KNOWLEDGE DRIVEN DATA MINING FOR CAUSAL RELATIONSHIPS BETWEEN NEWS AND FINANCIAL INSTRUMENTS

WANG SHAN SHAN

DOCTOR OF PHILOSOPHY
CITY UNIVERSITY OF HONG KONG
AUGUST 2009
Knowledge driven data mining for causal relationships between news and financial instruments

Submitted to
Department of Information Systems
in Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy

by

Wang Shan Shan

August 2009
ABSTRACT

As news becomes more and more important in financial instruments trading algorithms, financial industry observers, investors and other analysts in financial markets are paying more attention to news. Some financial services companies have designed services which filter unrelated news information and documents and text information such as SEC or regulatory filings, and corporate web site information, which is another data source that financial investors can consider into their trading models. Besides financial services companies, some news publishers also provide similar services for their customers. However, after filtering, news served to financial brokerages and investors still need further human judgements for exploring the implications of news content and distinguishing significant from non-significant news, and for finding out the impact polar type of each kind of significant news. But these judgements are always limited by human information processing capability. Thus, in order to support more objective decision making, an ontology based framework, for investigating the relationships between news and financial instruments trading activities qualitatively and quantitatively, is proposed. This thesis contains two separate, but interrelated parts.

The first part is about an ontology, provided for demonstrating the domain knowledge about news in financial markets. The ontology model comprises two components. One is represented using OWL DL (which is a sub-language of Web Ontology Language), which provides a hierarchical framework for the domain knowledge, including primary classes of news, classes of financial markets participants, classes of financial instruments, and relations between these classes. This component is a specification of domain-specific vocabulary terms. The other component is a causal map, used to demonstrate the impact of different classes of
news events on financial instruments. It is of either a direct or an indirect “cause-effect” form, which can be written as rules using OWL rules language.

The second part is about an ontology based data mining framework designed to study the quantitative relationships between news and financial instruments trading activities. The framework is made of three components. The first is editing of the ontology from the first part with Protégé software tool. It is used to classify news and stocks into different groups according to the nature of businesses and financial instruments, and the news categories defined in the ontology model, when news and financial instruments data come into the framework. The second part is an expert-rules reasoning system implemented in Jess Shell, a plug-in for the Protégé tool. For a given financial instrument trading activity, it can indicate the possible significant news, and generate a Bayesian network model for the specific financial instrument. The third part is Bayesian network algorithm. Combined with the data mining model, this algorithm can specify the quantitative relationships between the possibly significant news and the given financial instrument trading activity.

The major contributions of this research is that the ontology helps understand the knowledge about news in financial markets, helps build trading models based on news, and builds systems for prediction of stock prices based on news. The ontology based data mining framework provides an ontology method for classifying news and financial instruments data, besides an expert reasoning system to integrate the background (domain) knowledge with current news.
# TABLE OF CONTENT

ABSTRACT.................................................................................................................................................. i
ACKNOWLEDGEMENTS.......................................................................................................................... iii
TABLE OF CONTENTS............................................................................................................................. iv
LIST OF FIGURES.................................................................................................................................. vii
LIST OF TABLES....................................................................................................................................... viii

1 INTRODUCTION.................................................................................................................................... 1
   1.1 RESEARCH QUESTIONS.................................................................................................................. 2
   1.2 RESEARCH OBJECTIVES.............................................................................................................. 5
   1.3 RESEARCH METHODOLOGIES.................................................................................................... 7
       1.3.1 DESIGN AS AN ARTIFACT.................................................................................................... 9
       1.3.2 PROBLEM RELEVANCE..................................................................................................... 10
       1.3.3 DESIGN EVALUATION....................................................................................................... 11
       1.3.4 DESIGN AS A SEARCH PROCESS...................................................................................... 12
   1.4 OVERVIEW OF THE THESIS........................................................................................................ 14

2 LITERATURE REVIEW......................................................................................................................... 16
   2.1 NEWS AND FINANCIAL INSTRUMENTS...................................................................................... 16
       2.1.1 NEWS IN THE FINANCIAL MARKET.................................................................................. 16
       2.1.2 FINANCIAL INSTRUMENTS................................................................................................ 20
       2.1.3 CAUSAL RELATIONSHIPS BETWEEN NEWS AND FINANCIAL INSTRUMENTS.............. 24
   2.2 INFORMATION SYSTEMS AND ALGORITHMS APPLIED IN THIS DOMAIN............................... 27
2.3 ONTOLOGY-----------------------------------------------33
  2.3.1 ONTOLOGY-----------------------------------------------33
  2.3.2 CAUSAL MAPS--------------------------------------------39
2.4 DATA MINING---------------------------------------------41
  2.4.1 DATA MINING METHODS----------------------------------42
  2.4.2 BAYESIAN NETWORK TECHNIQUE-----------------------------44

3 AN ONTOLOGY FOR CAUSAL RELATIONSHIPS BETWEEN NEWS
AND FINANCIAL INSTRUMENTS----------------------------------50
  3.1 THE REASONS FOR REPRESENTING THE RELATIONSHIPS
BETWEEN NEWS AND FINANCIAL INSTRUMENTS USING
ONTOLOGY------------------------------------------------------50
  3.2 ONTOLOGY-----------------------------------------------53
    3.2.1 CLASSES OF THE ONTOLOGY-------------------------------53
    3.2.2 RELATIONSHIPS BETWEEN NEWS AND PARTICIPANTS------55
    3.2.3 RELATIONSHIPS BETWEEN PARTICIPANTS AND FINANCIAL
    INSTRUMENTS------------------------------------------------56
    3.2.4 RELATIONSHIPS BETWEEN NEWS AND FINANCIAL
    INSTRUMENTS------------------------------------------------58
  3.3 CAUSAL MAPS IN THE DOMAIN---------------------------------60
    3.3.1 DOMAIN SCENARIOS ANALYSIS-----------------------------60
    3.3.2 CAUSAL MAPS IN THE DOMAIN------------------------------62
  3.4 CASE ANALYSIS---------------------------------------------67

4 AN ONTOLOGY BASED FRAMEWORK FOR MINING THE CAUSAL
RELATIONSHIPS BETWEEN NEWS AND FINANCIAL
INSTRUMENTS-----------------------------------------------74
  4.1 LIMITATIONS IN TRADITIONAL METHODS USED TO INVESTIGATE
LIST OF FIGURES

FIGURE 1 TYPES OF ARROWS……………………………………………………54
FIGURE 2 CLASSES OF THE ONTOLOGY………………………………………..54
FIGURE 3 RELATIONSHIPS BETWEEN NEWS AND FINANCIAL MARKET PARTICIPANTS…………………………………………………………….55
FIGURE 4 RELATIONSHIPS BETWEEN FINANCIAL INSTRUMENTS MARKET PARTICIPANTS AND FINANCIAL INSTRUMENTS….57
FIGURE 5 RELATIONHIPS BETWEEN NEWS AND FINANCIAL INSTRUMENTS………………………………………………………………………………59
FIGURE 6. CAUSAL MAP FOR PARTIAL RELATIONS BETWEEN NEWS AND FINANCIAL INSTRUMENTS………………………………………………66
FIGURE 7 GRAPH FOR 9/11 CASE ANALYSIS………………………………..68
FIGURE 8 GRAPH FOR INVESTORS.MOOD(+).………………………………..70
FIGURE 9 THE LOGICAL ARCHITECTURE OF THE FRAMEWORK………..82
FIGURE 10 THE EXAMPLE FOR THE DATA MINING MODEL……………..84
LIST OF TABLES

TABLE 1 OWL DL DEFINITION OF THE CLASS “NEWS” ......................... 54
TABLE 2 OWL DL DEFINITION OF THE OBJECTPROPERTY “PERCEIVED
BY THROUGH MEDIUM” ............................................................. 56
TABLE 3 OWL DL DEFINITION OF THE OBJECTPROPERTY
“RECOMMEND” ......................................................................... 58
TABLE 4 OWL DL DEFINITION OF THE DATATYPE “IMPORTANT FACTORS
FOR ECONOMY” ...................................................................... 59
TABLE 6 DESCRIPTION FOR CLASS “FINANCIAL INSTRUMENT MARKET
PARTICIPANTS” ......................................................................... 63
TABLE 7 DESCRIPTION FOR CLASS “EUIITY RELATED INSTRUMENTS” ..63
TABLE 8 DESCRIPTION FOR CLASS “COMPANY” .............................. 63
TABLE 9 DATAPROPERTY OF VARIABLES REPORTED IN NEWS .......... 65
TABLE 10 DATAPROPERTY OF VARIABLES IN PARTICIPANTS .......... 65
TABLE 11 DATAPROPERTY OF VARIABLES IN COMPANY ............... 65
TABLE 12 DATAPROPERTY OF VARIABLES IN EQUITY RELATED
INSTRUMENTS ......................................................................... 66
TABLE 13 PARTIAL OWL DL DEFINITION OF INSTANCE
“TERROR-ATTACK-9/11 NEWS” ............................................... 67
TABLE 14 A RULE FOR JUDGING THE RELEVANT NEWS EVENTS OF CPCC
STOCK PRICE .......................................................................... 87
TABLES 15 A RULE FOR JUDGING THE IMPACT POLAR TYPE OF
RELEVANT NEWS ...................................................................... 87
TABLE 16 A RULE FOR CONSTRUCTING DATA MINING MODEL FOR CPCC
STOCK PRICE .......................................................................... 87
TABLE 17 DESCRIPTION OF THE STOCK DATA SAMPLE ................. 92
<table>
<thead>
<tr>
<th>Table</th>
<th>Description/Results</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 18</td>
<td>Description of the News Data Sample</td>
<td>93</td>
</tr>
<tr>
<td>Table 19</td>
<td>Results of Total Precision for Stock Price</td>
<td>96</td>
</tr>
<tr>
<td>Table 20</td>
<td>The Significance Order of News Events on Stock Price</td>
<td>97</td>
</tr>
</tbody>
</table>