维普资讯 http://www.cqvip.com

信息检索自然语言查询问句处理框架*)

熊文新 宋 柔
(北京语言大学信息科学学院语言信息处理研究所 北京 100083)

摘 要 以自然语言形式提出的查询问句不同于通常的关键词或主题词查询,需要提取用户真正要检索的信息内容。该文提出一个自然语言查询语句的处理框架,由3个部分构成:(1)离析查询问句的操作表述和信息内容;(2)凸显真正的信息需求内容;(3)对不同信息内容采取不同的词语实现方法。这一处理可望为自然语言信息检索提供准确的用户需求分析。

关键词 信息检索,查询请求,查询表述

A Framework for the Natural Language Request Processing in Information Retrieval

XIONG Wen-Xin SONG Rou
(Center for Language Information Processing, College of Information Science, Beijing Language and Culture University, Beijing 100083)

Abstract A framework to deal with the requests expressed in natural language is proposed, which aims to understand the users' real information content. The process consists of three steps, (1) to differentiate the expressions between operation and information content, (2) to extract the salient concept in information need, (3) to implement the different word forms for the information content. It is hoped that it will be beneficial for better grasp to users' information need.

Keywords Information retrieval, Information request, Query expression

网络发展促使上网信息越来越多,WWW成为一个巨大的知识宝库。信息检索日益受到关注。由于网上信息内容和用户查询语句多数以自然语言形式存在,信息检索在某种程度上可归结为一个语言工程问题。

用户表达查询意图,可采用不同表述方式:简单地排列几个关键词或短语,如例1;也可用复杂短语或句子形式,框定查询范围,如例2;还可进一步界定目标内容和所需排除的内容,如例3;甚至提交一段或整篇文档内容,寻找类似文档。

例1 青少年犯罪

例2 青少年犯罪相关的教育、预防、原因研究问题

例3 检索媒体或专家对于青少年犯罪的根源的分析,包括社会原因、家庭原因、学校教育原因,可以结合具体青少年犯罪事件给予说明,简单介绍单个犯罪事件不在检索范围之内。

例1中的关键词组合可直接用作检索项。例2需要过滤结构助词"的"和表概括义的实体词"问题"等。例3既有查询操作表述词语"检索",又有限制目标的表述词语"包括,范围"等;还有信息内容词语"青少年,犯罪"。因此应区分对查询操作及条件的表述和信息内容词语,前者属于通用查询操作,后者才是在目标文档中真正体现的检索内容。当前搜索引擎和数字图书馆领域由于系统架构和用户界面等原因,经常出现例1这种主题词或关键词短语的情形[1]。而我们主要处理后两类需要更多语言处理的情况。可以预期,随着用户多通道界面的发展,降低用户使用门槛,以自然语言形式出现的查询将会越来越普遍。

1 问题与对策

基于向量空间(VSM)的检索系统由于方便易行而得到广泛应用。然而这种方法在对用户提问的处理却有几个方面值得重新考虑,主要有:(1)默认用户输入就是要查询的信息内容,因而直接将输入语句分词的结果作为关键词处理,模糊了查询操作表述和信息内容的区别。由于问题表述的查询语句较短,对真正意义的检索项干扰影响更大。(2)对于分词结果表述的信息内容,没有"必须出现"一说,所有词语同样只是向量空间的一维,降低了某些需要凸显的检索内容词语的重要性。(3)对于检索内容在目标文本的词语实现形式缺乏多样化的处理,虽然采用WordNet等词语知识库,但多数局限于同义、反义和上下位的词语扩展,并且受制于所用通用知识库的编排原则。

由此,我们相应提出一个正确理解用户查询需求的处理框架,希望能够解决信息检索前端查询问句的分析问题。该方案如图1所示。

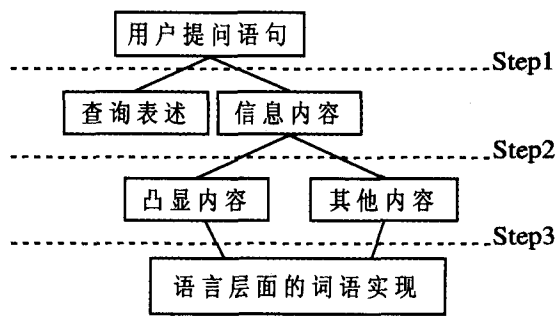

图1 查询问句处理框架

由图1可以看出,用户的自然语言查询表述首先被区分成信息内容和需求表述两部分。针对信息内容,凸显主要概念内容和一般信息内容,并对概念内容采用不同的词语实现形式。将处于凸显位置的信息内容权重提升,直至采用严格的布尔查询限制。

*)国家自然科学基金资助项目(60272055)、国家863计划资助项目(2001AA114111)、教育部科学技术研究重点资助项目(00128)。熊文新博士生,主要研究方向:中文信息处理、语言工程;宋 柔 教授,博士生导师,主要研究领域:智能软件工具、语言信息处理、人工智能。

维普资讯 http://www.cqvip.com

2 需求表述和信息内容

自然语言查询与关键词查询的最大不同在于，前者可能既包括后者，同时又包含有对信息需求的表述，因此不能简单地将自然语言查询语句囫囵地送入检索系统。区分需求表述和信息内容，提取信息内容词语作为检索项是形成系统输入的重要一步，可以从以下几个方面考虑。

2.1 停用词的确定

在单篇文档中的词语频次(TF)和同一文档集中不同文本出现的文档频次倒数(IDF)是衡量词语重要程度的指标。某个词语若在单篇文档反复出现，则重要性较高；若只在某些文档出现，则区分能力较强。综合 TF 和 IDF 考虑，一个词语如果既有较强区别性又有较高重要性，则可作为较好的检索项[2]。

考察用于信息检索的查询语句词语的不同分布，我们选用 863 信息检索测试问题集(2004)中的 30 个提问作为小型查询语料库。规模虽小，但尽量模拟真实用户的信息需求编写，并不试图对句式结构刻意描写，基本反映了普通用户对主题信息的表述。

首先对该语料库按每个问题一个文档的方式，拆分成 30 个子文档，采用北京语言大学语言信息处理研究所开发的分词系统(具有新词发现等功能)[3]实施分词处理；进而构造词语倒排表，再统计该语料库（总库，未拆分版本）各词语在这 30 个子文档(拆分后文档)中的 DF 值和在整个语料中的 TF 值，并对最终结果逆序排列。

表 1 不同语料的 TF，DF 对比

排序	查询 DF	查询 TF	新闻体 TF	大众体 TF
1	的	的	的	的
2	检索	检索	在	了
3	不在	范围	和	一
4	范围	不在	了	在
5	内	相关	是	是
6	相关	介绍	一	我
7	介绍	内	个	他
8	内容	以及	为	你
9	以及	对	有	和
10	在	在	中	有

为确定查询语料和通用语料的用词差别，以 1993 年全年《人民日报》的新闻体语料和容量 14G 的网页文本的大众体语料作为参照，取出词语分布特征。前 10 个 TF 值高的词语与查询语料 DF 值和 TF 值排序前 10 的词语参见表 1。

考察该表不难发现，查询语料库与通用语料库中的高频词语并不一致。新闻体的人民日报语料和大众体的网页文本相差不大，多是一些虚词和常用动词或代词；而查询语料库中更多出现表述查询操作和范围限制的词语。可见信息检索的查询表述用词和传统意义的停用词有区别。虽然都不是信息检索内容用词，但前者是对检索行为和对象要求的限定，与检索相关，称为专用停用词(specific stopword)；后者则是汉语的普遍现象，称为通用停用词(general stopword)。我们计划结合词语的左右熵及其词语在查询问句中的位置信息，构建这两类停用词候选词表。

应该注意到并非所有停用词候选词表的词语都在一个具体的查询问句中实现为停用词。比如“检索”对于查询语句而言具有表述指令意义，可用作专用停用词，但如果查询语句中用于“信息检索系统”则其不成为停用词。因此，实时动态识别当前查询语句中的停用词，是其中一个重要课题。

2.2 句法模板

自然语言查询与离散的关键词区别还表现在对问题的描述上。我们曾经对北京语言大学语言信息研究所开发的一个自动问答系统采集的近 20 万语句的用户查询日志进行过分析，发现在用户单句形式的查询中，至少有 16.5% 含有除疑问词语之外需要滤除的需求表述用词[4]。

首先利用《同义词词林》和知网中具有疑问义的词语构造一个疑问词语集，同时参照疑问句式如“X 不 X”“是否”和疑问语气词“吗”等，作为疑问句的形式特征，如“怎样联系北京市轮胎实业公司”、“治疗风湿病能不能用激素”、“订阅索易电子刊物是否需要付费”、“脂肪瘤能用中药治疗吗”；采用“请 X”等作为祈使句格式标准，如“请介绍 MTV 的特点”；采用“我想 X”等作为陈述句格式特征，如“我想知道中国版画艺术的起源”等，对查询语句分类，得到表 2 中的数据。

表 2 不同句式在查询表述中的百分比

句式	句子数	词语数	词语/句	百分比
疑问句	160589	1201060	7.48	82.06%
陈述句	30877	267257	8.66	15.78%
祈使句	3613	28467	7.88	1.85%
短语	595	4497	7.56	0.31%
总计	195676	1501377	7.67	100%

信息检索类的查询语句与日常所用的汉语句子差异较大，以获取信息为目的，对信息需求表述的句法语义相对简单，有一定规律可循，甚至可用表层模式匹配(pattern matching)的方法[5]实现信息内容和操作表述的分离。

2.3 功能词语分类

在借鉴数据库标准查询语言 SQL 语言后[6]，发现以下几类属于通用查询操作的语词不表示信息内容，不参与文档的内容匹配。

(1)检索操作词语。多是具有指令性的动词，通常用于谓语位置，如“检索、查找、列出、查询”，不是信息内容，不会出现在 Web 用户或联机检索的主题词组合检索，可直接过滤。

(2)目标对象词语。由指称文档类的词语充当，常在句中作宾语，如“文本、文章、新闻、描述、信息、评论、资料”。文本信息检索返回结果缺省是整篇文档，该部分词语可能在用户请求时从缺。但有些具有特别提示词语如“描述/评论”等，对目标文档的表述类型有影响，需依据实际文本内容进行取舍。

(3)查询对象词语。具有抽象地域属性，可由介词“从、在”引导，如“WWW 测试集、语料库”。自由文本的检索对象可以是 Web，或是某一特定语料集合。因其具有指定检索对象的特殊作用，不宜进行简单过滤。

(4)表示查询条件的词语或固定结构。如“与此相关、关于、属、范围、方面、包括、论及”等标识特定信息的词语；“必须、只、仅、单独、均、凡”等表示表述查询内容模态和量词逻辑等的语词；“不相关、与此无关、除外、没有、之外、不”等排除不相关内容的语词；这些词语或格式指定限制检索条件，对离析信息内容具有重要意义，可借用模板触发词解决。

3 信息内容的凸显

通过分离操作表述，可以将信息内容归入到不同检索条件。这其中存在需要检索和需要排除的区别，分别称为正条件(positive)和负条件(negative)部分[7]。对归属不同条件的概念内容分析，可以凸显真正的信息需求，并对其实施权值加强，直至对凸显的概念采用关键词的“必现”策略，以防止向量

维普资讯 http://www.cqvip.com

空间模型中由于其他检索项竞争造成的弱化。

3.1 查询概念凸显

一个查询概念被凸显，指的是如果概念 A 处在正条件（即检索范围内）中，同时又在负条件部分反方向（限制非 A 内容不属于被检索范围）制约后，相应被突出强化的概念内容，在目标文档中必须出现其语言体现形式。设 Salience(A) 表示概念 A 处于被凸显的位置。以逻辑为基础，如何构造、表示描述概念凸显的运算规则，来实现不同条件下查询概念的提取，是这一步工作的重点。

$A, \sim(\neg A) \rightarrow \text{Salience}(A)$

如针对查询需求“查询与电影《英雄》票房收入有关的内容，对电影的评价等与票房无关的内容不在检索范围内”，正条件要求检索票房收入(A)，负条件从反方面指出非票房(～A)内容不是检索目标。由此强化“票房收入”为凸现的检索概念内容。

有时检索条件当中出现在同一概念体系下具有互斥性的两个对象，若正条件肯定其中某一个对象，同时负条件否定其对立面，则相应地被肯定的对象处于概念凸显位置。即存在这样一种推导关系：

$$(\alpha(A), \sim\alpha(B)) \wedge (A \cap B = \text{NULL}, A, B \in U) \rightarrow \text{Salience}(A)$$

如检索需求为“介绍减肥药物和采用节食减肥等方法介绍不在检索范围内，仅检索运动减肥相关内容”。正条件要求运动减肥。而减肥手段，可以分成若干种，与之并列的是药物减肥、节食减肥等。负条件强调的正是非运动减肥不是检索内容，由此运动减肥得以凸显。

有时信息需求是检索几个个体对象之间的关系，单独某一个个体具有的性质不是其检索目标。这种情况常发生在要求若干行为主体共同实施某项行为动作，或是主体之间存在某种关联的检索。强调其共有性，则有

$\text{Mutual}(\alpha\beta), (\sim\text{Single}(\alpha) \wedge \sim \text{Single}(\beta)) \rightarrow \text{Salience}(\text{Mutual})$

3.2 概念出现的类型

由于向量空间没有“必须”一说，所有检索用词语都在权重数值化后，当作空间中的一个向量参与相似度计算，检索项的重要性不如布尔模型。放宽匹配固然能查全更多内容，同时导致精度下降。通过对用户检索条件中核心概念的分析，可以找出凸显的主题检索内容，采用关键主题概念和限制概念必现的策略，过滤无关检索结果。在这一过程中，那些用来表示限制范围的副词“仅”、“只”等具有强烈的提示(cue)意义，应该给予特别关注。

关于凸显的概念在目标文档中是否出现，我们区分如下类型：

(1)必须出现。检索概念必须在目标文档中体现，相应的概念意义属于绝对必要；

(2)可选出现。检索概念可在也可不在目标文档中体现，概念意义属于相对必要；

(3)必须不现。概念必须不在目标文档中体现；

(4)可选不现。概念不必在目标文档中体现。

应该指出，上述 4 条指的都是概念意义，而非具体词形，下一节概念意义的语言形式体现反映的是如何将概念意义实现为具体词语。因此，(1)处在主题层面的概念或不可再分解的原子概念属于必现范围。典型例子就是经过分析凸现的主题概念和人名地名等具有专指的概念，如强调“北京、房价、原因”中的“原因”和“北京”；(2)是向量空间模型的典型处理模式，所有表示概念都数值化后作为向量参与相似度的计算，但没有强制出现的规定；(3)是由负条件排除的禁用概念，任何出现该禁用概念的文档即判定为非相关，主要用于信息过滤，比如过滤某些含有敏感内容的文档；(4)表示该概念意义与主题无关，但禁忌程度不如(3)来得那么严重。这些概念在文档中出现并不意味不合要求，还需要考察目标文档是否同时有符合正条件的检索内容出现。比如“检索北京房价原因，关于上海房价内容不属检索范围”，“上海”虽然在文档中出现，但如果不是处于文档表述的主题部分，而只是作为“北京”的对比出现，则也满足检索条件。

考察 4 类情形，我们可以看到：(1)、(3)与布尔查询的关键词检索类似，形式要求严格；(2)、(4)则更像是向量空间模型的一部分。在对检索语句进行分析时，应根据检索条件凸显的不同概念内容，对不同类别词语采用不同检索策略。

3.3 不同条件部分的连接关系

研究检索条件之间不同概念的内部和外部逻辑关系对检索系统确定检索项有益。一般而言，处于正条件范围的并列性概念多数属于“或”关系，体现为在目标文档中表示该概念的词语为可选出现。例如“查询与奥运场馆建设进度、建设规划或者赛后场馆的利用安置等相关文档”则可归结为“奥运场馆∧(建设进度∨建设规划∨赛后利用安置)”，其中由“、”等形式关联表示的概念，只要任何一个与“奥运场馆”相关出现就满足需求。

而处在排除范围的负条件中的相关并列概念，则实现起来为所有这些概念都不能在目标文档出现。比如“生物公司介绍、涉及克隆的科幻电影、电脑术语、克隆、克隆人的讨论等内容不在检索范围之内”，则可描述为“～(生物公司∨电影∨电脑术语∨克隆人讨论)”，进而可表述为“～生物公司∧～电影∧～电脑术语∧～克隆人讨论”。

4 检索概念的语言形式实现

得到凸显概念内容后，在现有信息检索系统框架内实现，需要确定该信息内容的语言形式作为查询词语，以此实现查询与文档的匹配。根据调查，不同作者采用同一词语表述相同概念的情况不超过 20%[8]。因此检索系统在处理用户真实问题时，面临查询词语失配(mismatch)[9]。这是由同义词和多义词造成的：一个概念有多种不同表示，一个词语也有不同义项表示不同概念。

信息检索界曾利用 Wordnet 等词语知识库进行问句查询词语扩展，收效并不太好[10]。词典中的同义词一般是同义互释，同义词集中的同义词是由表达相同意义的一组词构成。而用户表述可能存在不同层级的抽象概括(generalization)，难以简单利用词语同义扩展。

对信息检索系统而言，具体语境中对每一表示信息需求的内容词语都应找到其合适的扩展办法。我们认为概念检索义在词语扩展上应区分 3 种情况。

4.1 不能扩展

针对某些信息内容，往往需要目标文本必须出现某些查询语句中的词语。这体现出信息检索的刚性部分。这类词语多数是表述具有唯一所指、目标明确或具有典型区别意义的对象。如查询“姚明在 NBA 中的表现”、“北京三星级酒店的介绍”，则文档必须体现“篮球赛是 NBA，并且运动员是姚明”；“酒店级别标准是三星，并且地理位置处于北京”等内容，其他篮球比赛或运动员，地域非北京及级别非三星的酒店介绍都不属于信息需求范围。

不能扩展的概念在形式体现上多是命名实体(Name Entity)，如人名、地名、机构名或者数量短语等。经过分词标注

维普资讯 http://www.cqvip.com

后发现的许多未登录词语也多属于此类。

应该指出，虽然有特定的明确所指，指称对象和意义不能替换，但有时也可以在词形上实施一定程度的扩展，只是扩展后的词语与原词语应该在所指内容意义上等价。如“美国”、“白宫”、“华盛顿”、“布什”等在表述国家政权时可以看作是“美国政府”的等价词(equivalent word)，若检索美国政府针对某一事件的立场看法时，这些词语应该视作对同一所指的不同表述；然而在表述自然人的纳税业务场景时，显然“布什”与表示地理名称或指称国家机关的“华盛顿”不具有不损失等价意义的可替换性。这部分信息内容需要构建知识库来利用。

4.2 需要扩展

当前信息检索底层核心依然是基于字符串匹配的全文检索技术[11]。查询词语扩展是防止查询语句和文档词形不匹配，造成查全率下降的常用手段之一。

查询词语扩展，一般有两种。一是采用只知识库的方法：利用词语资源库中记录的关系信息，常见的有同义(synonymy)、上下位的层级(hierarchy)关系。二是采用语料库统计方法：根据语料库中与检索项词语同现的词语，如互信息(MI)，获取与检索项词语意义相关联的其他词语；或是在较大语料库中与检索项词语具有相似分布的上下文的词语作为扩展项，如 KL 距离和信息半径(information radius)计算[12]。前者如“医生”与“护士”、“医院”等经常聚集出现，把具有关联(associative)意义的词语纳入检索范围；后者如“大夫”、“当班医生”等经常分享在医疗场所诊治病人这一典型场景的表述中，因此也可以将其纳入扩展词语。必须指出：单纯考虑词语在语料库中组合和聚合关系的出现强度，并不一定能准确地判定扩展词语和检索项之间确切的语义关系。这就是有时词语扩展对检索有较大促进作用，而相反有时影响检索效果的原因。

此外，不同颗粒度的分词策略也制约文档检出。检索“北京 大学生 就业”，单纯用词形匹配将漏检信息，尤其对已分词但没标注词内结构的文本。如“北京”和“北京大学”就词条而言，并不一致。如果检索要求不限制对具体某所高校学生就业的限制，而词条“北京大学”没有词内结构{北京 大学}的表示，则仅含有“北京大学”学生就业的分词文本，由于词形不匹配将被漏检，从而导致查全率降低。因此，词条内部结构的标注将有助于提升系统指标。

4.3 必须扩展

某些情况下，由于查询语词表述抽象概括义，而非具体的原子概念，不太可能直接在文档中找到其词形，因此必须扩展[13]。

概括词语，即其所指称的主题意义蕴含在文档语句的表述内容中，但该词语本身并不出现在文档内。这些概括词语与文本中实际体现该词语意义的其他词语显然分属不同层次。典型的例证就是对段落大意和文章主旨的归纳和概括。非基于关键词和关键语句抽取的自动文摘、图书情报学领域的主题标引等与此类似。叙词等是对文档主题内容的概括，而叙词本身并不一定出现在被标引检索的文档中。对于此类概念的检索，需要对文档表述的意义进行加工处理，实现高层匹配处理，而不能另取一个同样表述抽象概括意义的同义词来扩展。

设想用户要查询“北京的气候”，其信息需求是关于北京市的气候状况。根据词典定义，气候是“给定地区的天气状况和天气发展所示的变动着的大气状态”。对状态和状况可从不同方面角度来说明。具体落实到该例，按照百科全书的知识体系，需要从“降水量”、“无霜期”等能够体现气象因素的框架构成成分着手。即用户输入的条件是检索“北京的气候”问题，而实际目标文档可以不直接出现检索主题词“气候”的词形(token)本身，只要出现能反映该查询词语主题意义的其他词语(如降水量、无霜期)。其他类似查询还包括“举措”、“简历”等对某一类事物/事件的概括词语，需要扩展为体现其概括意义的相关词语，而非“措施”、“履历”等同义概括词语。人工智能的脚本(script)、框架(frame)和新近兴起的 ontology 的知识表示可以借用来构造扩展知识库[14]。

4.4 扩展方式

被检索概念扩展不仅体现在词语资源库中与查询词语有关联的其他词条上。当前资源多围绕人的常识来构造。而信息检索是任务性很强的应用，与用户特定兴趣和需求相关，通用知识对个性化检索意图无能为力。Ontology 技术瞄准特定领域的概念及关系的知识表述，可望在受限范围内取得一定成果。

比如关于抽象事物“举措”应用在不同场合有不同的体现方式。在“北京筹办 2008 年奥运会的举措”，由于涉及到地方政府这一权力机构，可以有制定法律条文法规等行为，因此围绕这类行为事件的词语能够体现这一概念意义；但对“学校、企业、学生三方对大学生就业的举措”这一检索需求，由于涉及三方都不具备立法资格，表现颁布条例等的语词就不能作为本例“举措”的扩展形式。

抽象概念意义的扩展具体外化为实际词语，可能包括不同类型的词语或格式表现。比如针对“北京房价变化原因”，价格的“变化”可以由“上涨”、“下调”、“稳定”等表示状态转变或维持的词形体现，而“原因”是“造成某种结果或者引发某种事情的条件”，体现事件间的关联。这种虚化的关系，若不借助世界知识或特定专业领域的知识，很难通过实词形式来体现。相反，通过一些特定结构信息，可以显性表述出来。像汉语本体研究中有关复句关系的辨识，就可以列举典型的关联词语来处理。本例中表示原因的因果复句，常见的格式有“因为…所以…”“之所以…是因为…”等。因此如果不深入引入句法分析，则可诉诸语言表述的这些形式特征，来检索此类虚化的概念意义。

结束语 当前自然语言处理技术，特别是句法、语义分析精度尚难以支撑实际运行系统，响应时间不能满足实时应用。相对来说，词语层面的处理更为实际，以词袋(bag of words)作为处理对象是目前信息检索系统的一个主要特征。从语言工程角度出发，我们适度引入基于词语层面的查询语句分析，即突破当前单纯把用户查询分词结果作为检索项的纯粹字面检索或词形出现次数的相似度计算，并不只局限仅对检索内容项作上下位或同义关系的词语扩展(传统的词语扩展法)，而是对用户查询问句进行分析，提取意义层面的检索凸现条件，并完成从意义到实际词形层面的转换，以期真正理解用户需求，并找出符合用户需求的文档。

参 考 文 献

1 Baeza-Yates R, Ribeiro-Neto B, et al. Modern Information Retrieval. ACM press, 1999

2 Vires A, Roelleke T. Relevance Information: A Loss of Entropy but a Gain for IDF? SIGIR'05, Salvador, Brazil, 2005

3 罗智勇，宋柔. 一种基于可信度的人名识别方法. 中文信息学报，2005，19(3)：67～72，86

4 熊文新，宋柔. 信息检索查询语句的表述分析. 见：中国应用语言学会(筹)，第四届全国语言文字应用学术研讨会论文，成都：四川大学，2005

(下转第 204 页)

维普资讯 http://www.cqvip.com

机制的优势，具有很强的灵活性、鲁棒性和局部更新能力。此外，多 agent 集成系统是闭环控制系统，在工程应用中能有效地进行实时在线控制，不断修正系统隐患，使系统始终有效地运行，使系统的运行状态能达到相应的稳定性和鲁棒性要求。因此，多 agent 集成系统是一种适用于不确定性环境的自组织、自学习系统。

3 实证分析

本文将所设计的多 agent 集成系统应用到电力设备励磁装置系统的故障预报中，不断地对子 agent 预测模块进行赋值训练，训练的周期设为 1000 秒，该系统的整体运行情况如曲线图 2 所示。

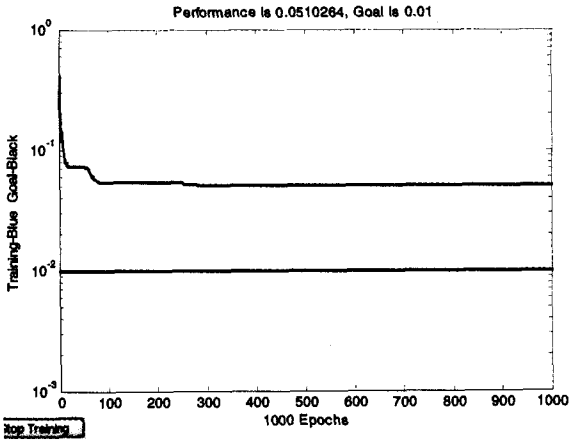

图 2 励磁系统整体运行性能示意图

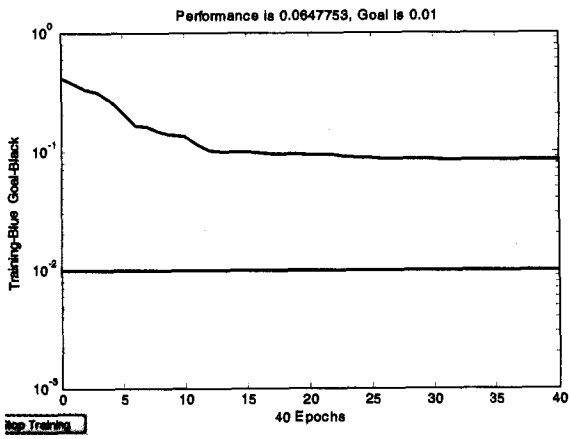

图 3 励磁系统部分运行性能示意图

从系统运行情况看，所设计的安全报警系统分别在第 20、60 和 249 秒左右发生了异常现象，每次经 agent 修正后均可“正常”运行一段时间，这说明系统要达到目标必须不断地修正。

从图 2 看，系统安全报警的异常现象似乎变化不大，实际上在报警期间的运行情况非常不稳定。0 到 40 秒的系统运行详细情况如图 3 所示，从中可看出，系统的内部变化非常剧烈，这与实际情况相吻合，同时由于报警开始到系统“正常”运行之间存在缓冲带，满足了 agent 智能修正所需的时间要求。

上述实证分析表明，多 agent 集成系统故障预报技术在复杂系统的故障预报中的应用是有效的。

结论 本文提出了一种基于多 agent 集成方法的复杂系统故障预报技术，并将其应用到实际工程中。实践表明，这种故障预报技术方法是有效的，可实现智能故障预报。各预测 agent 之间是一种宏观层次上的高内聚松耦合结构，解决了单一 agent 不能同时考虑各方面因素的缺点，使复杂系统内部结构的信息交换成为可能，具有良好的开放性和自适应性。

人作为人工智能系统的使用者，必然对系统产生影响，如何来协调复杂系统故障预报中信息资源的合理分配，是需要进一步深入研究的课题，这也将是笔者下一步研究的重点。

参 考 文 献

1 Jennings N R, Sycar K, Wooldridge M. A roadmap of agent research and development [J]. Auton Agents & Multi-agent syst, 1998,1:7～38

2 马笑潇，黄席樾，柴毅，等. 免疫 agent 概念与模型[J]. 控制与决策，2002,17(4)：509～512

3 黄席樾，刘卫红，马笑潇，等. 基于 agent 的人机协同机制与人的作用[J]. 重庆大学学报，2003,23(9)：32～35

4 曹立军，杜秀菊，秦俊奇，等. 复杂装备的故障预测技术[J]. 飞航导弹，2004(4)：24～27

5 杨天社，杨开忠，李怀祖. 基于知识的卫星故障诊断与预测方法[J]. 中国工程科学，2003,15(6)：63～67

6 张阳，曹迎春，黄皓，等. 移动 Agent 系统中的安全问题和技术研究综述[J]. 计算机科学，2005,32(3):21～25

7 胡昌华，许华龙. 控制系统故障诊断与容错控制的分析和设计[M]. 国防工业出版社，2000

（上接第 147 页）

5 Soubbotin M, Soubbotin S. Use of Patterns for detection of Likely Answer Strings: A Systematic Approach Answer. In: Proceedings of TREC-2002, 2002. 175～182

6 许龙飞，杨晓昀，唐世渭. 基于受限汉语的数据库自然语言接口技术研究. 软件学报，2002,13(4)：537～544

7 Dkaki T, Mothe J. Combining Positive and Negative Query Feedback in Passage Retrieval. In: Proceedings of RIAO' 2004, 2004. 661～672

8 Deerwester S, Dumais S T, Furnas G W, et al. Indexing by latent semantic analysis. JASIS, 1990,41(6)：391～407

9 Crestani F. Exploiting the Similarity of Non- matching Terms at Retrieval Time. Information Retrieval, 2000, 2(1)：27～47

10 Voorhees E. Query Expansion Using Lexical- Semantic Relations. In: Proceedings of 17th ACM-SIGIR, Dublin, Ireland, 1994. 61～69

11 张琪玉. 关于自然语言检索问题. 图书馆论坛，2004，24(6)：211～213

12 Meng H, Siu K. Semiautomatic Acquisition of Semantic Structure for Understanding Domain Specific Natural Language Queries. IEEE Transactions on Knowledge and Data Engineering. 2002,14(1):172～181

13 Yin L. Topic Analysis and Answering Procedural Questions: [Technical Reports]. ITRI-04-14. Univ of Brighton, 2004

14 Sowa J. Knowledge Representation Logic, Philosophical, and Computational Foundations. Brooks/ Cole, 2000