Module Interface Specification for Stock Prediction System

Renjie Zhang

December 18, 2017

1 Revision History

Date	Version	Notes
22/11/2017	1.0	Create
28/11/2017	1.1	Update
16/12/2017	1.2	Update

2 Symbols, Abbreviations and Acronyms

NA

Contents

3 Introduction

The following document details the Module Interface Specifications for Stock Prediction System which is used to predict the future stock price based on the historical data. Complementary documents include the System Requirement Specifications and Module Guide. The full documentation and implementation can be found at https://github.com/renjiezhang/CAS-741.

4 Notation

The structure of the MIS for modules comes from **?**, with the addition that template modules have been adapted from **?**. The mathematical notation comes from Chapter 3 of **?**. For instance, the symbol := is used for a multiple assignment statement and conditional rules follow the form $(c_1 \Rightarrow r_1 | c_2 \Rightarrow r_2 | ... | c_n \Rightarrow r_n)$.

The following table summarizes the primitive data types used by Stock Prediction System.

Data Type	Notation	Description
character	char	a single symbol or digit
integer	$\mathbb{Z}$	a number without a fractional component in (-∞, ∞)
natural number	$\mathbb{N}$	a number without a fractional component in [1, ∞)
real	$\mathbb{R}$	any number in (-∞, ∞)
List	list	a list of objects
Stock Record	record	a Record has two elements a date by string and a price by real number
Date	date	a date type with format yyyy-mm-dd

The specification of Stock Prediction System uses some derived data types: sequences, strings, and tuples. Sequences are lists filled with elements of the same data type. Strings are sequences of characters. Tuples contain a list of values, potentially of different types. In addition, Stock Prediction System uses functions, which are defined by the data types of their inputs and outputs. Local functions are described by giving their type signature followed by their specification.

5 Module Decomposition

The following table is taken directly from the Module Guide document for this project.

M1: Hardware-Hiding Module

M2: Main Module

M3: Data Input Module

M4: Price Volatility Module

M5: Price Momentum Module

M6: Prediction Module

M7: Kernelling Module

M8: Plot Module

M9: RDD Module

Level 1	**Level 2**
Hardware-Hiding	
Behaviour-Hiding Module	Main Module Data Input Module Price Volatility Module Price Momentum Module Prediction Module Data Plot Module
Software Decision Module	RDD Module Kernelling Module

Table 1: Module Hierarchy

6 MIS of Main Module (M??)

6.1 Module

Main

6.2 Uses

Hardware-Hiding Module (M??)

6.3 Syntax

NA

6.3.1 Exported Access Programs

Name	In	Out	Exceptions
main	-	-	-

6.4 Semantics

6.4.1 State Variables

NA

6.4.2 Enviroment Variables

ndxtPrices: list of real numbers ndxtDates: list of dates numDaysArray : list of integers numDayAheadArray : list of integers

6.4.3 Access Routine Semantics

main():

- transition: Calls the Data Input Modules, Volatility Module, Momentum Module and Predict Module with parameters and retrieve the returned data from them.
  -DataInput(filePath)
  -ndxtVolatilityArray = sc.parallelize(GetPriceVolatility(daysAhead,numDayStock, ndxtPrices)).colle
  -ndxtMomentumArray = sc.parallelize(GetMomentum(daysAhead,numDayStock, ndxtPrices)).collec
  -Predict(company,daysAhead,numDayStock,ndxtVolatilityArray,ndxtMomentumArray)
- output: NA
- exception: NA

7 MIS of Data Input Module (M??)

7.1 Module

Input Module

7.2 Uses

Main Module (M??)

7.3 Syntax

7.3.1 Exported Access Programs

Name	In	Out	Exceptions
ReadCSV	string	record list	IOError

7.4 Semantics

7.4.1 State Variables

NA

7.4.2 Environment Variables

dataSet : record list

7.4.3 Access Routine Semantics

ReadCSV():

- transition: NA
- output: A record list of the date and price
- exception:
  -IOError: Invalid file name and path. Invalid column name and data format in the file.

8 MIS of Price Volatility Module (M??)

8.1 Module

Volatility Module

8.2 Uses

Data Input Module (M??)
Main Module (M??)
RDD Module (M??)

8.3 Syntax

8.3.1 Exported Access Programs

Name	In	Out	Exceptions
GetPriceVolatility	$\mathbb{R}^n$,$\mathbb{N}^n$	$\mathbb{R}^n$	NotFittedError

8.4 Semantics

8.4.1 State Variables

NA

8.4.2 Access Routine Semantics

GetPriceVolatility():

- transition: NA
- output:
  volatilityArray: A list of real numbers represents the price volatility
- exception:
  NotFittedError : Improper price such as negative number.

9 MIS of Price Momentum Module (M??)

9.1 Module

Momentum Module

9.2 Uses

Data Input Module (M??)
Main Module (M??)
RDD Module (M??)

9.3 Syntax

9.3.1 Exported Access Programs

Name	In	Out	Exceptions
GetPriceMomentum	$\mathbb{R}^n$,$\mathbb{N}^n$	$\mathbb{R}^n$	NotFittedError

9.4 Semantics

9.4.1 State Variables

NA

9.4.2 Access Routine Semantics

GetPriceMomentum():

- transition: NA
- output:
  momentumArray : A list of real number for the price momentum
- exception:
  NotFittedError : Improper price such as negative number.

10 MIS of Prediction Module (M??)

10.1 Module

Predict Module

10.2 Uses

Data Input Module (M??)
Main Module (M??)
Volatility Module(M??)
Momentum Module(M??)
Kernelling Module(M??)
RDD Module(M??)

10.3 Syntax

10.3.1 Exported Access Programs

Name	**In**	**Out**	**Exceptions**
Predict	string, $\mathbb{R}^n$,$\mathbb{N}^n$	$\mathbb{R}$	-

10.4 Semantics

10.4.1 State Variables

NA

10.4.2 Environment Variables

Predict()

volatilityArray: A list of real numbers for the price volatility list calculated from the Price Volatility Model

momentumArray: A list of real numbers for the price momentum list calculated from the Price Momentum Model

featureX: A list of array which consists four real number elements : price volatility, price momentum, index volatility and index momentum

featureY: A list of integers (1 or -1)

Kernel mode: kernel='rbf

10.4.3 Access Routine Semantics

Predict():

- transition: NA
- output:
  score :A real number for the percentage of the possibility
- exception: NA

11 MIS of Plot Module (M??)

11.1 Module

Plot Module

11.2 Uses

Data Input Module (M??)
Main Module (M??)

11.3 Syntax

11.3.1 Exported Access Programs

Name	In	Out	Exceptions
Plot	char, $\mathbb{R}^n$	-	-

11.4 Semantics

11.4.1 State Variables

Dates: The array of the dates of each record

Prices: The array of the prices of each record

11.5 Semantics

11.5.1 Enviroment Variables

Dates: The array of the dates of each record

Prices: The array of the real number for the prices of each record

11.5.2 Access Routine Semantics

Plot():

- transition: NA
- output: NA
- exception: NA

11.6 Reference

Modeling high-frequency limit order book dynamics with support vector machines PDF 2013
Predicting Stock Price Direction using Support Vector Machines PDF 2015