

# 基于本体的股票主题事件案例推理系统研究

何岩新<sup>1,2</sup>,倪丽萍<sup>1,2</sup>,曹琳<sup>1,2</sup>,马驰宇<sup>1,2</sup>

(1. 合肥工业大学 管理学院,安徽 合肥 230009;  
2. 过程优化与智能决策教育部重点实验室,安徽 合肥 230009)

**摘要:**股票市场中一些事件容易影响股票的价格及收益。为辅助投资者积累经验性知识及快速做出决策,文中在收集主题事件的基本信息及市场波动数据的基础上,结合本体理论和案例推理方法,通过对事件案例的本体结构定义以及对金融事件类型本体的构建,开发了股票市场主题事件案例的推理系统原型。该系统用 OWL-DL 语言描述本体,采用 Protégé4.2 构建本体库;通过分析主题事件的特征属性,用相似性度量方法检索相似案例,进而对目标事件做出详细分析。通过特定事件对股票市场的走势分析,验证了用相似案例进行市场短期预测具有可行性,同时引入本体后在推理系统的查全率上有了明显的提高,从而说明该方法是合理的、有效的。

**关键词:**主题事件;本体;案例推理;案例库;股票市场

中图分类号:TP302.1

文献标识码:A

文章编号:1673-629X(2016)01-0033-06

doi:10.3969/j.issn.1673-629X.2016.01.007

## Research on Case Based Reasoning System of Stock Theme Events Based on Ontology

HE Yan-xin<sup>1,2</sup>, NI Li-ping<sup>1,2</sup>, CAO Lin<sup>1,2</sup>, MA Chi-yu<sup>1,2</sup>

(1. School of Management, Hefei University of Technology, Hefei 230009, China;  
2. Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of  
Education, Hefei 230009, China)

**Abstract:** In stock markets there are some events that their occurrences will significantly affect stocks' prices and profits. In order to help investors accumulate experimental wealth and quickly make decisions, in this paper a case-based reasoning system for subject events in stock markets was constructed with the combination of ontology theory and case-based reasoning method, then with information about subject events and data about market fluctuations, the ontology structures were defined for event cases and type ontologies were set for financial events. In this system, OWL-DL was used to describe ontology, and Protégé4.2 was applied to build the ontology base. By studying the feature properties and using similar measuring method to retrieve similar cases, a target event was finally analyzed in detail. In the end a trending analysis for a certain event to the stock markets was given, the results showed that using similar cases to predict short-term market effect was feasible, and meantime the recall ratio was obviously improved after introducing ontology into the reasoning system, which means that the proposed method is reasonable and effective.

**Key words:** theme events; ontology; Case-Based Reasoning (CBR); case base; stock market

### 1 概述

主题事件是指在股票市场中经常发生的可能影响市场波动的特定类别事件,如,银行调息、公司公布重大事件、举行世界性盛会等。在类似的市场环境下,发生类似的主题事件往往会对市场产生类似的效应。因此,对股票市场,构建主题事件案例库显得非常必要。通过记录已经发生的主题事件详细内容,包括事件的

描述、事件前的市场情况及事件导致的股票市场波动情况,将有利于投资者积累经验,提高投资决策能力。

案例推理(Case-Based Reasoning, CBR)是将过去求解问题的经验定义为案例并使用案例来求解新的类似的问题<sup>[1]</sup>,而对于事件案例推理系统来说,就是将事件表示成案例。通过对相似事件案例的检索,为新事件的发生结果提供辅助决策。然而,传统的 CBR 系统

收稿日期:2015-02-13

修回日期:2015-06-04

网络出版时间:2016-01-04

基金项目:国家自然科学基金青年基金项目(71301041)

作者简介:何岩新(1989-),男,硕士研究生,研究方向为数据挖掘、人工智能。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160104.1607.066.html>

大多是用静态框架来描述案例,存在难以扩展、可重构和学习性差,且对理论知识和领域资源缺乏良好的支持,不能有效利用语义关系等缺点<sup>[2]</sup>。本体(ontology)明确了领域内共同认可的概念术语,利用领域知识的语义模型表达了概念含义,并在内部层次中规定了这些概念之间的关系。它描述了概念及概念之间关系的模型,具有较强的知识表达能力和支持较强的推理能力。把本体与 CBR 相结合,可以把案例所表达的知识与用本体表示的一般领域知识模型进行整合,这样基于本体的案例推理过程可以充分利用领域知识来获取更精确和全面的结果<sup>[3]</sup>。

在过去的研究中,有不少学者提出将本体和案例推理相结合以提高决策能力。文献[4]将本体与 CBR 相结合开发了电子政务项目审批决策支持系统,与传统基于关键字的项目案例检索相比,引入本体技术后,其案例检索的查准率和查全率明显提高。文献[5]针对 CBR 系统缺乏语义理解的问题,提出将本体与 CBR 相结合,通过在注塑模具设计案例系统中的应用,说明该方法优于现有的 CBR 系统。文献[6]针对传统 CBR 存在的案例库难以扩展、表述词表不统一等异构问题,将本体技术引入 CBR,研究了基于本体 CBR 的旅游产品案例表示方法和案例检索策略。文献[7]构建基于本体的高速交通突发事件案例的本体模型,提供了对突发事件的发展过程与应急处置全过程知识的语义一致性描述,屏蔽了不同交通信息管理系统之间的异构性。文献[8]通过构建基于本体的应急响应决策支持系统,把本体与 CBR 相结合,应用到信息安全领域,从而共享成功的经验知识。文献[9]将基于本体的范例推理技术应用到常见问题回答服务领域,取得了不错的效果。文献[10]针对传统的基于句法的相似性度量方法存在匹配度不高的现象,提出采用本体理论和基于特征的相似性度量方法来确定两个问题的相似性,并用基于案例的推理实现产品信息的检索和再利用。

事件对股票市场的影响是毋庸置疑的,过去的研究主要通过事件研究法<sup>[11-13]</sup>实证分析某个事件的影响作用,缺乏系统性和全面性。因此文中将本体与 CBR 相结合,通过建立基于本体的股票市场主题事件案例模型及案例推理系统,利用本体和案例推理技术的优势对股票市场上发生的事件影响结果进行预测。相关实例说明该系统能够检索出相似的事件案例,对投资者规避风险具有一定的现实参考价值。

## 2 基于本体的案例推理系统

### 2.1 系统总体架构

基于本体的股票市场主题事件案例推理系统基于

B/S 架构,整个系统分为三层:应用层、业务逻辑层和数据存储层。其中应用层主要包括案例管理模块、案例学习模块、市场波动趋势预测模块和金融事件类型本体管理模块等。业务层用 Protégé 对本体进行定义,通过支持基于 RDFS 和 OWL 等语义推理的 Jena2 工具包来实现本体模型的解析和推理<sup>[4,6]</sup>。数据存储层以 MySQL 为主要的数据库管理系统,实现案例知识的存储。其体系架构图如图 1 所示。

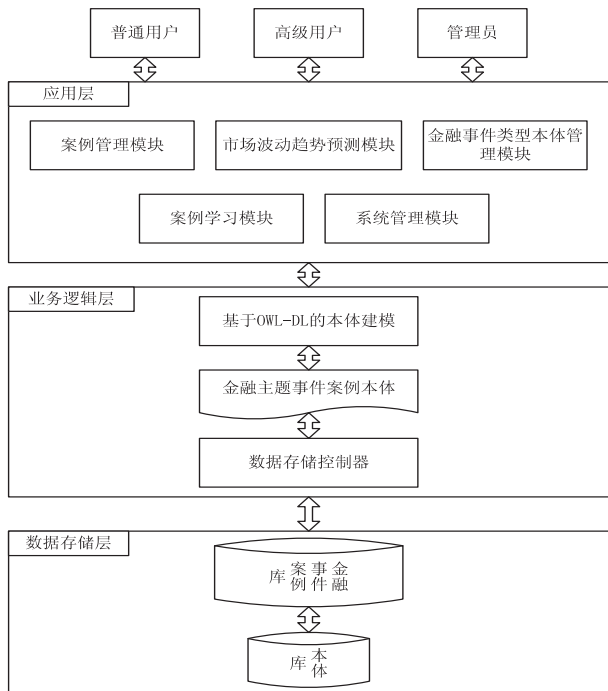


图 1 基于本体的案例推理系统体系架构图

### 2.2 系统主要功能模块介绍

(1)案例管理模块:该模块根据案例的表示结构,建立和维护事件案例,包括案例的添加、删除、修改等操作。

(2)市场波动趋势预测模块:该模块通过检索案例库,寻找与待查询事件相似的事件案例,根据检索得到的事件案例对市场短期波动趋势做出预测。

(3)案例学习模块:该模块通过对案例的学习和修正形成新的案例,丰富案例库。

(4)金融事件类型本体管理模块:该模块是高级用户操作的,即为领域知识维护模块。负责股票市场事件类型本体知识的管理。

(5)系统管理模块:该模块主要负责系统用户的管理以及系统设置。该系统根据权限,把用户划分为三类:普通用户可以进行市场波动趋势预测;高级用户是针对股票市场领域专家而设置的,其负责案例库的管理和维护以及金融事件类型本体管理;系统管理员负责管理用户的权限。

### 2.3 系统推理流程

股票市场主题事件案例推理系统的基本流程

如下:

Step1: 根据用户所提交的事件信息,包括事件时间、事件摘要、事件发生前市场波动情况、事件类型以及趋势等,在主题事件案例库中寻找相似案例;

Step2: 若没有找到相似案例,则存入最新事件案例,更新案例库;如果存在相似案例,则根据事件的相似度的计算结果排序,按照案例的相似度大小排序展现给用户,并列出生相似事件案例的详细信息;

Step3: 用户通过分析案例信息,对市场做出预测反应。

### 3 基于本体的主题事件案例表示与检索策略

#### 3.1 基于本体的案例表示

(1) 主题事件特征属性选择。

通过对股票市场主题事件的分析,发现主题事件的描述一般是基于文本的非结构化形式。因此需要收集这些事件知识,并抽取出事件属性来规范化事件案例,这样更有利于案例库的存储、维护以及推理功能的实现。

对于一个事件,一般应包含两类信息:事件的描述和事件的影响。此外不同的市场环境下事件的影响作用可能不同。因此,针对股票市场主题事件,文中将其抽象为一个三元组表,分别为事件描述、事件前市场描述和事件影响结果,即:  $Event\_Case: \{ Event\_Descrip-$

$tion, Market\_Description, Event\_Results \}$ 。

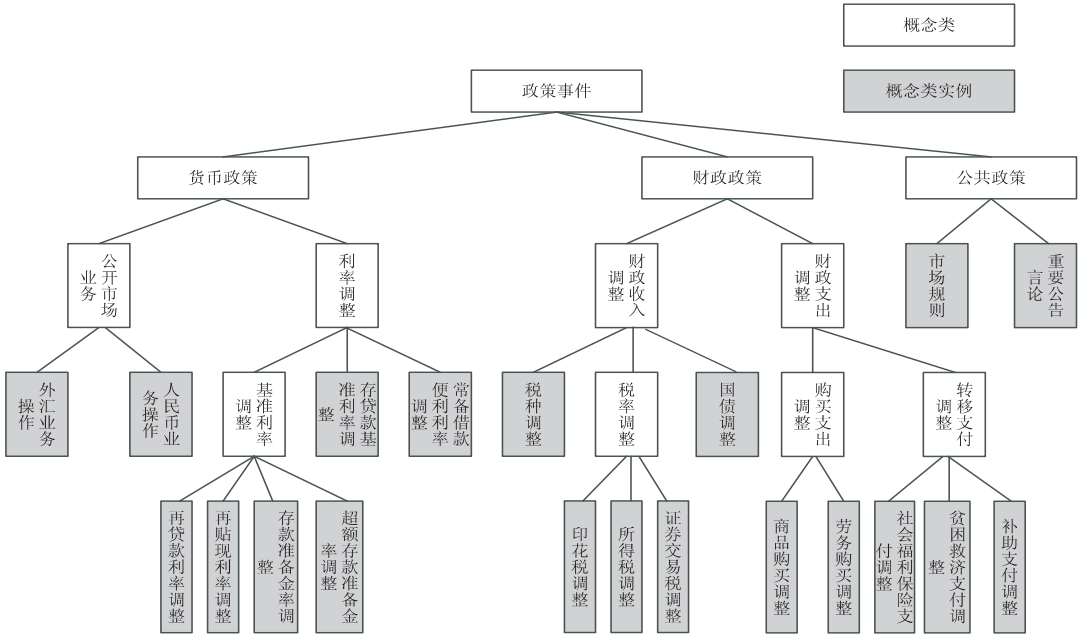
事件描述表示事件的基本信息,可以描述为  $Event\_Description: \{ Event\_ID, Event\_Time, Event\_Abstract, Event\_Type, Event\_Trend \}$ 。其中,  $Event\_ID$  表示“事件编号”,  $Event\_Time$  表示“事件发生时间”,  $Event\_Abstract$  表示“事件摘要”,  $Event\_Type$  表示“事件类型”,  $Event\_Trend$  表示“趋势”。对于事件发生时间,文中将事件公布日当做事件发生时间。

事件前市场描述表示为  $Market\_Description: \{ updown\_five, updown\_ten \}$ 。其中,  $updown\_five$  表示“事件前五天平均涨跌”,  $updown\_ten$  表示“事件前十天平均涨跌”。在股票市场中,股票涨跌反应了股价波动情况。通过计算事件发生前一段时间股票的涨跌值,能够了解当前市场的波动情况。

事件影响结果表示为  $Event\_Results: \{ Ten\_Day\_Market \}$ 。其中,  $Ten\_Day\_Market$  表示“十天股指表现”。文中选取事件公布后十天股指表现作为事件对股市短期影响的描述。

(2) 主题事件案例本体结构定义。

基于对主题事件属性选择后,通过分析发现,对于事件类型其涉及到的种类多,文中经过归纳整理将其分为货币政策事件、财政政策事件以及公共政策。在此基础上,构建了主题事件类型本体,通过该本体给出了影响市场主要事件类型的基本术语以及相互间的关系,其结构如图2所示。



综上,在主题事件案例推理系统中,案例本体结构如图3所示。

该系统所包含的两类本体—主题事件案例本体和主题事件类型本体,两者均属于轻量级的应用本体,

结构较简单,易于表达和实现<sup>[4]</sup>。

#### 3.2 基于本体的事件案例检索策略

对于基于本体的案例推理系统,其检索过程中相似度计算是很重要的环节,通过相似度的计算来衡量

查询案例和检索到的案例之间的相似关系。

似度计算公式可以表示成公式(1)。

一般地,基于本体的案例推理系统,其案例间的相

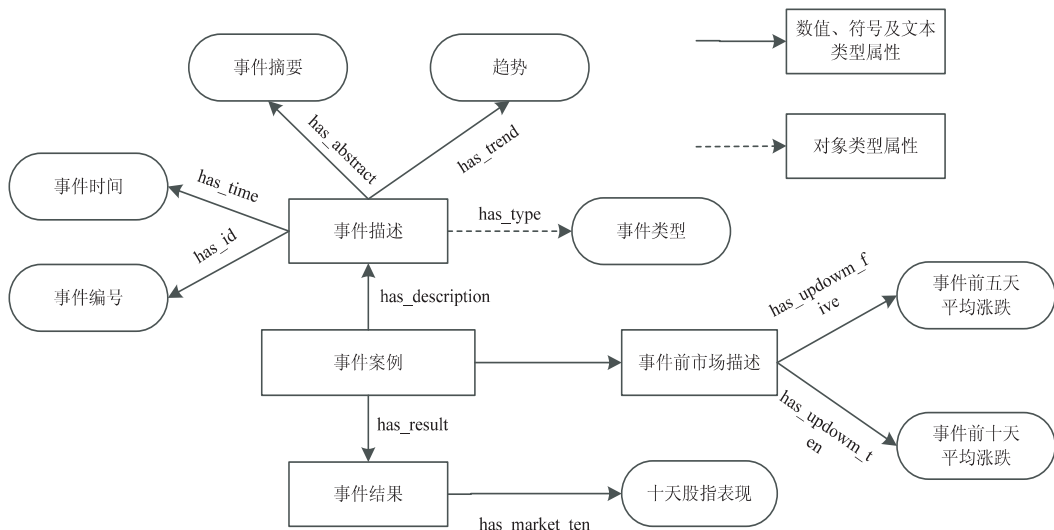


图3 案例本体结构

$$\text{Sim}(C_1, C_2) = \sum_{i=1}^n [w_i \times \text{Sim}(X_i, Y_i)] / \sum_{i=1}^n w_i \quad (1)$$

其中,  $\text{Sim}(C_1, C_2)$  表示两个案例的总体相似度;  $w_i$  表示第  $i$  个属性的权重;  $\text{Sim}(X_i, Y_i)$  表示第  $i$  个属性的相似度即案例局部相似度。

具体地,主题事件案例相似度计算可以分解为两个步骤:先计算案例间的局部相似度,再计算案例间的总体相似度。该系统中,对于属性分为四种类型:数值型、文本型、符号型和本体类中定义的本体类型属性。对于不同类型属性以下给出不同相似度计算公式。

### 3.2.1 数值型属性相似度计算

文中的属性“事件前五天平均涨跌”和“事件前十天平均涨跌”即为该类型。数值型属性的局部相似度可表示为公式(2)。

$$\text{Sim}(X_i, Y_i) = 1 / (1 + \sqrt{(x_i - y_i)^2}) \quad (2)$$

其中,  $x_i$  和  $y_i$  分别表示案例  $X$  和案例  $Y$  的第  $i$  个属性值,当  $|x_i - y_i|$  值越大时,其相似度值越小。

### 3.2.2 文本型属性相似度计算

文中的属性“事件摘要”即为文本型属性。首先对该文本属性值进行预处理,包括中文文本分词、去除停用词以及文本特征项选择。并通过构造文本的特征向量,用向量间的夹角余弦值来衡量文本相似度。

假设给定文本  $d$ , 其特征项集合表示为  $d(t_1, t_2, \dots, t_n)$ , 特征向量表示为  $\vec{V}_d = (w_1, w_2, \dots, w_n)$ 。其中,  $t_i$  是文本的特征词;  $w_i$  表示  $t_i$  在文本  $d$  中的权重,这里选择词频作为文本特征项的权重。

假设两个文本  $d_i$  和  $d_j$  的特征向量分别为  $\mathbf{V}_{d_i} = (w_{i1}, w_{i2}, \dots, w_{in})$  和  $\mathbf{V}_{d_j} = (w_{j1}, w_{j2}, \dots, w_{jn})$ 。其相似度计算可表示为公式(3)。

$$\text{Sim}(d_i, d_j) = \cos \theta = \frac{\sum_{k=1}^n w_{ik} * w_{jk}}{\sqrt{(\sum_{k=1}^n w_{ik}^2)(\sum_{k=1}^n w_{jk}^2)}} \quad (3)$$

其中,  $\theta$  表示两特征向量在空间中的夹角;  $\text{Sim}(d_i, d_j)$  表示  $d_i$  和  $d_j$  的相似度,  $\text{Sim}(d_i, d_j)$  越大表示两个文本之间的相似度越高。

### 3.2.3 符号型属性相似度计算

符号型属性值属于一种简单的枚举值,其值列举了该属性所有可能的取值。文中属性“趋势”即为该类型,其值包含三个值:上调、下调和无。其相似度计算表示为公式(4)。

$$\text{Sim}(X_i, Y_i) = \begin{cases} 1, & X_i = Y_i \\ 0, & X_i \neq Y_i \end{cases} \quad (4)$$

### 3.2.4 本体类型属性相似度计算

对于本体类型属性的,其值是本体中的某个概念或者实例,文中属性“事件类型”即为该类型。可以利用概念(或实例)的相似度来衡量相应属性值的相似度,而概念(或实例)的相似度通过计算树状结构中的节点距离来计算。计算本体类型特征值的局部相似度,假设两个实例  $i_1$  和  $i_2$ , 根据文献[14], 计算相似度可表示为公式(5)。

$$\text{Sim}(i_1, i_2) = \begin{cases} \frac{2 \times (\text{Depth}(\text{Msc}(i_1, i_2)) - 1)}{\text{Dis}(i_1, i_2) + 2 \times (\text{Depth}(\text{Msc}(i_1, i_2)) - 1)}, & i_1 \neq i_2 \\ 1, & i_1 = i_2 \end{cases} \quad (5)$$

其中,  $\text{Msc}(i_1, i_2)$  表示两个实例  $i_1$  和  $i_2$  的最近父节点;  $\text{Depth}(\text{Msc}(i_1, i_2))$  表示最近父节点在树状结构中的深度(父节点到根节点经过的节点数);  $\text{Dis}(i_1,$



$i_2$ ) 表示两个实例的最短距离(实例  $i_1$  和  $i_2$  经过的最少节点数)。当  $\text{Depth}(\text{Msc}(i_1, i_2))$  或  $\text{Dis}(i_1, i_2)$  值越小时,两个实例间的相似度也就越大。

根据图 2,利用公式(5)可以得到表 1 中一组实例间的相似度值。

表 1 事件类型概念树中实例相似度计算结果

相似度	相似度值
Sim(存款准备金率调整,再贷款利率调整)	2/3
Sim(存贷款利率调整,存款准备金率调整)	1/2
Sim(再贷款利率调整,印花税调整)	0
Sim(印花税调整,补助支付调整)	2/9

4 系统实现与实例分析

基于本体的股票市场主题事件案例推理系统采用的开发语言是 Java。编程开发环境是 MyEclipse,数据库采用的是 MySQL,操作系统为 Windows 7。推理系统采用 OWL-DL 语言描述本体,采用 Protégé4.2 作为本体库构建工具。

文中系统主要考察事件对 A 股短期走势的影响,结合主题事件案例本体结构,构建了主题事件案例本体和相关实例。

图 4 是在 Protégé4.2 下创建的金融主题事件案例本体及实例。



图 4 在 Protégé4.2 金融主题事件案例本体及实例

以图 4 中构建的事件案例为输入案例进行检索。检索出来相似事件案例的详细信息及相似度值如图 5 所示。其中权重通过专家调查并结合实验数据,设置为  $\vec{W} = (0.2, 0.1, 0.20, 0.25, 0.25)$ , 分别对应“事件前五天平均涨跌”、“事件前十天平均涨跌”、“事件摘要”、“趋势”和“事件类型”的权重值。

由于查询案例发生后市场短期内表现为下跌,因此由检索结果可知,相似度值高的案例走势与查询检索案例发生后市场短期走势具有一定的相似性,通过案例检索进行市场短期预测具备可行性。同时,引入本体后能够有效提高案例推理系统的查全率,能够检

索出如“存贷款利率调整”等相关事件类型的案例,从而提供更全面的参考。表明该系统设计思想是合理的,实现方法是有效的。

5 结束语

基于本体的股票市场主题事件案例推理系统,充分利用本体的优势来提高事件案例检索的查全率和准确率。通过对事件案例的本体结构定义以及对金融事件类型本体的构建,开发了股票市场主题事件案例推理系统原型,通过具体事件案例进行了仿真测试,验证了该系统在案例推理中的合理性。



图 5 案例检索结果界面截图

下一步的工作重点是对案例库的进一步完善以及对属性权重设置方法的研究。

[8] Liu Ping, Yu Haifeng, Ma Guoqing. An incident response decision support system based on CBR and ontology[C]//Proc of international conference on computer application and system modeling. [s. l. ]:[s. n. ],2010:337-340.

参考文献:

[1] Kolodner J L. An introduction to case-based reasoning[J]. Artificial Intelligence Review,1992,6(1):3-34.

[2] 高俊杰,邓贵仕. 基于本体的范例推理系统研究综述[J]. 计算机应用研究,2009,26(2):406-410.

[3] 谢红薇,李建伟. 基于本体的案例推理模型研究[J]. 计算机应用研究,2009,26(4):1422-1424.

[4] 刘晓文,胡克瑾. 基于本体和 CBR 的电子政务项目审批决策支持系统[J]. 计算机应用,2009,29(3):896-899.

[5] Guo Yuan,Hu Jie,Peng Yinghong. A CBR system for injection mould design based on ontology:a case study[J]. Computer Aided Design,2012,44(6):496-508.

[6] 虞娟. 基于本体的 CBR 及其在旅游产品智能推荐系统的应用研究[J]. 哈尔滨师范大学自然科学学报,2013,29(6):67-70.

[7] 李文雄,闫茂德,王建伟. 高速公路交通突发事件案例的本体模型及应用[J]. 微电子学与计算机,2013,30(3):55-58.

[9] Yang Shengyuan,Liao Penchin. An ontology-supported case-based reasoning technique for FAQ proxy service[C]//Proc of international conference on software engineering and knowledge engineering. [s. l. ]:[s. n. ],2005:693-644.

[10] Akmal S,Shih L H,Batres R. Ontology-based similarity for product information retrieval[J]. Computers in Industry,2014,65(1):91-107.


[11] 孙伶俐. 股票市场上利率政策公告效应实证研究[J]. 中南财经政法大学学报,2008(6):65-70.

[12] 刘洋. 存款准备金率调整对我国证券市场的影响[J]. 统计研究,2008,25(3):42-45.

[13] 谢敏. 谈印花税下调对沪市的影响[J]. 知识经济,2009(6):36-37.

[14] Wu Z,Palmer M. Verb semantics and lexical selection[C]//Proceedings of the 32nd annual meeting of the associations for computational linguistics. LasCruces, New Mexico:[s. n. ],1994:133-138.

基于本体的股票主题事件案例推理系统研究

作者：[何岩新](#)，[倪丽萍](#)，[曹琳](#)，[马驰宇](#)，[HE Yan-xin](#)，[NI Li-ping](#)，[CAO Lin](#)，[MA Chi-yu](#)  
作者单位：[合肥工业大学 管理学院](#)，[安徽 合肥 230009](#)；[过程优化与智能决策教育部重点实验室](#)，[安徽 合肥 230009](#)  
刊名：[计算机技术与发展](#)  
英文刊名：  
年，卷(期)：2016(1)

引用本文格式：[何岩新](#).[倪丽萍](#).[曹琳](#).[马驰宇](#).[HE Yan-xin](#).[NI Li-ping](#).[CAO Lin](#).[MA Chi-yu](#) [基于本体的股票主题事件案例推理系统研究](#)[期刊论文]-[计算机技术与发展](#) 2016(1)