文章编号:1001 - 9081(2010)04 - 1015 - 04

基于属性选择的半监督短文本分类算法

蔡月红[1,2],朱 倩[1],孙 萍[1],程显毅[1]
(1. 江苏大学 计算机科学与通信工程学院,江苏 镇江 212013; 2. 江苏大学 外语学习中心,江苏 镇江 212013)
(caiyh@ ujs. edu. cn)

摘 要:针对海量短文本分类中的标注语料匮乏问题,提出了一种基于属性选择的半监督短文本分类算法。通过基于 ReliefF 评估和独立性度量的属性选择技术选出部分具有较好的属性独立关系的属性参与分类模型的学习,以弱化朴素贝叶斯模型的强独立性假设条件;借助集成学习,以具有一定差异性的分类器组去估计初始值,并以多数投票策略去分类未标注语料集,以减低最大期望算法(EM)对于初始值的敏感。通过真实语料上进行的比较实验,证明了该方法能有效利用大量未标注语料提高算法的泛化能力。

关键词:属性选择;半监督学习;短文本;文本分类;集成学习
中图分类号: TP391 **文献标志码**:A

Semi-supervised short text categorization based on attribute selection

CAI Yue-hong[1,2], ZHU Qian[1], SUN Ping[1], CHENG Xian-yi[1]
(1. *School of Computer Science and Communication Engineering, Jiangsu University, Zhengjiang Jiangsu* 212013, *China*;
2. *Foreign Language Learning Center, Jiangsu University, Zhengjiang Jiangsu* 212013, *China*)

Abstract: In order to solve the data scarcity problem of massive short text categorization, a semi-supervised short text categorization method based on attribute selection was presented. An attribute selection algorithm based on ReliefF and independence measures was used to overcome the limitation of the attributes independence assumption by deleting irrelevant or redundant attributes, and an ensemble algorithm based on Expectaion-Maximization (EM) was used to resolve the problems of sensitivity to initial values in semi-supervised EM algorithm. The experiments on real corpus show that the proposed method can more effectively and stably utilize the unlabeled examples to improve classification generalization.

Key words: attribute selection; semi-supervised learning; short text; text categorization; ensemble learning

0 引言

全面资讯时代,电子邮件、即时通信、社区论坛、博客、在线聊天室、手机短信已成为主要的信息交流方式。在每天产生的大量信息流中,存在着大量长度很短的文本数据。海量的短文本信息的传播使得信息传播格局正在发生深刻变化。因此短文本信息挖掘技术在舆情监控、话题识别与跟踪、流行语分析等领域有着广泛的应用前景。分类分析是信息挖掘技术中最基本和最重要的方法之一。

目前的短文本分类算法[1-2]大多是基于监督学习的。监督学习必须对所有的学习样本做类别标记,而且为了保证泛化能力,通常需要大量的标注样本作训练集。语料库的人工标注是很费时费力的,而大量的未标注语料却很容易获取,这就是所谓的标注瓶颈问题。由于短文本的不规范性和海量性,在短文本分类中标注语料匮乏问题尤为突出。因此如何整合已标注样本和未标注数据的学习,成为短文本分类中一个现实的问题。

半监督学习是综合利用已标注数据和未标注数据的主流学习技术之一。现有的半监督学习算法可分为四大类:生成式模型[3]、基于图的方法[4]、协同训练算法[5]和直推式学习[6]。其中生成式模型在文本分类中得到了广泛应用,文献[3]首先将最大期望算法(Expectaion-Maximization, EM)用于结合已标注语料和未标注语料的文本分类研究,利用测试样本改进了 Bayes 分类器的分类效果,一般称为 EM-NB 算法。但这种方法的不足之处在于朴素贝叶斯(Naive-Bayes, NB)模型的强独立性假设和 EM 算法对于初始值的敏感。这些局限性影响了 EM-NB 算法的合理性,使得最终的分类效果受到影响。

本文提出了一种基于属性选择的半监督 EM 短文本分类算法(AS-EM),通过在半监督 EM 学习算法中嵌入基于属性选择的集成学习框架来提高性能,构造了一个鲁棒的、准确度高的半监督短文本分类算法。

1 半监督 EM 文本分类算法

1.1 朴素贝叶斯文本分类器

假定文档集 $D = \{(d_1, l_1), (d_2, l_2), \cdots, (d_{|D|}, l_{|D|})\}$,其中 d_i 为属性变量,$l_i \in C = \{c_1, c_2, \cdots, c_{|C|}\}$ 为类变量,特征集合 $W = \{w_1, w_2, \cdots, w_{|W|}\}$。文本分类时,通过计算文本 d_i 的后验概率 $P(c_j \mid d_i)$ 最终将 d_i 分到后验概率概率最大的类别中,根据贝叶斯公式:

$$P(c_j \mid d_i; \hat{\theta}) = \frac{P(c_j \mid \hat{\theta}) \prod_{k=1}^{|d_i|} P(w_{d_{i,k}} \mid c_j; \hat{\theta})}{\sum_{r=1}^{|d_i|} P(c_r \mid \hat{\theta}) \prod_{k=1}^{|d_i|} P(w_{d_{i,k}} \mid c_r; \hat{\theta})} \quad (1)$$

收稿日期:2009 - 10 - 12;**修回日期**:2009 - 12 - 07。 **基金项目**:国家自然科学基金资助项目(60702056)。

作者简介:蔡月红(1969 -),女,江苏兴化人,工程师,硕士研究生,主要研究方向:自然语言处理、人工智能; 朱倩(1979 -),女,江苏镇江人,讲师,博士研究生,主要研究方向:自然语言处理、机器学习、模式识别; 孙萍(1983 -),女,山东菏泽人,硕士研究生,主要研究方向:自然语言处理、人工智能; 程显毅(1956 -),男,黑龙江哈尔滨人,教授,博士生导师,主要研究方向:模式识别、自然语言理解。

其中：

$$P(w_t \mid c_j;\hat{\theta}) = \frac{1+\sum_{i=1}^{|D|}N(w_t,d_i)P(c_j \mid d_i)}{|W|+\sum_{s=1}^{|W|}\sum_{i=1}^{|D|}N(w_s,d_i)P(c_j \mid d_i)} \quad (2)$$

$$P(c_j \mid \hat{\theta}) = \frac{1+\sum_{i=1}^{|D|}P(c_j \mid d_i)}{|C|+|D|} \quad (3)$$

其中：$|d_i|$ 为文本 d_i 的长度；$N(w_t,d_i)$ 表示文本 d_i 中特征 w_t 出现的次数；$P(c_j \mid d_i) \in \{0,1\}$。

1.2 EM-NB 算法

基于完全数据概率分布（例如朴素贝叶斯假设的概率模型）建模，再利用 EM 算法对模型参数估计是半监督学习中的一种重要方法。假设 D^L 和 D^U 分别代表已标注的样本集和未标注的样本集。EM-NB 算法将未标注样本与朴素贝叶斯的学习像结合：首先采用已标注样本集 D^L 作为训练集进行初始化训练，得到第一个中间分类器 $\hat{\theta}$；接下来在 E-Step 中利用现有的分类器 $\hat{\theta}$ 和式(1) 估计未标注样本集 D^U 中样本的类别；随后在 M-step 中将估计出类别的样本加入到训练集中，利用式(2)、(3) 训练新的中间分类器 $\hat{\theta}$。迭代地重复 EM 步骤，直至 $\hat{\theta}$ 收敛。

2 基于属性选择的半监督文本分类算法

由于实际数据一般很难完全满足 NB 模型理论上的假设条件，所以 EM-NB 算法对数据比较敏感，平均分类错误率较大，容易陷入局部最优。提高 EM-NB 算法的性能，如何放松 NB 的属性独立性假设是个关键问题。研究人员在 NB 的属性独立性假设放松方面进行了很多工作，使用的方法主要有：

1）通过属性选择选出部分属性参与分类模型的学习，在选出的具有较好属性独立关系的属性子集上构建贝叶斯分类器，即选择性贝叶斯分类器[7]。

2）修正贝叶斯分类器所使用的概率估计[8]。

3）将若干贝叶斯分类器集成起来以提高其分类性能[9]等。本文中为了提高贝叶斯分类能力，采用基于属性选择的贝叶斯分类器的集成来放松属性独立性假设。

2.1 基于 ReliefF 和独立性度量的属性选择算法

属性选择即从输入特征集合中选择使某种评估标准最优的属性子集，在选出的具有较好的属性独立关系的属性子集上构建分类器能有效地提高分类器的性能。主要的属性选择方法有：嵌入式、包装式（Wrapper）和过滤式（Filter）。ReliefF 算法[10]是公认的性能较好地用于多类别数据的过滤式特征评估方法，因为运行效率较高而适用于大规模数据集，但 ReliefF 算法不能去除冗余特征。为克服这个缺点，本文设计了基于 ReliefF 评估和独立性度量的属性选择算法（MID-ReliefF）。算法首先采用 ReliefF 去除无关特征，然后再采用特征间独立性分析的方法，去除冗余特征。特征间的独立性度量采用条件互信息。

2.1.1 独立性度量

在信息论中常用互信息来衡量两个随机变量之间的统计独立关系。如果设 $p(x_i, y_j \mid c_k)$ 为特征 x 取值 x_i，特征 y 取值为 y_i 时在 c_k 类文本中共同出现的条件概率，$p(x_i \mid c_k)$ 为特征 x 取值 x_i 时在 c_k 类文本中出现的条件概率，则 x_i、y_i 的条件互信息定义如下：

$$I(x_i,y_j \mid c_k) = \mathrm{lb}\frac{p(x_i,y_j \mid c_k)}{p(x_i \mid c_k)p(y_j \mid c_k)} \quad (4)$$

定义 1 条件独立性。设 x 和 y 是两个特征，在考虑类别信息的情况下，两个特征的条件独立性由平均条件互信息 $MI(x,y \mid C)$ 来度量，计算公式如下：

$$MID(x,y) = MI(x,y \mid C) = \sum_{i,j,k}p(x_i,y_j \mid c_k)\mathrm{lb}\frac{p(x_i,y_j \mid c_k)}{p(x_i \mid c_k)p(y_j \mid c_k)} \quad (5)$$

$MID(x,y)$ 反映了 x 与 y 之间的独立程度。为了使独立性度量取值在[0,1]，参照文献[11]的方法对式(5) 进行归一化处理：

$$SUMID(x,y) = 2\left[\frac{MID(x,y)}{H(x \mid C)+H(y \mid C)}\right] \quad (6)$$

其中 $H(x \mid C) = \sum_{i,k}p(x_i \mid c_k)\mathrm{lb}(x_i \mid c_k)$ 为 x 在考虑类别信息情况下的信息熵。

$SUMID(x,y)$ 的值为 0 表示两个特征间独立，$SUMID(x,y)$ 的值为 1 则表示两个特征完全冗余。

2.1.2 MID-ReliefF 算法

MID-ReliefF 算法首先运行 ReliefF 算法，计算每个特征的权重，由权值大于设定阈值 δ 的特征组成一个相关特征集合 S_{list}，然后将集合 S_{list} 中的特征按照权重从大到小排序；把权重最大的特征加入初始状态为空的目标集合 S_{goal} 中，从有序集合 S_{list} 中取下一个特征，将该特征与集合 S_{goal} 中的特征进行独立性度量，如果冗余度小于阈值 Ct，则将特征加入到目标集合 S_{goal}，否则将该特征看作相对于 S_{goal} 的冗余特征，继续取下一个特征，反复循环直到 S_{list} 为空。MID-ReliefF 算法的详细步骤见算法 1。由于算法中未涉及到分类算法，因此该算法可视为双层 Filter 模式。

算法 1 MID-ReliefF 算法

输入：训练数据集 Tr，测试数据集 Te，初始特征集 S，Relief F 过滤阈值 δ，相关度阈值 Ct。

输出：特征子集 S_{goal}。

```
1.  S_list = NULL, S_goal = NULL;
2.  Relief F( Tr );                    //得到特征权值 W = {W_i}
3.  for i := 1 to n do begin
4.    if ( W_i > δ )
5.      append f_i to S_list;          //得到新的特征集合 S_list
6.    end if
7.  end for
8.  SortByDesc(S_list) ;               //按权值从大到小排序
9.  S_goal←f_i = getFisrtElement (S_list);
                                       //第一个权值最大的特征到 S_goal 中
10. f_j = getNextElement (S_list, f_i) ;
11. while ( f_j <> NULL)
12.   for each f_i in S_goal
13.     if SUMID(f_i, f_j) ≥ Ct then
14.       remove f_j from S_list; jump to 18;
15.     end if
16.   end for
17.   S_goal←f_j;
18.   f_j = getNextElement (S_list, f_j);
19. end while
```

2.2 基于属性选择的半监督 EM 算法

为了改进 EM-NB 算法的性能，可以考虑以下两方面：1）放松朴素贝叶斯模型的强独立性假设；2）改善 EM 算法对于初始值的敏感。针对这两点，本文通过在 EM 训练框架嵌入基于属性选择的贝叶斯分类器集成，提出了一种基于属性选择集成的半监督 EM 算法（AS-EM）。首先利用 bootstrap 生成特征子集，然后在子集上运用 MID-ReliefF 算法进行特征选择，在选出的具有较好属性独立关系的属性子集上构建贝叶斯分类器以弱化 NB 的强独立性假设条件；并借助集成学习，以具有一定差异性的分类器组去估计初始值，并以多数投票策略去分类未标注语料集，以减低 EM 算法对于初始值的敏

感。算法具体步骤如下:

算法 2　AS-EM 算法。

输入:已标注语料集 L,未标注语料集 U, 个体子模型的个数 N。

输出:分类结果。

初始化:

1)for $i = 1:N$

①运用 bootstrap 取样在 L 上产生训练子集 S_i;

②利用特征选择从训练子集 S_i 中选取其最优特征子集,从而得最优训练子集 $S_{i-optional}$;

③在最优训练子集上训练个体贝叶斯子模型 h_i;

end

循环:

2)E-Step:使用当前分类器组 $h_i, i = 1,2,\cdots,i$ 以多数投票策略分类未标注语料集 U,并计算未标注样本集中各样本最大后验概率;

3)M-Step:将估计出类别的样本加入到训练集中利用最大后验概率统计更新训练分类器 $h_i, i = 1,2,\cdots$;

直到收敛或参数变化小于某一给定的阈值为止。

3　实验结果及分析

3.1　实验语料

语料是从天涯、中国大学生论坛、西祠等 BBS 网站人工采集的共约 5 591 篇论坛帖子,时间跨度为 2008 年 12 月到 2009 年 3 月,并按版面名称分为教育、就业、IT、音乐、体育、经济、文学、生活、游戏、留学等 10 个类别,将该语料记为 BBS-corpus。对每类数据,都从中取出 25% 作测试集,其余的则按已标注样本的不同比例(分别为 20% 、40% 、60%)划分为已标注语料集 L 和未标注语料集 U。

3.2　性能评测

实验中的性能评侧指标采取了综合指标 F_1,综合所有类别的评测指标为宏平均 $MacroF_1$。

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \tag{7}$$

$$MacroF_1 = \frac{\sum_{k=1}^{K} F_{1,k}}{K} \times 100\% \tag{8}$$

其中:

$$Precision = \frac{\text{正确分为某类的文本数}}{\text{测试集中分为该类的文本总数}} \times 100\%$$

$$Recall = \frac{\text{正确分为某类的文本数}}{\text{测试集中属于该类的文本总数}} \times 100\%$$

3.3　实验结果及分析

为了研究属性选择对 EM 算法性能的影响,将本文提出的 AS-EM 算法与 Bagging-EM、EM-NB、NB 三种算法进行了比对实验。其中 Bagging-EM 是基于朴素贝叶斯分类器 Bagging 集成的 EM 算法,在生成属性子集时只简单地运用了 bootstrap 技术。本实验在 BBS-corpus 语料集上进行了充分的实验,对每个给定的已标注样本比例产生 3 个不同的 L 和 U 随机划分,取 3 次独立运行的实验的平均值作为最终实验结果。

表 1 ~3 是不同的已标注样本比例下,4 种算法分类质量的统计比较。其中 F_1 值是迭代训练结束后联合假设的 F_1 值。

从表 1 ~3 可以看出,三种半监督 EM 算法中,AS-EM 算法的性能最好。在已标注样本比例为 20% 的情况下,AS-EM 算法的 $MacroF_l$ 比 EM-NB 提高 4.54% ,与 NB 相比提高了 8.16% ,这充分说明了基于属性选择的集成框架对提高半监督学习的泛化性能显著有效。而 Bagging-EM 与 EM-NB 相比,没有明显的改善,在有的实验中 $MacroF_l$ 甚至降低。这说明 Bagging 集成对稳定的学习算法效果不明显。

表 1　不同算法分类 F_1 比较(已标注样本比例 = 20%)

类别	NB	EM-NB	Bagging-EM	AS-EM
教育	61.13	61.00	60.94	64.80
就业	65.19	65.31	65.60	69.66
IT	79.71	84.01	84.12	87.78
音乐	80.72	83.25	83.52	86.54
体育	72.53	76.44	76.33	78.57
经济	67.41	69.52	69.35	71.79
文学	77.79	84.27	84.18	87.59
生活	60.26	61.61	61.58	63.69
游戏	85.95	87.96	88.18	94.03
留学	75.66	78.19	78.21	81.26
$MacroF_l$	72.64	75.16	75.20	78.57

表 2　不同算法分类 F_1 比较(已标注样本比例 = 40%)

类别	NB	EM-NB	Bagging-EM	AS-EM
教育	58.46	59.5	59.38	61.06
就业	65.94	67.98	67.98	70.56
IT	74.88	79.51	79.36	81.96
音乐	83.63	86.96	86.96	89.5
体育	75.26	77.64	78.33	80.76
经济	67.14	69.14	69.14	71.44
文学	74.17	79.63	79.49	82.62
生活	59.08	60.14	60.08	62.01
游戏	84.83	90.58	90.58	93.72
留学	78.97	81.71	81.24	84.26
$MacroF_l$	72.24	75.28	75.25	77.79

表 3　不同算法分类 F_1 比较(已标注样本比例 = 60%)

类别	NB	EM-NB	Bagging-EM	AS-EM
教育	61.92	70.61	70.40	71.50
就业	71.55	74.27	74.55	76.37
IT	81.05	83.97	83.64	85.27
音乐	82.12	85.49	85.41	86.82
体育	74.15	76.97	77.12	78.39
经济	67.00	67.00	67.20	68.40
文学	79.33	83.05	82.90	86.02
生活	60.26	64.66	64.55	66.58
游戏	85.67	89.01	89.01	90.48
留学	73.85	74.96	74.96	76.13
$MacroF_l$	74.36	77.00	76.97	78.60

文中图 1 ~图 3 给出不同已标注比例下三种 EM 算法的 $MacroF_l$ 值随训练过程的迭代变化情况。需要指出的是每个算法每次训练结束的迭代次数都不相同,为了便于比较 $MacroF_l$ 值变化过程,考虑到三种算法训练结束时迭代次数的不同,图中横轴最大值取不同算法训练结束时迭代的最大次数,训练结束后相应的 $MacroF_l$ 值保持不变。

从图 1 ~图 3 可以看出:

1)所有情形下,三种算法所得最终 $MacroF_l$ 值都高于初始假设,这说明三种半监督 EM 算法都能利用未标注样本提高泛化性能。其中,AS-EM 算法对泛化性能提高最显著

2)对比算法各自迭代过程,发现 AS-EM 和 Boosting-EM 的宏平均调和率基本是持续升高直至迭代结束达到最高值,几乎不会出现 EM-NB 的波动现象,这说明了集成框架能有效地减低 EM 算法对于初始值的敏感,融合集成学习框架的半

监督 EM 算法每次迭代受误标记噪声影响较小,能更稳定地保证泛化性能的提高,有更好的健壮性。

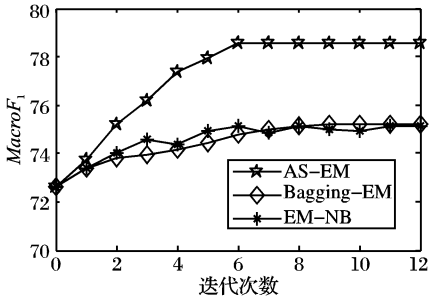

图 1 已标注样本比例 = 20%

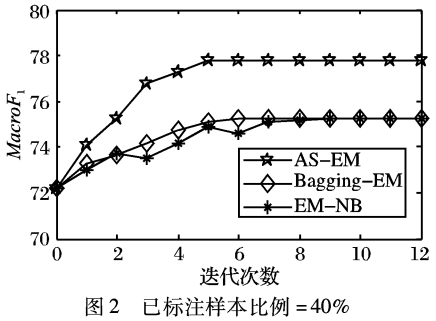

图 2 已标注样本比例 = 40%

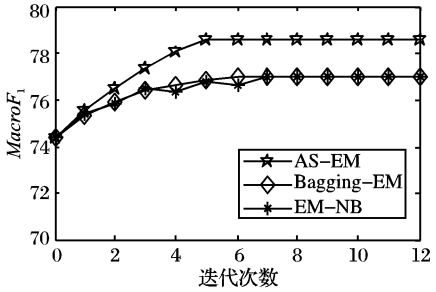

图 3 已标注样本比例 = 60%

4 结语

本文提出了基于属性选择集成的半监督 EM 算法(AS-EM)。该算法的改进之处在于:不是采用单个的贝叶斯分类器在已标注样本集上得到似然函数参数的初始值,而是利用 bootstrap 进行子集生成,然后运用基于 ReliefF 评估和独立性度量的属性选择算法在子集上进行属性选择,再在选择后的子集上分别建立子模型;并且在 E-Step 中,使用当前分类器组以多数投票策略分类未标注语料集。实验证明 AS-EM 算法可有效地利用大量未标注语料提高泛化性能,并在性能和效率上明显优于 EM-NB 算法。

参考文献:

[1] 宁亚辉, 樊兴华, 吴渝. 基于领域词语本体的短文本分类[J]. 计算机科学, 2009, 36(3): 142 - 145.

[2] 王细薇, 樊兴华, 赵军. 一种基于特征扩展的中文短文本分类方法[J]. 计算机应用, 2009, 29(3): 843 - 845.

[3] NIGAM K, MCCSLLUM A K, THRN S, *et al.* Text classification from labeled and unlabeled documents using EM[J]. Machine Learning, 2000, 39(2/3): 103 - 134.

[4] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, 7: 2399 - 2434.

[5] ZHOU Z H, LI M. Semi-supervised regression with co-training style algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(11): 1479 - 1493.

[6] WANG Y, SHANG S T. Training TSVM with the proper number of positive samples[J]. Pattern Recognition Letters, 2005, 26(14): 2187 - 2194.

[7] 陈景年, 黄厚宽, 田凤占, 等. 用于不完整数据的选择性贝叶斯分类器[J]. 计算机研究与发展, 2007, 44(8): 1324 - 1330.

[8] WAN Z H, WEBB G I, ZHENG F. Adjusting dependence relations for semi-lazy TAN classifiers[C]// Advances in Artificial Intelligence. Berlin Heidelberg: Springer-Verlag, 2003: 453 - 456.

[9] 石洪波, 黄厚宽, 王志海. 基于 Boosting 的 TAN 组合分类器[J]. 计算机研究与发展, 2004, 41(2): 340 - 345.

[10] ROBNIK-ŠIKONJA M, KONONENKO I. Theoretical and empirical analysis of ReliefF and RreliefF[J]. Machine Learning, 2003, 53(1 - 2): 23 - 69.

[11] YU L, LIU H. Efficient feature selection via analysis of relevance and redundancy[J]. Journal of Machine Learning Research, 2004, 5: 1205 - 1224.

(上接第 1010 页)

BAQP 能较好地解决当前 MOOC 系统面临的负载平衡问题,从整体降低系统成本。该处理器根据资源状态和查询信息动态建立资源负载耗费矩阵,并结合 Min-Min 的执行高效性和 Max-Min 的负载均衡性,提出新的 A-MM 查询调度算法。不管取静态权值,还是动态权值,该算法都能较好地降低系统负载。以后要改进的工作主要集中在将 BAQP 组件移植到“国家精品课程”项目中,进行实际性能评测。

参考文献:

[1] PRESTON J, BOOTH L, CHASTINE J. Improving learning and creating community in online courses via MMOG technology[C]// Proceedings of the 35th SIGCSE Technical Symposium. Norfolk: ACM Press, 2004, 3: 175 - 180.

[2] VLAD N, ALEXANDRU I, RADU P, *et al.* Efficient management of data center resources for massively multiplayer online games[C]// International Conference on High Performance Computing, Networking, Storage and Analysis. Austin: IEEE Computer Society Press, 2008: 212 - 244.

[3] ZHAO C, CAO J, WU H, *et al.* Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications [M]. [S.l.]: IGI Publishing, 2009.

[4] LIU SHUO, SKARIMI L. Grid query optimizer to improve query processing in grids [J]. Future Generation Computer Systems, 2008, 24(5): 342 - 353.

[5] ZHOU Y, TAN K. An adaptable distributed query processing architecture [J]. Data & Knowledge Engineering, 2005, 53(3): 283 - 309.

[6] BLYTHE J, JAIN S, DEELMAN E, *et al.* Task scheduling strategies for workflow based applications in grids[C]// CCGrid 2005: IEEE International Symposium on Cluster Computing and the Grid. Los Alamitos: IEEE Computer Society, 2005, 2: 759 - 767.

[7] MAHESWARAN M, ALI S, SIEGEL H, *et al.* A comparison of dynamic strategies for mapping a class of independent tasks onto heterogeneous computing systems[C]// Eighth Heterogeneous Computing Workshop. [S.l.]: IEEE, 1999: 30 - 44.

[8] 怀进鹏, 胡春明, 李建欣. CROWN: 面向服务的网格中间件系统与信任管理[J]. 中国科学: E 辑(信息科学), 2006, 36(10): 1127 - 1155.

[9] HOU YONG, YU JIONG, TURGUN. NDA-MM: A new adaptive task scheduling algorithm based on the non-dedicated constraint grid [C]// IEEE Transactions on Sixth International Conference on Grid and Cooperative Computing. Wulumqi: IEEE, 2007: 275 - 281.