Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2013, 5(9):478-482

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

A novel multi-scale qualitative trend analysis algorithm

Dong Gao, Xin Xu and Xin Ma*

College of Information Science and Technology,Beijing University of Chemical Technology, Beijing, China

---

ABSTRACT

Qualitative trend analysis (QTA) is a data driven approach and has been widely used in data compression, data monitoring, fault diagnosis and etc. However, for the traditional qualitative trend analysis, the qualitative trend is usually extracted by a fixed window and thresholds used are always dependent on human experience. So the qualitative trend extracted cannot represent the real trend reasonably. A novel multi-scale qualitative trend analysis algorithm is proposed to extract the qualitative trends. The process data is fitted linearly in a sliding window. The window can be extended or reduced to determine a segment. After all the segments for the data are determined, a qualitative trend of the data has been extracted. And then the initial width of the sliding window is changed to repeat above work in different scales. All the trends in different scales are ranked according to the fit index (F.I.) in decreasing order. The segments of the first one are sent to trend identification. The segments are indentified as "Increasing", "Decreasing"," Steady" and the qualitative trend is obtained. The result of case study proved that the trend can be extracted and identified efficiently with high accuracy using the algorithm.

Keywords: Multi-scale; Qualitative trend analysis; Trend extraction; Sliding window

---

INTRODUCTION

In chemical process industry, the data usually contain a lot of information, such as the state of system, the response of system, hand operation and so on. The information is very important and useful for state monitoring, data compression, fault diagnosis. However, due to increasing complexity of modern chemical plants, the useful information is usually hidden in the huge data. Qualitative trend analysis (QTA) is a data driven technique, which works by extracting the trend, identifying the trend [1]. QTA has been widely used in process data analysis, fault diagnosis, data compression and so on.

The most important issue of qualitative trend analysis is trend extraction. Trend extraction is to fit a constant, a first-order, second-order or higher order polynomial to the data in a window (a segment). And then a primitive is assigned to the segment. J.T.Cheung et al. proposed an approach to represent the process trends using triangular segments [2]. M. Janusz et al. proposed seven primitives(A(0,0),B(+,+),C(+,0),D(+,-),E(-,+),F(-,0),G(-,-)) to describe the qualitative trend . Rengaswamy and Venkatasubramanian described this language of trend, making recognition of primitives as a way of classification through neural network [3]. B.R. Bakshi et al. combined triangular segments and wavelet [4]. H. Vedam et al. proposed B-Splines based trend extraction algorithm using a sliding window with fixed width [5]. S. Charbonnier et al. proposed a method using QTA for fault diagnosis, in which it extracts semi-qualitative temporal episodes on-line from any time series, uses three primitives to describe the segments [6]. In the method, some thresholds depend on human experience. Based on the interval halving algorithm, M.R. Maura et al. proposed a framework for on-line trend extraction and fault diagnosis [7]. C. Scali et al. proposed improved qualitative shape analysis technique for automatic detection of valve stiction in flow control loops [8].Chen et al. improved the interval halving algorithm by adding the identification of step signal [9].

For the algorithms above, most of them fitted the data to a polynomial in a fixed window and some thresholds depend on human experience, so the segments cannot represent the real trend well. In addition, the complexity of the algorithm is high.

In the article, a novel multi-scale qualitative trend analysis algorithm is proposed to extract and identify the qualitative trend. The algorithm fits the data in a sliding window with an initial width to a constant or first-order polynomial. According to the fitting result, the window is extended or reduced adaptively to determine a segment. After all the segments for the data are determined, the qualitative trend of the data has been extracted. And then a sliding window with new initial width is used to repeat above work in a different scale. Every initial width is equal to $L/2^n$ (n = 1, 2…$\log_2 L$-3). L is the length of data. All the trends are ranked according to the fit index (*F.I.*) in decreasing order. The segments of the first result are used for trend identification. The segments are indentified as "Increasing", "Decreasing", "Steady" and the qualitative trend is obtained. The case study presented shows that the algorithm can extract trend efficiently with high accuracy.

MULTI-SCALE QUALITATIVE TREND ANALYSIS

The multi-scale qualitative trend analysis includes three steps: (1) extracting trends using an adaptive window in different scales; (2) Ranking the trends according to *F.I.*; (3) Indentifying segments of the first trend as "Increasing", "Decreasing", and "Steady".

1. Multi-scale trend extraction

The trends are extracted as follows:

(1) The original data is y_1, y2,...y_L,the length of data is L. A window is used to store the data, whose width M is equal to $L/2^n$ (n = 1, 2…$\log_2 L$-3). The data $y_1,y_2,...y_M$ is added to the window.

(2) Fit the data in the current window to a constant or first-order polynomial:

$$y(t) = a * t + b \tag{1}$$

t is time. And then F-test is used to determine whether linear fitting is acceptable.

If it is acceptable, the window will be extended to add a new data from the original data until it is not acceptable. To (3).

Otherwise, 3 σ rule is used to determine whether the data in the window is steady. If the data is steady, the window will be extended to add a new data from the original data until it is not steady, a segment is determined. To (4). Otherwise, the window is reduced and the last data in the window is sent back to original data. The data in the window is fitted to formula (1) again until fitting result is acceptable or is steady. To (4).

(3) The data in the current is split into 2 parts to get a more accurate result. The data in current window is $y_1,y_2,\ldots y_w$ (w > M). y_1, y_2, y_i, $y_{i+1}\ldots y_w$ (1< i<w) are fitted as formula (1) separately. The sum of fit error is calculated. The splitting schedule (i = P) with min fit error is found. The second part of data ($y_{P+1},\ldots y_w$)is sent back to original data and the first part of data($y_1,y_2,\ldots y_P$) as a segment is to (4).

(4) A new segment is determined. The data in the window is cleared. The other original data is moved into the window. To (2).If all the original data has been fitted, a trend consisting of several segments has been extracted and stored into the result set S. The initial width of window is changed (n = n+1).To (2).If n > $\log_2 L$-3, trend extraction ends.

2. Ranking the trends according to *F.I.*

This step is to rank the trends in the result set S according to *F.I.* *F.I.*is an index which can represent how suitably using the qualitative trend to represent the real data in the algorithm.

For the trends (S_1, $S_2\ldots S_N$, N = $\log_2 L$-3) in the set S, fitting error is calculated between every trend and original data. Obviously, the miner the error is the same between trend and original data is. On the other hand, that always makes too many short segments and these segments cannot reveal the general trend of data well. So it is reasonable to make compromise between fitting error and revealing the general trend.

F.I. is calculated as formula (2):

$$F.I._i = \begin{cases} 1/m, Fittingerror_i \le averageerror \\ 0, Fittingerror_i > averageerror \end{cases} \quad (2)$$

i = 1, 2… N; N is number of trends in the set; *F.I.*$_i$ is the index of the trend S_i; m is the number of segments in trend S_i; Fittingerror$_i$ is the fitting error of the trend S_i; averageerror is the average error of the trends.

If the fitting error of trend S_i is greater than the average error, the fitting error is unacceptable and the *F.I.*$_i$ is set to 0; otherwise, *F.I.*$_i$ is equal to 1/m. It is reasonable that fewer segments should be used when the error is acceptable, which can reveal the general trend and the complexity is lower.

So the trends are ranked according to the *F.I.* in decreasing order, the first one is the most suitable and chosen as the result for trend identification.

3. Segments identification
A algorithm is used to identify segments in the trend as “Increasing”, “Decreasing”, “Steady”.

For a segment, if it is steady (the value of data have no change), it is identified as “Steady”. If the parameter “a” calculated by formula (1) is positive, it is identified as an “Increasing”, otherwise it is “Decreasing”. When all the segments have been identified, the qualitative trend has been obtained. The result can be applied in data compression, data monitoring, fault diagnosis and etc.

4. The whole flowchart of algorithm
The whole flowchart of algorithm is shown in Fig.1.

CASE STUDY
The algorithm is used to extract and identify qualitative trend from real data in chemical industry. The original data is shown in Fig.2.

The algorithm of multi-scale qualitative trend analysis is used to extract and identify trends from the original data. The result is shown in table1 in different scales.

As the result shown in table1, the fitting error decreases as the initial window width decreases. However, the num of segments increases at the same time. Too many segments can not reveal the general trend. On the other hand, too high fitting error may miss the detail of trend. So according to the rank strategy, the *F.I.* of 3th trend is the greater

Table 1 Result of qualitative trend analysis

NO.	Initial window width	Num of segments	Fitting error	*F.I.*
1	512	3	2995.31	0
2	256	4	2885.2	0
3	128	5	2035.45	0.2
4	64	6	2020.1	0.17
5	32	9	1930.64	0.11
6	16	18	1429.74	0.05
7	8	78	1155.45	0.013

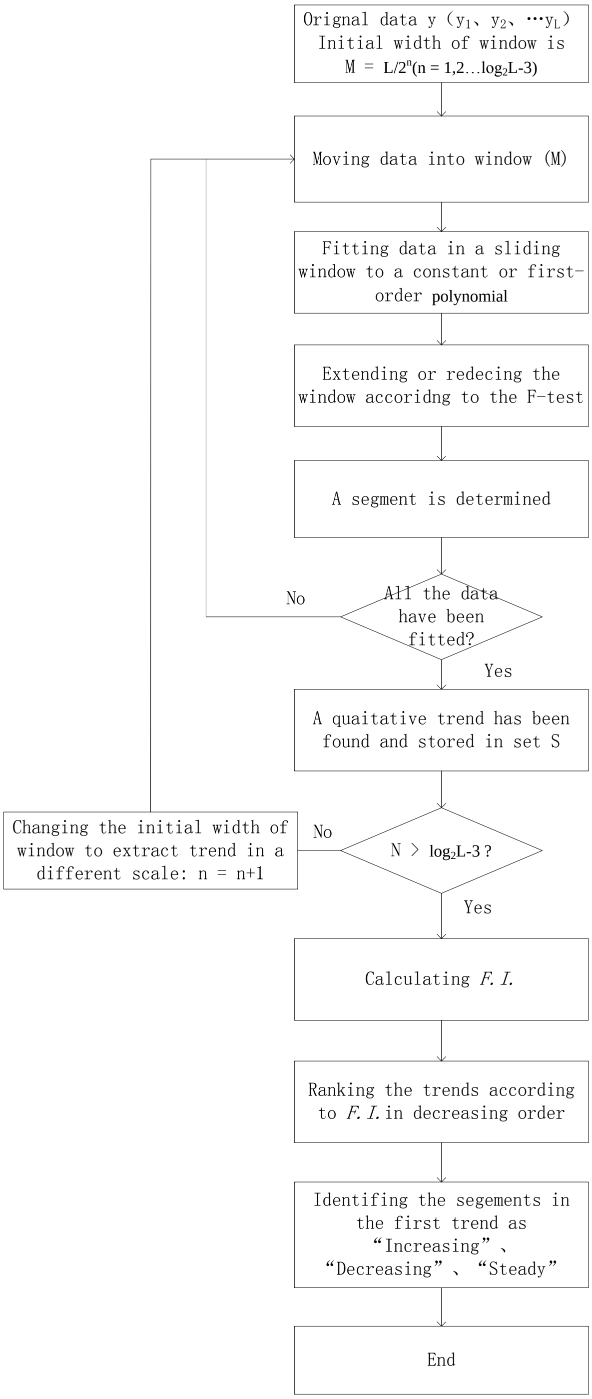

Fig 1: Flowchart of algorithm

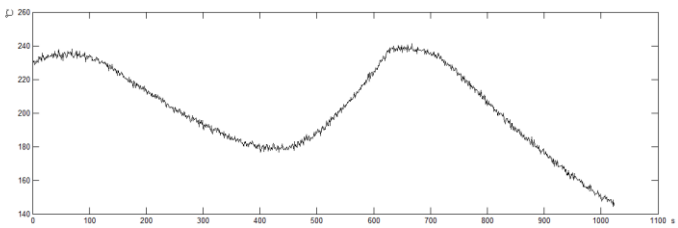

Fig 2: Original data

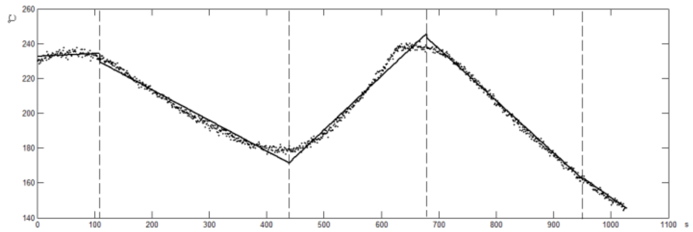

Fig 3: Qualitative trend and original data

than others, so the final qualitative trend is obtained. It is shown in Fig. 3.

As it shown in Fig.3, the qualitative trend extracted includes 5 segments (real line). They are: “Increasing”, “Decreasing”, “Increasing”, “Decreasing”, and “Decreasing”. The qualitative trend can represent the real trend well. If more detail is needed, the other trends in table 1 with more segments and lower fitting errors can be selected.

The result of case study shows that the trend can be extracted reasonably and accurately using the algorithm. It can be applied for data compression, data monitoring, fault diagnosis and etc.

CONCLUSION

In this article, a novel multi-scale qualitative trend analysis algorithm is proposed. The algorithm includes three steps: (1) extracting trends using a adaptive window in different scales; (2) Ranking the trends according to *F.I.*; (3) Indentifying segments of the first trend as “Increasing”, “Decreasing”, “Steady”.

Compared to traditional qualitative trend analysis, the trend extracted in the algorithm can represent trend more reasonably due to the sliding window in different scales. In addition, the complexity is lower due to only linear fitting is used. The case study showed that trend can be extracted and identified efficiently with high accuracy using the algorithm.

REFERENCES

[1] Janusz M., Venkatasubramanian V.,**1991**, *Eng. Appl. Artif. Intell*., 4,329-339.
[2] Cheung J.T., Stephanopoulos G., **1990**, *Comput. & Chem. Engng*.,14,495-510.
[3] Rengaswamy R., Venkatasubramanian V.,**1995**, *Eng. Appl. Artif. Intell*., 8,35-51.
[4] Bakshi B. R., Stephanopoulos G., **1994**, *Comput. & Chem. Engng.*,18,267-302.
[5] Vedam H., Venkatasubramanian V., Bhalodia, **1998**, *Comput. & Chem. Engng*., 22, S827-S830.
[6] Charbonnier S., Garcia-Beltan C., Cadet C., Gentil S., **2005**, *Engineering Applications of Artificial* Intelligence, 18, 21–36.
[7] Maurya R.M., Paritosh P.K., Rengaswamy R.,**2010**, *Engineering Application of Artificial intelligence*, 23, 950-960.
[8] Scali C., Ghelardoni C., **2008**, *Control Engineering Practice*, 16, 1501–1508.
[9] Chen J.P, Shui A., Song Z.H., Liu S.B., Zhou R., **2011**, *Instrumentation Technology*, 2, 57-60.