# Graduation Paper Collection

wugouzi

March 6, 2021

## Contents



# 1 Mining stock price using fuzzy rough set system

## 1.1 Preliminaries

To store data, we have to define our fuzzy relational model. A fuzzy relational schema $R(A_1, \dots, A_n, u_r)$ is made up by a relation name $R$ and a list of attributes $A_1, \dots, A_n, u_r$. Each attribute $A_i$ is the name of a role played by some domain, $\text{dom}(A_i)$, and $u_r$ is characterized by the following membership function:

$$u_r : \text{dom}(A_1) \times \cdots \times \text{dom}(A_n) \to [0, 1]$$

**Definition 1.1.** Let $R(A_1, \dots, A_n, u_r(t))$ be a fuzzy relational schema. An $n$-ary fuzzy relation $r$ over $R$ is a set of $\text{dom}(A_1) \times \cdots \times \text{dom}(A_n)$, $t[A_i]$ refers to the value in $t$ for attribute $A_i$, and each tuple $t \in \text{dom}(A_1) \times \cdots \times \text{dom}(A_n)$. That is

$$\begin{aligned} r =& \{(t, u_r(t)) \mid t = ((t[A_1], u_1(t[A_1])), \dots, (t[A_n], u_n(t[A_n]))), \\ & \quad \text{and for } i = 1, \dots, n \\ & \quad t[A_i] \in \text{dom}(A_i), u_i(t[A_i]) \in [0, 1], \\ & \quad u_r(t) = \min(u_1(t[A_1]), \dots, u_n(t[A_n]))\} \end{aligned}$$

In this paper, we use a fuzzy relation EQUAL(EQ) to compute the membership degree of an object w.r.t. a given criteria, $\varphi(A)$

**Definition 1.2.** A fuzzy relation EQUAL(EQ) over dom$(A_j)$ is characterized by the membership function $u_{\mathrm{EQ}}$, where $u_{\mathrm{EQ}}$ satisfies the following conditions

$$\text{For all } x, y \in \mathrm{dom}(A_j), u_{\mathrm{EQ}}(x, x) = 1 \text{ and } u_{\mathrm{EQ}}(x, y) = u_{\mathrm{EQ}}(y, x)$$

According to Zadeh's possibility theory, $u_{\mathrm{EQ}}(x, y)$ can be interpreted as the possibility of treating the two values $(x, u(x))$ and $(u, u(y))$ equally under the same fuzzy term.

Since an object may partially match a certain property in many situations, we define a prespecified threshold value, $\alpha_j$, and let $u_r(t_i) \geq \alpha_j$. When $u_r(t_i) < \alpha_j$, we assume that the tuple $t_i$ doesn't satisfy to the given property

For example, let $\varphi(A) = y$. Then a fuzzy relation EQUAL(EQ) over an attribute domain, dom$(A_j)$, w.r.t. $y$ can be defined as

$$u_{\mathrm{EQ}}(x, y) = \\ \bigtimes \begin{cases} 0 & \text{for } 1 - \left|u_j(x) - u_j(y)\right| < \alpha_j \\ 1 - \left|u_j(x) - u_j(y)\right| < \alpha_j & \text{for } 1 - \left|u_j(x) - u_j(y)\right| \geq \alpha_j \end{cases}$$

**Definition 1.3.** Let the criteria $\varphi(A) = \varphi(A_i)$ and $\varphi(A_i) = y_i$. Then, a fuzzy set $X$ w.r.t. the criteria $\varphi(A)$ in a fuzzy relation $r$ can be defined as

$$X = \{(t, u_r(t)) \mid t \in r \text{ and } u_r(t) = u_{\mathrm{EQ}}(t[A_i], y_i)\}$$

Let $X = \{x_1, \dots, x_n\}$ be a set of objects. Denote $P(X)$ the set of all crisp subsets from $X$ and by $F(X)$ the set of all fuzzy subsets from $X$. Any subset $A \subset X$ will be called a **concept** in $X$. $u_A$ is the membership function of $A$. Let $C = \{C_1, \dots, C_n\}$, $C_i \in P(X)$ be a family of concepts in $X$ that forms a partition of $X$ and let $Y$ be a crisp subset of $X$.Lower and upper approximation of $Y$

$$\underline{C}(Y) = \bigcup_{C_i \subset Y} C_i, \quad \overline{C}(Y) = \bigcup_{C_i \cap Y \neq \emptyset} C_i$$

The pair $(\underline{C}(Y), \overline{C}(Y))$ is called a **rough set**.

Consider a situation where concepts $C_1, \dots, C_n \in F(X)$ from a weak fuzzy partition and concept $Y$ is a fuzzy set on $X$.

Then the lower and the upper approximation of $Y$ by means of $C$ are defined as fuzzy sets of $X/C$ with membership functions

$$u_{\underline{C}(Y)}(C_i) = \inf_x \max\{1 - u_{ci}(x), u_Y(X)\}$$
$$u_{\overline{C}(Y)}(C_i) = \sup_x \min\{u_{ci}(x), u_Y(x)\}$$

where $u_{\underline{C}(Y)}(C_i)$ is the degree of certain membership of $C_i$ in $Y$ and $u_{\overline{C}(Y)}(C_i)$ is the corresponding degree of possible membership. The pair $(u_{\underline{C}(Y)}(C_i), u_{\overline{C}(Y)}(C_i))$ is called a **fuzzy rough set**

The stock price at a specific time can be influenced by numerous factors. For example

| Date | Time | Stock price | $u_{price}(t[price])$ |
|---|---|---|---|
| 20010205 | 09:00 | 35.81 | 1 |
| 20010205 | 10:00 | 35.20 | 0.9662 |
| 20010205 | 11:00 | 34.29 | 0.9169 |
| 20010205 | 12:00 | 33.48 | 0.8741 |
| 20010205 | 13:00 | 35.20 | 0.9662 |
| 20010205 | 13:30 | 35.01 | 0.9777 |

| Date | Time | Stock price | $u_{price}(t[price])$ |
|---|---|---|---|
| 20000815 | 09:00 | 86.91 | 1 |
| 20000815 | 10:00 | 85.23 | 0.9617 |
| 20000815 | 11:00 | 83.02 | 0.9125 |
| 20000815 | 12:00 | 80.92 | 0.8669 |
| 20000815 | 13:00 | 82.47 | 0.9004 |
| 20000815 | 13:30 | 82.15 | 0.9452 |

The value $u_{price}(t[price])$ is the degree of the stock price at a specfic trading hour on that day, and the membership function of $u_{price}$ is given as

$$u_{price}(x) = (x/y)^2$$

where $x$ is the stock price at a specific trading hour on one day, and $y$ is the highest stock price during trading hours on the same day

Particularly, the change rate at an hour interval of the stock price may vary little each day. Since different times present different specific behavior, we employ the roughness technique to pre-classify stock price data into different groups by time and consider only the behavior of stock price of trading hours on weekdays. Each group has the same values for the attribute time.

| | Time | $u_{price}(t[price])$ |
|---|---|---|
| Group 1 | 09:00 | 0.9864 |
| | ⋮ | ⋮ |
| | 09:00 | 1 |
| Group 2 | 10:00 | 0.9327 |
| | ⋮ | ⋮ |
| | 10:00 | 0.9665 |
| ⋮ | ⋮ | ⋮ |
| Group 3 | 14:00 | 0.9887 |
| | 14:00 | 0.9678 |

For concentrating the change rate at hour intervals of the stock price, we also used roughness technique and assume the change rate of stock price is in following ranks

| |
|---|
| Rank 0: 6~7% |
| Rank 1: 5~6% |
| Rank 2: 4~5% |
| Rank 3: 3~4% |
| Rank 4: 2~3% |
| Rank 5: 1~2% |
| Rank 6: 0~1% |
| Rank 7: -1~0% |
| Rank 8: -2~-1% |
| Rank 9: -3~-2% |
| Rank 10: -4~-3% |
| Rank 11: -5~-4% |
| Rank 12: -6~-5% |
| Rank 13: -7~-6% |

A **linguistic summary** is a linguistically quantified proposition containing metaknowledge about a set of particular objects.

The general form of a linguistically quantified proposition is usually written as:

$$\mathbf{Q}\,\mathbf{X} \text{ are } \mathbf{F}$$

where **Q** is a (fuzzy) linguistic quantifier, **X** is a set of objects, and **F** is a property

To a relational database, let '$\mathbf{Q}\{t_1, \dots, t_n\}$ are **F**' be a linguistically quantified proposition, $t_1, \dots, t_n$ be a set of tuples in a fuzzy relation $r$. Then according to the semantics of the property, the truth value of '$\mathbf{Q}\{t_1, \dots, t_n\}$ are **F**' over the fuzzy relation $r$ can be computed by the following definition

**Definition 1.4.** Let $\{t_1, \dots, t_n\}$ be a set of tuples in a group over a relation $r$. Then

$$\text{Truth}(\mathbf{Q}\{t_1, \dots, t_n\} \text{ are } \mathbf{F}) = u_Q \left( \frac{1}{n} \left( \sum_{i=1}^{n} u_r(t_i) \right) \right)$$

where $u_t(t_i)$ is a degree membership value of $t_i$ w.r.t. **F**. When $\mathbf{F} = F_1 \vee \dots \vee F_m$, $u_r(t_i) = \max_{j=1}^{m}(u_{F_j}(t_i))$ and when $\mathbf{F} = F_1 \wedge \dots \wedge F_m$, $u_r(t_i) = \min_{j=1}^{m}(u_{F_j}(t_i))$.

## 1.2 A fuzzy rough set system

### 1.2.1 Mining agent using fuzzy rough set method

In the system, the corresponding domain knowledge is represented as

$$\{\text{Table}.A_1, \dots, \text{Table}.A_k\} \approx \mathbf{B}$$

where $\approx$ stands for 'relate to', and for $i = 1, \dots, k$ attribute Table.$A_i$ relates to knowledge **B**. For example, when we want to predict the ranks of stock price at a specifc time, the corresponding domain knowledge w.r.t. the WSP relation in table can be represented as

$$\begin{aligned}&\{\text{WSP.Date,WSP.Time,WSP.Stock\_Price,}\\ &\text{WSP.}(t[\text{price}])\} \approx \mathbf{B}\end{aligned}$$