

# 市场分析与预测的一种文字报告 生成方法与实现\*

郑彤 马少平 陆玉昌

(清华大学计算机系 北京 100084)

**摘要** 市场分析与预测系统是一种为企业决策者提供市场信息的决策支持系统. 本文对一种市场分析与预测系统的总体结构进行了讨论, 并在此基础上提出和实现了基于模板的文字报告生成方法.

**关键词** 决策支持系统, 市场分析, 市场预测, 专家系统, 报告生成, 规范数据, 模板.

准确、有效地掌握市场行为对于企业的生存和发展至关重要, 而市场行为的决策很大程度上依赖于决策者对市场情况所作的分析和判断. 如何最大限度地利用已掌握的各种信息和数据, 对市场情况及其趋势作出正确的分析和预测, 进而改善营销策略, 提高市场行为决策的质量, 就成为一个普遍关注的研究课题. 市场分析与预测正是基于这种出发点, 以管理科学、系统科学、行为科学为基础, 以数学、信息技术、计算机技术为手段, 及时、准确地提供信息, 辅助管理者进行决策活动. 在这种意义下, 市场分析与预测的作用在于辅助支持而不是取代决策者, 决策的最终方案仍由决策者决定.

所谓市场分析与预测报告就是指在对已掌握的经济信息、市场信息以及生产数据进行分析并对未来情况作出预测之后得到的文字报告.

从报告内容上看, 市场分析与预测报告可分为常规报告和随机报告 2 种. 其中, 常规报告是指内容相对固定、每次都要做的报告, 如基本经济情况(人口、国民生产总值等)、企业经营概况等; 随机报告则是由数据情况触发的报告, 如市场竞争态势及其变化趋势等.

本文将介绍一种市场分析与预测系统的结构模型, 并对其中的文字报告生成方法以及数据规范化处理进行较为详细的讨论.

## 1 总体结构与框图

### 1.1 市场分析与预测系统

市场分析与预测系统 MAFS(market analysis forecast system), 是致力于提高市场分

\* 作者郑彤, 1971 年生, 硕士生, 主要研究领域为人工智能及知识工程. 马少平, 1961 年生, 副教授, 主要研究领域为知识工程, 汉字识别. 陆玉昌, 1937 年生, 教授, 主要研究领域为人工智能.

本文通讯联系人: 陆玉昌, 北京 100084, 清华大学计算机系

本文 1996-01-30 收到修改稿

析与预测信息质量的一个综合信息处理系统.从功能上来讲,MAFS 实际上是一个决策支持系统.MAFS 从数据库中接收所需的原始数据,进行分析和预测之后,再将所得结果存入数据库中以便查询,同时生成相应的文字报告.在必要的情况下还可以加上一定的辅助信息,以丰富输出画面.

## 1.2 系统总体结构

MAFS 从功能上分,由数据规范化处理模块(DSP)、专家系统(ES)、市场分析模块(MA)、市场预测模块(MF)以及报告生成模块(RG) 5 部分组成.

(1) 数据规范化处理模块(DSP) 出于决策的需要,企业必然尽可能地收集各种市场信息和生产数据,因此也就存在原始数据来源多、结构混乱的问题.DSP 的目标旨在对原始数据进行一定的预处理,从而得到结构简单、格式单一的规范化数据,供专家系统、分析、预测和报告生成使用.

(2) 专家系统(ES) 通过长期的工作实践,有经验的企业管理人员和市场营销人员必然会对市场的变化和发展形成一套较为完整的看法,将这些知识总结成规则的形式即可通过专家系统加以运用.ES 的功能是运用已定义的经验性规则,基于实际的数据情况进行推理,以发现一些值得注意的问题.

(3) 市场分析模块(MA) 不难想到,有限的经验性知识无法覆盖可能出现的全部情况,因此,还需要从实际数据中发掘出知识中没有提到但有可能对企业行为产生影响的信息.在这种情况下,除了采用经验性知识进行分析推理之外,基于数学方法的统计分析也是必不可少的.MA 包括运用几种常用的统计分析方法(因果分析、趋势分析等)以及一些经验性公式,在已有数据的基础上进行一定的数学运算,以得出一些决策者可能会关心的分析结果.

(4) 市场预测模块(MF) MF 的功能为在现有数据的基础上,对未来(主要是短期内)的基本经济情况、市场销售情况等作出预测,以便决策者在考虑下一步行动时有所借鉴,不至于对将来的情况一无所知.由于市场情况的多样性和复杂性,单纯使用数学上的预测方法难以收到满意的效果,因此,在本系统中我们引入了一种数学方法与经验知识相结合的综合组合预测方法.

(5) 报告生成模块(RG) RG 的功能即为根据原始数据以及专家系统、分析和预测所得到的结果,生成市场分析和预测的文字报告,必要时配以辅助信息.在本系统中我们采用的是一种基于模板的文字报告生成方法.

## 1.3 系统流程及其框图

在运行本系统之前,一定的数据准备工作是必要的,所有需要用到的原始数据都应按照 DSP 所能处理的格式在数据库中存放好.在此之后,即可通过一定的参数(如时间等)调用本系统.系统的工作步骤如下所示:

- (1) 调用 DSP,对原始数据进行规范化处理,得到规范数据;
- (2) 调用 ES,在规范数据的基础上运行专家系统,得到规则结果;
- (3) 调用 MA,在规范数据的基础上进行分析,得到分析结果;
- (4) 调用 MF,在规范数据的基础上进行预测,得到预测结果;
- (5) 调用 RG,在上述结果的基础上生成相应的文字报告(以及必要的辅助信息).

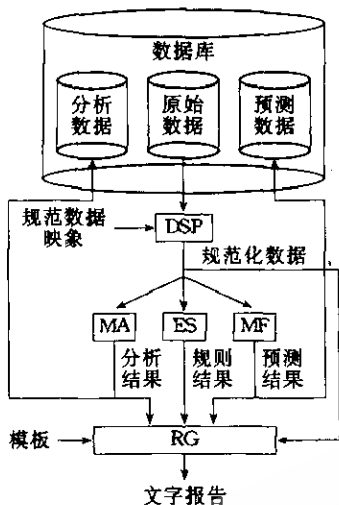


图1

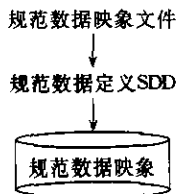


图2

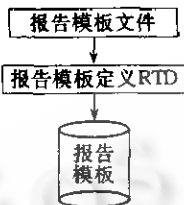


图3

MAFS 的总体框图见图 1.

## 2 一种数据规范化的方法

### 2.1 数据规范化的意义

从上面的介绍可以看出,数据的规范化处理是整个市场分析和预测的第 1 步.此作法具有以下优点:

(1) 数据共享 专家系统、分析、预测以及报告生成都使用预处理后得到的规范化数据,从而使整个市场分析和预测系统具有一致的数据格式,与原始数据只需唯一的数据接口.

(2) 数据独立性 ①市场分析和

预测系统所使用的数据通过一个数据预处理程序(规范数据生成程序)与原始数据隔离开来,从而具有数据独立性.②数据独立性使得数据查询也相应地具有独立性.当原始数据的格式改变时,只需相应地修改数据预处理程序(规范数据生成程序)中的数据源描述,规范数据表结构以及相应的查询程序则保持不变或仅需做少量修改.

(3) 一致的数据格式 规范数据的存储具有相对一致的数据格式,使得数据查询过程简单、清晰.

### 2.2 规范数据映象

(1) 规范数据映象的定义 原始数据的规范化处理是以规范数据映象为依据进行的.规范数据映象,就是从原始数据到规范数据的一种对应关系.鉴于 MAFS 对数据要求的灵活性,规范数据映象对用户来说是可维护的.

(2) 规范数据映象的维护 规范数据映象的维护是通过规范数据映象文件来完成的.当用户觉得有必要修改规范数据映象时,只需在该文件中做相应的改动,然后再重新安装即可.其过程如图 2 所示.

## 3 基于模板的文字报告生成方法

### 3.1 模板定义

MAFS 是以文字报告以及其他一些辅助信息作为分析和预测结果的最终表示形式.为了生成文字报告,一份文字脚本是必不可少的.同时,为了生成必要的辅助信息,还必须要有相应的辅助定义.

报告模板是规定文字报告内容及其辅助信息形式的一种规范化脚本知识.

在报告模板中,必须对以下内容进行定义:报告的文字内容;数据查询方式;相应的辅助信息.

### 3.2 模板结构

在 MAFS 中,我们将报告模板的结构划分为 4 个层次:

报告模板 = (专题,段,短语,单元).

专题,指针对某一个特定主题的所有报告.

段,指专题中描述某一个特定问题的一段报告.

短语,指一个段中存在相互关系的一些句子组合.

单元,指报告模板的最小单位.一个单元最多只能涉及一次数据查询.

报告模板的基本构架可用巴科斯范式表示如下:

报告模板 ::= [专题]

专题 ::= [段]

段 ::= (填充项说明) (段控制结构) [短语]

短语 ::= (短语相关条件) [单元]

单元 ::= 文本单元 | 控制单元

文本单元 ::= (单元文本) (填充项)

[...]:可重复多次 (...):可缺省

一份报告模板由多个专题组成,一个专题由多个有顺序的段组成,一个段由多个有顺序的短语组成,一个短语则由多个有顺序的单元组成.其中短语体现报告的灵活性:每个段的短语和短语之间可能存在相关关系,这种关系在短语相关条件中进行说明.生成报告时,某些短语的存在或不存在会影响到其他一些短语是否加入.一般地,单个短语不会影响整个段.

此外,一个单元可以是文本单元或控制单元.其中文本单元分为单元文本和填充项,分别定义报告文本内容及其数据查询方式,控制单元则用于定义其他的辅助信息.文本单元中使用的填充项必须在填充项说明中进行定义,同样,控制单元中所用到的辅助控制信息也必须在段控制结构中进行定义.

### 3.3 模板维护

模板的维护是通过模板文件来完成的.当用户觉得有必要修改模板时,只需在该文件中做相应的改动,然后再重新安装即可.其过程如图 3 所示.

## 4 系统实现

本系统以玉溪卷烟厂的“信息管理与决策支持系统(IMDSS)”为背景,以 DEC VAX 10000 AXP/610(OSF/1 2.0,SYBASE SYSTEM 10)和 SGI 的 INDIGO2 及 INDY(IRIX 5.2)为开发平台进行了系统实现.

IMDSS 是为玉溪卷烟厂领导提供全国经济状况,全国卷烟生产、销售形势,本厂生产、经营、销售等信息的一个综合系统.该系统可以提供对各种经济形势及卷烟市场的分析和预测,对全国卷烟生产、市场形势及发展趋势以及玉溪卷烟厂生产状况和玉溪地区烤烟种植情况给出评论,并对这些评论给出解释,为领导决策提供支持;同时为全厂的各级领导和各科室提供各种信息查询.本系统也具有制定生产计划和销售计划的功能.输出有语音报告、可视化模型、统计图形、简表、卡片、打印等多种形式.

MAFS 作为 IMDSS 中的一个重要组成部分,接收来自玉溪卷烟厂的 MIS 数据以及来

自国家统计局的全国卷烟生产和销售数据(均已转换成 SYBASE 格式),按照给定的时间、地区范围和专题(如:基本经济情况、市场销售情况、烟厂生产情况等)自动生成市场分析和预测的文字报告,同时给出相应的辅助信息。

**致谢** 本项工作在进行过程中,自始至终得到石纯一、周立柱老师的关心和指导,王令赤、刘卫东老师也给予了很大帮助和支持,在此表示衷心感谢。

### 参考文献

- 1 Mittal Vibhu O, Paris Cecile L. Automatic documentation generation; the interaction between text and examples. Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, Chambéry, 1993.
- 2 Mittal Vibhu O, Paris Cecile L. Generating natural language descriptions with examples; differences between introductory and advanced texts. Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI'93), Washington DC, 1993.
- 3 Mittal Vibhu O, Paris Cecile L. The placement of examples in descriptions; before, within or after the text. Proceedings of the First Conference of the Pacific Association for Computational Linguistics, Vancouver BC, 1993.

## A TEXT REPORT GENERATION METHOD AND ITS IMPLEMENTATION FOR MARKET ANALYSIS AND FORECAST SYSTEM

Zheng Tong Ma Shaoping Lu Yuchang

(Department of Computer Science Tsinghua University Beijing 100084)

**Abstract** Market analysis and forecast system is a decision support system that provides market information for decision makers of companies. In this paper, the architecture of a market analysis and forecast system is discussed, and a template based method of text report generation and its implementation in the system are presented.

**Key words** Decision support system, market analysis, market forecast, expert system, report generation, standardized data, template.