝叶斯分类器作为基础分类器,不进行特征选择和特征约简,使用所有单词作为特征。

在有监督的情感分类中,本文采用在全文上训练的分类器作为 Baseline。此外,如果一篇文本只有一个句子,则默认其为关键句。对于每一个领域,本文

选择 50% 有标注的数据作为训练集, 50% 有标注的数据作为测试集。

在半监督的情感分类中, 本文采用 10% 有标注数据作为训练集, 20% 有标注数据作为测试集, 70% 未标注数据作为添加集。有 4 组 Baseline 方法用于和本文的 Co-training 算法进行对比:

1) Self-learning. 使用 Baseline 的分类器作为基础分类器, 每轮迭代挑选最确信的样本加入训练集。

2) Transductive SVM. 学习机在训练过程中使用较少的有标签样本和较多的无标签样本进行学习, 测试集的样本分布信息从无标签样本转移到了最终的分类器中。

3) 基于全文分类器的 Co-training. 每轮迭代中, 由 3 个基础分类器来选择样本, 但是样本的最终标签由全文分类器来指定。

4) 基于随机关键句的 Co-training. 随机挑选关键句, 然后进行分类器融合, 每轮迭代中, 由 3 个基础分类器来选择样本, 由融合后的分类器来指定样本的最终标签。

## 5.2 实验结果

为了验证本文提出的关键句抽取算法的有效性, 我们将抽选出的关键句分别运用于有监督和半监督的情感分类中。在有监督学习中, 本文首先将分类器融合方法和 Baseline 进行对比, 实验结果如表 1 所示。

从表 1 可以看出, 将自动抽取出的关键句运用到分类器融合方法后, 分类正确率相比 Baseline 方法在 8 个领域中平均提高了 2.84 个百分点, 从而证明使用关键句可以提高有监督情感分类的性能。值得注意的是, 尽管关键句训练集的规模比全文训练集少很多, 但在关键句集合上训练的单分类器的平均性能却超过了 Baseline。在某些情况下, 关键句上

**Table 1 Comparison Results of Supervised Sentiment Classification**

**表 1 有监督情感分类对比实验结果**

| Domain     | Key Sentences Classifier | Detailed Sentences Classifier | Full-Text Classifier (Baseline) | Combined Classifier |
|------------|--------------------------|-------------------------------|---------------------------------|---------------------|
| Book       | 0.699                    | 0.691                         | 0.714                           | 0.742               |
| DVD        | 0.727                    | 0.729                         | 0.74                            | 0.773               |
| Electronic | 0.766                    | 0.741                         | 0.762                           | 0.780               |
| Kitchen    | 0.778                    | 0.792                         | 0.815                           | 0.837               |
| Network    | 0.812                    | 0.684                         | 0.739                           | 0.788               |
| Software   | 0.640                    | 0.645                         | 0.686                           | 0.715               |
| Pet        | 0.640                    | 0.623                         | 0.636                           | 0.650               |
| Health     | 0.659                    | 0.547                         | 0.568                           | 0.602               |
| Average    | 0.7151                   | 0.6815                        | 0.7075                          | 0.7359              |

训练的单分类器性能甚至可以超过多分类器融合的结果, 比如在“Network”和“Health”领域, 因为关键句分类器的特征空间更紧凑, 所以更容易对特定样本进行区分。

在所有分类器中, 细节句分类器的性能是最差的, 因为细节描述通常复杂多变而且存在歧义, 所以很难对整篇文档的极性进行正确预测。相反地, 关键句分类器却可以从另一个视角提供一些确信的知识, 因为它忽略了细节描述中的很多噪音。本文的关键句抽取算法可以将一篇文本划分成不同却互补的两部分, 所以适用于分类器融合的方法。

在半监督情感分类的实验中, 本文将在初始 10% 有标注数据上训练的分类器定为 Baseline, 并将 4 种 bootstrapping 类似方法与本文的 Co-training 算法进行比较, 实验结果如表 2 所示。迭代总共 100 轮, 在每一轮迭代中, 每一个类别选取 2 个最为确信的样本加入训练集, 即  $n_1 = n_2 = n_3 = 2$ 。

**Table 2 Comparison Results of Semi-Supervised Sentiment Classification**

**表 2 半监督情感分类对比实验结果**

| Domain     | Baseline | Self-Learning | Co-Training with Full-Text Classifier | Co-Training with Random View | Transductive SVM | Co-Training with Combined Classifier |
|------------|----------|---------------|---------------------------------------|------------------------------|------------------|--------------------------------------|
| Book       | 0.66     | 0.6875        | 0.6975                                | 0.6875                       | 0.6575           | 0.7225                               |
| DVD        | 0.575    | 0.7075        | 0.665                                 | 0.7025                       | 0.6525           | 0.6975                               |
| Electronic | 0.6875   | 0.685         | 0.7275                                | 0.6975                       | 0.6825           | 0.735                                |
| Kitchen    | 0.7025   | 0.775         | 0.7825                                | 0.790                        | 0.750            | 0.7975                               |
| Network    | 0.6825   | 0.7575        | 0.7475                                | 0.750                        | 0.755            | 0.7775                               |
| Software   | 0.6575   | 0.660         | 0.695                                 | 0.6825                       | 0.7475           | 0.7075                               |
| Pet        | 0.465    | 0.5875        | 0.5675                                | 0.5775                       | 0.500            | 0.605                                |
| Health     | 0.4125   | 0.405         | 0.520                                 | 0.4775                       | 0.410            | 0.5525                               |
| Average    | 0.6053   | 0.6581        | 0.6753                                | 0.6706                       | 0.6444           | 0.6994                               |

从表 2 可以看出,将关键句运用到 Co-training 算法后,分类器的正确率在 8 个领域中平均提高了 9.41 个百分点,效果非常显著.特别地,在“Pet”和“Health”领域,正确率甚至可以提高 14 个百分点.总体说来,在 5 种半监督学习方法中,基于关键句的 Co-training 算法效果最好.从实验中不难看出,把随机选取的关键句运用到 Co-training 算法后,效果不如基于全文分类器的 Co-training 算法,从而证明分类效果的提升并不是由于分类器融合,而是得益于正确的关键句的抽取.

本文提出的 Co-training 算法之所以能取得较好的性能,是因为关键句和细节句有着不同的特征空间,能被关键句分类器很好预测的样本不一定能被细节句分类器很好的预测,所以细节句分类器可以从这个样本上学习到有用的知识来提高自身性能,反之亦然.

## 6 结 论

本文引入了情感关键句这个概念,认为关键句有助于提高情感分类的性能,并且提出了一种新颖的关键句自动抽取算法.与此同时,本文还将抽取出的关键句分别应用于有监督和半监督的情感分类中.在有监督情感分类中,本文采用了分类器融合的方法来充分利用关键句和细节句之间的差异性和互补性.在半监督情感分类中,本文采用了 Co-training 算法,使得关键句分类器和细节句分类器可以互相学习互相促进.在 8 个领域中进行实验,结果表明引入了关键句后,分类器融合方法和 Co-training 算法的分类正确率都得到了显著提升.

## 参 考 文 献

[1] Du Weifu, Tan Songbo, Yun Xiaochun, et al. A new method to compute semantic orientation [J]. Journal of Computer Research and Development, 2009, 46(10): 1713-1720 (in Chinese)  
(杜伟夫, 谭松波, 云晓春, 等. 一种新的情感词汇语义倾向计算方法[J]. 计算机研究与发展, 2009, 46(10): 1713-1720)

[2] Wu Qiong, Tan Songbo, Xu Hongbo, et al. Cross-domain opinion analysis based on random-walk model [J]. Journal of Computer Research and Development, 2010, 47(12): 2123-2131 (in Chinese)  
(吴琼, 谭松波, 徐洪波, 等. 基于随机游走模型的跨领域倾向性分析研究[J]. 计算机研究与发展, 2010, 47(12): 2123-2131)

[3] Hu Yi, Lu Ruzhan, Li Xuening, et al. Research on language modeling based sentiment classification of text [J]. Journal of Computer Research and Development, 2007, 44(9): 1469-1475 (in Chinese)  
(胡熠, 陆汝占, 李学宁, 等. 基于语言建模的文本情感分类研究[J]. 计算机研究与发展, 2007, 44(9): 1469-1475)

[4] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques [C] //Proc of EMNLP. New York: ACM, 2002: 79-86

[5] Turney P. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews [C] //Proc of ACL. New York: ACM, 2002: 417-424

[6] Dasgupta S, Ng V. Mine the easy and classify the hard: A semi-supervised approach to automatic sentiment classification [C] //Proc of ACL. New York: ACM, 2009: 701-709

[7] Yessenalina A, Yue Y, Cardie C. Multi-level structured models for document-level sentiment classification [C] //Proc of EMNLP. New York: ACM, 2010: 1046-1056

[8] Gamon M. Sentiment classification on customer feedback data; Noisy data, large feature vectors, and the role of linguistic analysis [C] //Proc of the Int Conf on Computational Linguistics. New York: ACM, 2004

[9] Melville P, Gryc W, Lawrence R. Sentiment analysis of blogs by combining lexical knowledge with text classification [C] //Proc of SIGKDD. New York: ACM, 2009

[10] Li Shoushan, Huang Churen, Zhou Guodong, et al. Employing personal/impersonal views in supervised and semi-supervised sentiment classification [C] //Proc of ACL. New York: ACM, 2010: 414-423

[11] Taras Z, Carroll J. Automatic seed word selection for unsupervised sentiment classification of Chinese text [C] //Proc of the Int Conf on Computational Linguistics. New York: ACM, 2008: 1073-1080

[12] Li Tao, Zhang Yi, Vikas Sindhwani. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge [C] //Proc of ACL. New York: ACM, 2009

[13] Wan Xiaojun. Co-training for cross-lingual sentiment classification [C] //Proc of ACL. New York: ACM, 2009

[14] Becker I, Aharonson V. Last but definitely not Least: On the role of the last sentence in automatic polarity-classification [C] //Proc of ACL. New York: ACM, 2010: 331-335

[15] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training [C] //Proc of COLT. New York: ACM, 1998: 92-100

[16] Blitzer J, Dredze M, Pereira F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification [C] //Proc of ACL. New York: ACM, 2007: 440-447



**Lin Zheng**, born in 1984. PhD candidate. Her main research interests include natural language processing and sentiment analysis.



**Cheng Xueqi**, born in 1971. Professor and PhD supervisor. His main research interests include network information security, large-scale information retrieval and knowledge mining (cxq@ict.ac.cn).



**Tan Songbo**, born in 1978. PhD and associate professor. His main research interests include sentiment analysis (tansongbo@software.ict.ac.cn).

## 2013 年《计算机研究与发展》专题(正刊)征文通知

### ——“物联网的基础理论与新技术方向”

近年来,针对物联网的基础理论、核心技术、系统应用及产业化的研究与深入发展一直是国际国内学术与产业界集中关注的热点。

传感网物联网在国内受到极大的重视,近几年在全国各地开展的物联网建设实践活动也方兴未艾,并有越来越多的学者和学生参与到这个领域的研究中,人们对物联网的认识以及对其技术和理论方面尚存问题的认识也不断深入。与此同时,各类新型电子产品的不断涌现(比如智能手机等),也为物联网的应用提供了新的技术依托手段和日新月异的可能性,展现了更广阔更引人入胜的服务模式和未来商业模式,也极大地扩展和丰富了物联网研究领域的学术及技术范畴。

《计算机研究与发展》将在 2013 年 6 月出版“物联网的基础理论与新技术方向”专题,本专题以探讨理论研究和新技术发展如何真正支持未来物联网的大规模推广及普及应用为核心,以突出物联网相关基础理论和未来新技术发展方向为特色,重点探索与未来物联网相适应的新的计算理论、围绕物联网发展及应用特点而形成的未来新技术方向及物联网专用技术增长点,期望对未来的物联网研究、技术发展以及相关实践活动做出有意义的讨论和有价值和建设性的引导。

本专题将针对国内外物联网的基础理论与新技术方面具有创新性、突破性的研究成果,面向国内外征集论文,欢迎广大专家、学者、工程技术人员和研究生踊跃投稿。

#### 征文范围

本专题主要征集原创性研究论文和高质量的综述论文,包括(但不限于)以下内容:

- 1) 新型计算理论和物联网体系结构:人机物融合的计算理论、新型物联网体系结构及其组网关键技术、物联网中间件、支持环境和操作系统、物联网资源访问、管理与服务机制;
- 2) 物联网的设计方法与核心技术:物联网系统软件体系架构与开发技术、新型物联网系统编程模型与编程方法、物联网编程语言及其支撑运行环境、物联网系统评价模型与仿真模型及方法、新型智能终端、物联网新设备及其感知执行技术;
- 3) 移动计算和物联网的服务机制:物联网移动计算原理和技术、采用移动设备(如智能手机等)的综合感知技术及智能协同系统、以人为中心的感知及智能决策和智能协助系统、物联网应用服务机制和商业模式;
- 4) 物联网的多源异构数据融合及大数据处理:多源异构物联网的感知融合理论、计算模型和方法、协同计算理论与方法、大数据的存储、融合和挖掘的方法与技术;
- 5) 物联网的安全理论:物联网中的隐私保护、物联网的安全机制;
- 6) 物联网应用系统技术:新型行业应用系统构建技术与实例、新型物联网行业应用及系统部署和运行经验、物理模型与行业应用模型的设计与实现、系统支撑共性技术与行业应用的融合方法。

#### 投稿要求

- 1) 来稿应属于作者的科研成果,数据真实可靠,具有重要的学术价值与推广应用价值,未在国内外公开发行的刊物或会议上发表或宣读过,不存在一稿多投问题。作者在投稿时,需向编辑部提交投稿声明。
- 2) 论文一律用 word 格式排版,论文格式体例参考近期出版的《计算机研究与发展》的要求(<http://crad.ict.ac.cn/>)。
- 3) 论文通过期刊网站(<http://crad.ict.ac.cn/>)投稿,投稿时提供作者的联系方式,备注中务必注明“物联网 2013 专题”。(否则按自由来稿处理)。

#### 重要时间

截稿日期:2012 年 12 月 31 日

最终结果通知日期:2013 年 3 月 30 日

作者修改版提交日期:2013 年 4 月 10 日

出版日期:2013 年 6 月

特约编委 崔莉 研究员 中国科学院计算技术研究所 lcui@ict.ac.cn

#### 联系方式

北京 2704 信箱 《计算机研究与发展》编辑部 100190

电话:010-62620696, 010-62600350

E-mail: crad@ict.ac.cn