第 38 卷 第 3 期
Vol.38 No.3

计 算 机 工 程
Computer Engineering

2012 年 2 月
February 2012

·人工智能及识别技术· 文章编号：1000—3428(2012)03—0183—04 文献标识码：A 中图分类号：TP18

基于关键词抽取的自动文摘算法

蒋效宇

(北京服装学院商学院，北京 100029)

摘 要：针对生成文摘内容不完整的问题，利用相邻词的共现频率进行未登录词识别，提出一种通过词汇链的构建进行中文关键词抽取和文摘生成的算法，并给出一种采用《知网》为知识库构建词汇链的方法。通过计算词义相似度构建词汇链，结合词汇所在词汇链的强度、信息熵和出现位置等属性，进行关键词抽取和句子重要度计算。实验结果表明，与已有算法相比，该算法能够提高生成摘要的召回率和准确率。

关键词：自动文摘；向量空间模型；关键词抽取；词汇链；未登录词识别

Automatic Summarization Algorithm Based on Keyword Extraction

JIANG Xiao-yu

(Business School, Beijing Institute of Fashion Technology, Beijing 100029, China)

【Abstract】 In order to over the shortcoming of the incomprehensive of summarization, a new lexical chain-based keywords extraction and automatic summarization algorithm from Chinese texts based on the unknown word recognition using co-occurrence of neighbor words is proposed, and an algorithm for constructing lexical chain based on Hownet knowledge database is given in the method, lexical chain is constructed by calculating the semantic similarity between terms, keywords are extracted and the importance of each sentence is calculated according to the intensity of lexical chain, the entropy of terms and position. Experimental results show that the summarization generated by the improved algorithm gets better performance than other methods both in recall and precision.

【Key words】 automatic summarization; vector space model; keyword extraction; lexical chain; unknown word recognition

DOI: 10.3969/j.issn.1000-3428.2012.03.062

1 概述

随着各种媒介信息的快速增长，自动文摘已经成为处理日益增长海量信息的一个有效手段。所谓自动文摘就是利用计算机自动地从原始文献中生成准确全面反映文献中心内容、语言简洁连贯的摘要。目前常用算法有关键词词频统计法[1]、文档聚类法[2]、MMR 算法[3]等，但是这些算法只是通过计算词汇和句子的权重，简单地将那些权重大的句子提取出来组成文摘，从而导致文摘内容不简洁不全面。

本文鉴于关键词是自动文摘的一个特例，如果在事先知道文档关键词的情况下进行自动文摘，势必能够进一步提高摘要的质量[4]，因此提出了一种基于词汇链的关键词抽取的算法，并在关键词的基础上进行文摘生成和冗余消除，以此提高文摘内容的全面性。

2 关键词抽取

2.1 未登录词识别

本文采用的分词算法是中科院计算所的 ICTCLAS，但在实验过程中，发现分词后的文本中有很多个连续的单字独立成词，但经过观察分析，这些连续的单字往往是一些未登录词(领域术语等)，而这些未登录词对于关键词抽取有着重要影响，如果不能正确识别，则将会大大降低了关键词抽取的准确率。针对上述的不足，本文利用了相邻词[5]的共现频率进行未登录词的识别。

词 t 的相邻词是指对句子进行分词后，在 t 之前的一个词和之后的一个词。例如对“中文关键词抽取算法”分词后得到“中文/关键词/抽取/算法”，那么“关键词”的相邻词即为：“中文”(称为“前邻”)和“抽取”(称为“后邻”)。显然，由于词可能出现在句首或者句尾，因此前邻和后邻有可能是空。对文档中每个词 t 的相邻词的频繁程度进行考察，从而判断词 t 及其相邻词是否需要合并以成为语义完整的未登录词。未登录词识别的具体算法如下：

Step1 利用 ICTCLAS 对文档 d 进行分词和词性标注，去除停用词后，将剩余的所有词都加入集合 W 中。

Step2 I=0。

Step3 I=I+1，若 I 大于阈值 γ，则转至 Step7；否则，对词集合 W 中每个词 s 统计出它在文档 d 中的前邻和后邻的分布，并根据某种策略判断是否具备频繁前邻和频繁后邻。例如可以根据某个前邻 PR(前邻和后邻均不包含单字虚词)出现概率是否大于预定阈值 ζ 来认定它是一个频繁前邻。γ 通常取 4 或者 5，ζ 取 0.6。

Step4 若 s 具有频繁前邻 PR 和频繁后邻 BE，则将 PR+s+BE 拼成一个词加入未登录词候选集合 W^*中。

Step5 若 s 仅具有频繁前邻 PR，则将 PR+s 拼成一个新词加入词集合 W_{PR} 中。

Step6 若 s 仅具有频繁后邻 BE，则将 s+BE 拼成一个新词加入词集合 W_{BE} 中。

基金项目：北京市优秀人才培养资助专项科研基金资助项目(2009 D005001000005)

作者简介：蒋效宇(1979－)，男，副教授、博士，主研方向：人工智能，自动文摘

收稿日期：2011-06-03 **E-mail：**sxyjiangxiaoyu@bift.edu.cn

Step7 将 W_{PR} 和 W_{BE} 中共同出现的词加入 W^*，清空 W_{PR} 和 W_{BE}。

Step8 令 $W=W^*$，清空 W^*，转至 Step3。

Step9 利用未登录词候选集合 W^* 进行二次分词。

通过上述算法能够提高未登录词的识别率和分词的准确率，更利于统计词频和文档关键词的抽取。实验证明这种未登录词识别的方法是很有效的。

2.2 词汇链构建

文献[6]根据英文词汇间的关系，提出了基于 WordNet 的词汇链计算模型以及词汇链的生成算法，该算法针对英文，采用 WordNe 作为知识库，仅选择出现在 WordNet 中的所有名词作为候选词进行词汇链的构建。与之不同，本文采用了《知网》作为知识库来确定中文词汇间的关系，并构建词汇链，为了进一步提升关键词抽取的精度，选择的候选词是《知网》中收录的名词、动词、形容词(词频大于预定阈值)以及未登录新词，根据其与初始词汇链的相似度，加入相应的词汇链。具体算法如下：

Step1 对文本集进行分词、词性标注和未登录词识别，并统计每个词在文档集中的特征频率 *TF* 和文档频率 *DF*。

Step2 因为有些领域词汇并未被知网收录，而这些词汇相对比较重要，所以 *TF* 大于指定阈值 δ(一般 δ 取值为 3)的未登录词将单独作为一个词汇链 L_0。

Step3 选择文档集中的所有名词、*TF* 大于指定阈值 δ 的动词 $W_1, W_2, \cdots, W_n$ 作为候选词汇集，并取 W_1 构建初始词汇链 L_1。

Step4 依次从候选词汇集中选择词 $W_i(i\in[2,n])$，按照式(1)计算它与除词汇链 L_0 之外的每个词汇链的词义相似度 $S(W_i,L_j)$，即与该词汇链中所有单词的词义相似度和的平均值。

$$S(W_i,L_j)=\frac{1}{N}\sum_{k=1}^{N}Sim(W_i,W_k)=\frac{1}{N}\sum_{k=1}^{N}\max_{i=1,2,\cdots,n,\ j=1,2,\cdots,m}Sim(S_{1i},S_{2j}) \tag{1}$$

其中，N 为词汇链 L_j 中包含词汇的个数；W_k 为 L_j 中的词汇，$1\leqslant k\leqslant N$。

Step5 如果最大词义相似度 $S(W_i,L_k)$大于预设的相似度阈值 ζ，就把 W_i 插入词汇链 L_k 中。

Step6 如果最大词义相似度 $S(W_i,L_k)$小于预设的相似度阈值 ζ，就重新创建一个新的词汇链，并把 W_i 插入新的词汇链中。

Step7 重复 Step3~Step6，直至全部候选词汇计算完毕。

在上述算法中，不难看出相似度阈值 ζ 的选择与最后构建的词汇链的数目呈正比关系，即 ζ 越大，词汇链数目越多。

2.3 词汇链权重的计算

在计算词汇链重要性进行的时候，考虑了以下因素[7]：(1)词汇链的长度(链条包含的词的数目)；(2)构成词汇链的各个词初始权重；(3)词汇链覆盖文本的范围，词汇链覆盖的文本范围越大，则包含主题的内容就越多；(4)词汇链中词汇的分布密度，词汇分布越集中，整体的重要性就越高；(5)词汇链的拓扑结构，考虑词之间关联程度，加强核心节点的重要性。

至此完成了对文本的词汇链构建，并对词汇链进行了评价，赋给相应权值。每个文本表示成 $T=\{T_1,T_2,\cdots,T_n\}$，其中 T_i 表示各个词汇链的权值。词汇链的权值越大，表达文本主题越强；反之，权值越小，离主题就越远。本文预设了一个阈值，从中取出最强的几个词汇链(肯定包含未登录词所在的词汇链)来共同表示文本，当然关键词将从这几个词汇链包含的词汇中抽取。

2.4 关键词抽取算法

根据上述算法构建的词汇链 $L_i(0\leqslant i\leqslant n)$实际是若干个语义相近的词汇的集合，选择其中的哪些词汇作为关键词，需要考虑以下 4 个属性：

(1)首次出现位置：表示在词汇所在的文档中该单词首次出现位置之前的词汇数量占文档中所有词汇数量的比率，这一属性的取值在 0~1。

(2)所处文档区域：采用这个属性是基于如下的假设：出现在文档标题、文档摘要和章节标题中的词汇是文档关键词的可能性比其他词汇要大。

(3)所处词汇链的强度：即词汇所处词汇链的权值，权值越大，表明该词汇链表达文档主题的能力越强。

(4)词汇的信息熵：代表了词汇涵盖了多少文档内容，具体计算方法如式(2)所示；如果词汇几乎在所有的文档中出现，则信息熵将会很小，如果词汇只出现在个别文档中，则信息熵将会很大，这种思路与 *TFIDF* 是一致的。

$$E_i=\frac{1}{1\mathrm{b}(M)}\sum_{j=1}^{M}\left[\frac{f_{ij}}{df_i}1\mathrm{b}\left(\frac{f_{ij}}{df_i}\right)\right] \tag{2}$$

其中，E_i 表示词汇 W_i 的信息熵；M 表示单文档中的句子总数或多文档集中的文档总数；f_{ij} 表示 W_i 在句子 j 或者文档 j 中出现的次数；df_i 表示出现 W_i 的句子数或文档数。

在综合考虑了词汇首次在文档中的出现位置，所处文档区域、所处词汇链的强度和信息熵等 4 个属性，本文提出了如式(3)所示的词汇权重计算方法：

$$Weight_i=\alpha\times1\mathrm{b}(f_i+1.0)\times(1+E_i)+\beta\times T_i+\gamma\times\frac{Length_i}{Length}+\eta\times Area_i \tag{3}$$

其中，$W\mathrm{eight}_i$ 表示词汇 W_i 的权值；f_i 表示 W_i 出现的次数；T_i 表示 W_i 所在词汇链的权重；$Length_i$ 表示 W_i 所在文档中该词汇首次出现之前的词汇数；*Length* 表示文档中所有词汇数量；$Area_i$ 表示 W_i 所处文档区域的权重，如果 W_i 出现在文档标题中时，$Area_i$=5，出现在文档摘要中时，$Area_i$=4，出现在章节标题时，$Area_i$=2，否则，$Area_i$=0.5；α, β, γ 和 η 是用来平衡词汇权重计算中 4 个属性的调节因子，一般情况下均取值 1 即可。

至此，对词汇链 T_i 中包含的所有词汇进行权重计算后，可以表示为 $T_i=\{t_{i1}, t_{i2},\cdots,t_{im}\}$，而 t_{ij} 表示构成词汇链 L_i 的词汇 W_j 的权值信息。对所有词汇链中的所有词汇进行权值计算后，按照权值进行降序排列，并依次选择权值最大的词汇作为关键词，直至关键词的数目 N 达到预定的个数。

3 基于关键词抽取的自动文摘实现

关键词是自动文摘的一个特例，由于用户通过关键词可以基本了解文章的主题，说明了关键词蕴含了文档的重要信息。如果在事先知道文档关键词的情况下进行自动文摘，这势必能够进一步提高摘要的质量，所以本文在基于词汇链的关键词抽取的基础上，对单文档进行自动文摘的过程可以分为如下 3 个步骤：

Step1 每篇文档都是由若干个句子组成的，将抽取的关键词作为特征项作为文档表示的基本单位，用向量空间模型对每个句子进行表示。

Step2 对于每一个句子，根据其包含的关键词的数目和权重，及句子所处位置等信息对该句子的重要度进行计算。

Step3 选取预定比例的、权值递减排序的前面的若干句子，并按照它们在文章中原来的顺序输出，即可形成文摘。

3.1 句子表示

向量空间模型是 Salton 于 20 世纪 60 年代提出的，并成功地应用文本分类和信息检索等领域[4]，向量空间模型以其简单、有效的文本信息表示模型被广大研究人员采用，本文以 2.4 节抽取的关键字集合作为文档表示的基本单位，句子 S_j 的向量表示如式(4)所示：

$$V(S_j)=(k_1,W(k_1);\cdots;k_i,W(k_i);\cdots;k_n,W(k_n)) \tag{4}$$

其中，t_i 表示第 i 个关键词；$W(t_i)$表示关键词 t_i 的权重。

3.2 句子重要度计算

为了衡量句子的重要性，需要给文档中的每个句子 S_k 赋予权重 $W(S_k)$，$W(S_k)$主要由以下 5 个因素决定[8]：

(1)句子中包含的关键词的重要性。句子关键词权重之和越大则说明句子的重要度越大，为了弱化句子长度对权重的影响，采用关键词权重之和除以句子中关键词个数的方法。

(2)句子在文档中的出现位置。处于篇首、篇尾、段首和段尾等位置的句子通常比其他位置的重要度要高。

(3)句子中是否包含提示语。如果包含“综上所述”、“总而言之”等词语，那么句子往往是对主题内容的概括，因此该句子重要性相对较高。

(4)句子是否为标题句。标题通常是对下文的一个概括，无论在信息量还是重要性都比较高。

(5)句子是否以“例如”、“比如”等细节性词语开头，这些词语的出现意味着句子包含举例成分，并非概要性语句，因此重要性相对较低。

综合上述 5 个因素，句子的重要度计算 $W(S_k)$ ($1\leqslant k\leqslant m$) 定义如下：

$$W(S_k)=\lambda_1\times\sum_{i=1}^{n}W(t_i)/Len+\lambda_2\times W_{pos}+\lambda_3\times W_{hint}+\lambda_4\times W_{title}+\lambda_5\times W_{ex} \tag{5}$$

其中，$\sum_{i=1}^{n}w(t_i)$ 是句子 S_k 中关键词的权值和；Len 是 S_k 中包含关键词总数；W_{pos} 表示句子 S_k 的位置权值；W_{hint} 表示提示语权值；W_{title} 表示标题句权值；W_{ex} 表示细节性词语权值；λ 是加权系数，$\lambda_1\geqslant 0.5$，$\lambda_1\geqslant\lambda_2\geqslant\lambda_3\geqslant\lambda_4\geqslant 0$，$\lambda_5\leqslant 0$，$\sum_{i=1}^{5}\lambda_i=1$。

3.3 冗余消除与文摘生成

文档中的每个句子的重要度计算出来后，依据其重要度将各句降序排列。摘要构造方法是依次将重要度最大的句子抽取出来，直到摘要达到指定长度，摘要长度一般由用户确定，通常是原文的 5%~25%，接着将这些从原文抽取的文摘句按其在原文中的顺序排列输出即可生成文摘。但是由于反映文档中心内容的句子可能多次在文档不同位置处出现而被同时选入文摘，从而导致文摘内容过于重复，为了解决这一问题，在选择每个句子的时候，需先与已选句子进行相似度计算。设定一个阈值，相似度高于该阈值的 2 个句子认为是内容重复的，只保留其中权值较高的一句，舍弃另一句。

4 实验

4.1 评价标准

本文采用基于句子命中率的自动评价方法，该方法通过考察机器生成的自动文摘与专家文摘在句子层级上的重合率对文摘进行评价，从准确率(P)、召回率(R)和调和值 $F\_measure$ 3 个方面体现文摘的优劣，其中调和值 F 是准确率和召回率综合评价指标。

$$P=\frac{|S_m\cap S_t|}{|S_m|} \tag{6}$$

$$R=\frac{|S_m\cap S_c|}{|S_c|} \tag{7}$$

$$F\_measure=\frac{(\beta^2+1)P\times R}{\beta^2\times P+R} \tag{8}$$

其中，S_m 是计算机自动摘录的文摘句；S_t 是由多位专家文摘人员手工摘录的文摘句集合的并集；S_c 是是它们的交集，在本文的对比实验中令β=1。

4.2 实验方法

为了对基于关键词抽取的自动文摘方法进行评价，从 1998 年《人民日报》中选择了 100 篇文章进行测试，其中包括教育、财经、体育、军事 4 种类型的文章。将文摘压缩率分为 5%、10%、15%、20%、25%、30%共 6 种情况。首先请 4 位专业文摘人员独立地按照压缩比，手工从每篇文档中摘录出相应数目的句子，作为“理想文摘”，然后使用基于关键词抽取的方法和 Edmundson 的方法生成各种压缩率的文摘，并用上述的 3 个评价指标对 2 个方法生成的机器文摘与理想文摘的重合率进行评价。

4.3 实验结果与分析

为了评价基于关键词抽取的自动文摘的效果，以及关键词的数目对文摘效果的影响，本文对 4 类测试文档采用不同数目的关键词和 6 种不同的压缩率 m 进行文摘，并与“理想文摘”进行了比较，$F\_measure$ 实验结果如表 1 所示，其中设 x 为关键词数。

表 1 基于关键词抽取的自动文摘 *F_measure* 实验结果

类别	x	m=5%	m=10%	m=15%	m=20%	m=25%	m=30%
教育	3	0.404 5	0.482 5	0.472 9	0.532 5	0.551 4	0.484 7
	4	0.421 9	0.511 7	0.513 5	0.591 9	0.567 6	0.564 8
	5	0.380 0	0.494 0	0.483 1	0.559 8	0.548 3	0.537 5
财经	3	0.415 5	0.484 9	0.515 0	0.544 9	0.529 4	0.535 0
	4	0.447 7	0.564 7	0.563 7	0.611 9	0.600 2	0.623 8
	5	0.409 5	0.489 6	0.516 7	0.568 0	0.555 3	0.533 5
体育	3	0.430 6	0.486 0	0.535 3	0.529 3	0.589 3	0.513 1
	4	0.453 8	0.523 0	0.568 4	0.631 2	0.618 4	0.673 2
	5	0.437 3	0.481 8	0.517 2	0.584 2	0.552 0	0.606 8
军事	3	0.427 4	0.504 3	0.503 9	0.533 8	0.485 7	0.494 4
	4	0.443 6	0.534 2	0.604 8	0.610 9	0.604 4	0.630 1
	5	0.425 8	0.438 2	0.548 9	0.575 7	0.552 5	0.538 4

从表 1 的对比实验结果来看，对于 4 种类型的文档在不同关键词数目和不同文摘压缩比下的文摘的 $F\_measure$ 实验结果的整体评价比较理想，其中当关键词数目设置为 4 的时候，在各种压缩率下均得到了比较好的效果，尤其是体育类文章，分析其原因，主要是因为本文提出的关键词抽取的方法和文摘句摘录的方法在这类文章的关键词及文摘句的位置等因素方面得到了更多的体现，而在教育类文章上没有更多的吻合，所以应该在关键词抽取和文摘句摘录方面需要考虑更多的因素。为了客观地考察本文提出的自动文摘方法的实际效果，在 6 种不同的文摘压缩率下，将采用 3 个不同数目的关键词生成的文摘与 Edmundson 方法生成的文摘进行了比较，具体实验结果如图 1 和图 2 所示。

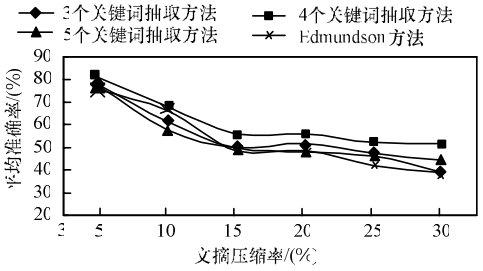

图 1 平均准确率结果对比

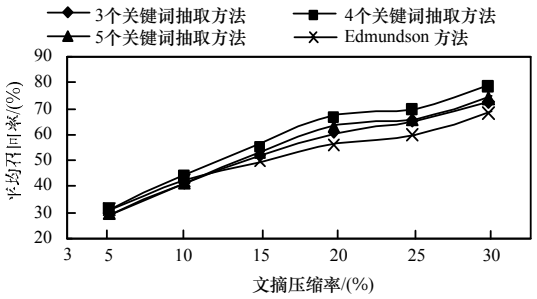

图 2 平均召回率结果对比

从上述结果可以看出，在压缩率较小(5%和 10%)的情况下，2 种方法的文摘平均召回率和平均准确率是比较接近的，但当压缩率增大以后，基于关键词抽取生成的文摘平均召回率和准确率均有明显的提高，这是因为基于词汇链的关键词抽取将词汇按照它们的语义进行了“聚合”，能够从语义分析的角度对文档词汇及其相关特征进行归纳，是一种基于理解的、深层分析的方法，将该方法抽取的关键词用于自动文摘更有利于把握文档的主题。

如果单纯从准确率来看，当压缩率较小时，文摘的准确率要明显高于压缩率较大时的准确率，这表明随着文摘长度的增长，其差异也在扩大。其实，人工摘录的情况也是如此，当文摘句很少时，摘录目标相对比较集中；当文摘句较多时，摘录目标就变得比较分散。

5 结束语

在对文档进行未登录词识别的基础上，本文提出了一个基于词汇链构建的关键词抽取的算法，并将该算法抽取的关键词用于文摘生成。由于词汇链是由一系列具有语义相关性的单词所构成，因此将文档中的词汇先组织成词汇链，再结合词汇所在词汇链的强度、信息熵和出现位置等属性从语义的角度提高关键词抽取的性能。该方法不受领域的限制，并且可以用于单文档或多文档的关键词抽取。实验结果表明，在关键词抽取的基础上，将采用不同数目的关键词生成的文摘与 Edmundson 方法生成的文摘进行了比较，在召回率和准确率方面均有所提高。

参考文献

[1] 傅间莲, 陈群秀. 基于规则和统计的中文自动文摘系统[J]. 中文信息学报, 2006, 20(5): 10-16.

[2] 郭玉箐, 万 敏, 罗振声. 面向非受限领域的综合式自动中文文摘方法[J]. 清华大学学报, 2002, 42(1): 139-142.

[3] Goldstein J, Kantrowitz M, Mittal V O, et al. Summaring Text Documents: Sentence Selection and Evaluation Metrics[C]//Proc. of Research and Development in Information Retrieval Conference. Berkeley, USA: ACM Press, 1999: 121-128.

[4] 张 虹. 基于自动文本分类的关键词抽取算法[J]. 计算机工程, 2009, 35(12): 145-147.

[5] 王灿辉, 张 敏, 马少平, 等. 基于相邻词的中文关键词自动抽取[J]. 广西师范大学学报: 自然科学版, 2007, 25(2): 161-164.

[6] Morris J, Hirst G. Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text[J]. Computational Linguistics, 1991, 17(1): 21-48.

[7] 尤文建, 李绍滋, 李堂秋. 基于词汇链的文本过滤模型[J]. 计算机应用研究, 2003, 20(9): 32-35.

[8] 王继成, 武港山. 一种篇章结构指导的中文 Web 文档自动摘要方法[J]. 计算机研究与发展, 2003, 40(3): 398-405.

编辑 索书志

(上接第 182 页)

行 3 次 CuBICA 算法和 FastICA 算法和 Infomax 算法，比较三者的执行效率，如表 4 所示。

表 4 3 种算法的运算时间 s

算法	第 1 次	第 2 次	第 3 次	平均
CuBICA 算法	0.002 214	0.002 254	0.002 246	0.002 238
FastICA 算法	0.007 991	0.007 024	0.007 873	0.007 629
Infomax 算法	0.021 970	0.021 643	0.021 904	0.021 839

由于计算机操作系统的多进程特性，3 次运算时间略有不同，取平均后可以发现，CuBICA 算法的运行时间明显小于 FastICA 算法和 Infomax 算法，具有更好的执行效率。

4 结束语

采用一路脑电信号和眼电、心电信号组成三维的输入信号，先计算当该输入信号的三阶和四阶累积量中非对角元素的平方和 $\overline{\Psi}_{34}(\boldsymbol{u})$ 取得最小值时的旋转矩阵，进而可求得解混矩阵，从而得出脑电信号的估计。为了验证算法的有效性，本文同时引入了信号分量的互相关系数和矩阵分量的互相关系数来验证 CuBICA 算法的有效性。因为 CuBICA 算法基于累积量的特性和简化的数学公式，在处理较低维信号时具有比 FastICA 等算法更好的处理效果和更短的处理时间。由于脑电信号伪迹去除过程中涉及的独立源信号较少，因此 CuBICA 能够有效地去除脑电信号中的伪迹。但是 CuBICA 算法在处理高维信号时所需的时间较长，影响了该算法的应用范围。本文把 CuBICA 算法应用于生物医学领域，成功地实现了脑电信号中眼电和心电伪迹的去除，证明 CuBICA 算法在脑电信号处理方面具有较好的发展前景。

参考文献

[1] Akhtar M T, James C J. Focal Artifact Removal from Ongoing EEG——A Hybrid Approach Based on Spatially-constrained ICA and Wavelet De-nosing[C]//Proc. of IEEE Engineering in Medicine and Biology Society. Shanghai, China: [s. n.], 2009: 4027-4030.

[2] Blaschke T, Wiskott L. CuBICA: Independent Component Analysis by Simultaneous Third-and Fourth-order Cumulant Diagonalization[J]. IEEE Transactions on Signal Processing, 2004, 52(5): 1250-1256.

[3] Comon P. Independent Component Analysis, a New Concept?[J]. Signal Processing, 1994, 36(3): 287-314.

[4] Wang Bin, Lu Wenkai. An In-depth Comparison on FastICA, CuBICA and IC-fast ICA[C]//Proc. of IEEE ANC'05. [S. 1.]: IEEE Press, 2005: 410-414.

[5] 郭 瑞, 宋海娜, 匡纲要. 基于定点 ICA 算法的人脸识别方法[J]. 计算机工程, 2004, 30(9): 159-161.

[6] Hyvarinen A, Oja E. Independent Component Analysis: Algorithms and Applications[J]. Neural Networks, 2000, 13(4/5): 411-430.

[7] 曲国庆, 党亚民, 章传银, 等. 小波包消噪方法分析及改进[J]. 大地测量与地球动力学, 2008, 28(4): 102-106.

[8] 姚志湘, 刘焕彬, 粟 晖. 盲信号分离输出与源信号的一致性判断[J]. 华南理工大学学报, 2007, 35(5): 50-53.

编辑 索书志