



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

АЛГОРИТМ УНИФИКАЦИИ ВРЕМЕННЫХ РЯДОВ

Сапунков А.А., Афанасьева Т.В.

*Ульяновский государственный технический университет,
г. Ульяновск, Российская федерация*

sapaslks@gmail.com

tv.afanasjeva@gmail.com

В работе приводится методика унификации временных рядов, то есть приведение временных рядов разной длины к временным рядам одинаковой длины для последующей обработки или анализа. Описывается наиболее популярный метод для решения данной задачи, метод F-преобразований, и приводится новая разработанная методика унификации временных рядов. В конце работы сравниваются обе методики, указываются преимущества и недостатки.

Ключевые слова: временной ряд, унификация временного ряда по длине, F-преобразование, обработка временного ряда.

Введение

Актуальность. Одной из важных проблем при анализе больших баз данных временных рядов (ВР) является обнаружение схожих по поведению ВР и их лингвистическое описание для упрощения последующего процесса моделирования. Данная проблема тесно связана с феноменами “Big Data” и “Internet of Thing”, при помощи которых предоставляются большие объемы данных с датчиков высокой частоты. Полученные ВР трудно оценить эксперту из-за их объема и количества в текстовом или графическом представлении. А для автоматизированных систем анализа, сравнения и оценки ВР необходимо приводить все ВР к единому виду. Специализированных методов для унификации ВР, т.е. для приведения ВР к общей длине и общему диапазону значений, на момент написания данной работы обнаружено не было. Обычно для выполнения данной задачи использовался метод F-преобразования [Перфильева 2003], поэтому с ним и будет сравниваться приведенный в данной статье алгоритм

Цель. Главной целью данной работы было разработать методику приведения ВР к единому формату для последующей обработки (анализа, сравнения, прогнозирования) ВР в автоматизированных системах. Под единым форматом подразумевается единый для всех ВР диапазон значений ВР и его длина.

Задача. Разработать алгоритм унификации ВР, т.е. алгоритм приведения ВР разных длин к ВР указанной длины путем сжатия или расширения ВР.

В первом разделе описывается применение алгоритма F-преобразований для сжатия или расширения ВР. В следующем разделе определяется новый метод унификации ВР, приводится пошаговый алгоритм преобразования заданного ВР в унифицированный и приводятся примеры работы алгоритма. В третьем разделе приводятся результаты обработки базы временных рядов при помощи F-преобразования и нового предложенного алгоритма, а также оценка эффективности нового алгоритма.

1. Описание F-преобразования

Методика F-преобразования успешно применяется для анализа ВР на основе моделей мягкого вычисления, к примеру [Novak 2010], [Perfilieva 2010]. Цель F-преобразования – предобработка ВР. Ниже приведены важные для F-преобразования свойства:

- удаление высокочастотных колебаний,
- сглаживание и разложение.

Далее мы коротко напомним основные положения F-преобразования, которые были представлены в [Перфильева 2003]. Данный метод уже был описан в ряде работ, поэтому детали опускаются.

Предположим, задан ВР X из n значений, где $x_t \in X$, $t=1,2,...,n$.

F-преобразование проходит в 2 этапа:

- Прямое F-преобразование.
- Обратное F-преобразование.

На первом этапе F-преобразования происходит разбиение заданного ВР X на $[1, n]$ отрезков с помощью m : $2 \leq m \leq n - 2$ нечетких множеств A_1, \dots, A_m : $[1, n] \rightarrow [0, 1]$, которые называются прямыми F-преобразованиями. Значение m получается путем деления длины ВР n на коэффициент преобразования c , который указывается при преобразовании, и округляется до целого. Далее компоненты вектора $Fm[X] = [F_1, \dots, F_m]$ рассчитываются следующим образом:

$$F_k = \frac{\sum_{t=1}^n x_t * A_k(t)}{\sum_{t=1}^n A_k(t)} \quad (1.1)$$

где $k = 1, 2, \dots, m$; A_1, \dots, A_m – это треугольные функции, определенные на отрезке $[1, n]$. согласно [19], для F-преобразования ВР необходимо определить значение m равноудаленных точек основной функции и рассчитать компоненты вектора $Fm[X] = [F_1, \dots, F_m]$ при помощи взвешенных средних.

Второй этап F-преобразования – обратный этап. На этом этапе аппроксимированное представление некой функции, определенной поведением временного ряда X предстает в новом виде:

$$f_t = \sum_{k=1}^m F_k * A_k(t) \quad (1.2)$$

Таким образом F-преобразование можно использовать для решения поставленной задачи: прямое F-преобразование для сжатия ВР, обратное – для расширения.

2. Приведение ВР к унифицированному виду

В данной методике приводится понятие унифицированного ВР. Предположим, задан ВР X из n значений, где $x_t, x_t \in X, t=1, 2, \dots, n$, тогда унифицированным ВР будем считать ВР Z из m значений где $z_p, z_p \in Z, z_p \in [0..K] p=1, 2, \dots, m$. Таким образом получаем, что привести ВР к унифицированному виду значить преобразовать все значения заданного ВР в диапазон от 0 до K и расширить/сократить сам ВР до m значений.

Получим ВР Y из n значений, где $y_t, y_t \in Y, y_t \in [0..K] t=1, 2, \dots, n$. Для этого преобразуем все значения ВР X по формуле (2.1):

$$y_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} * K, \quad (2.1)$$

где x_{\min} – минимальное значения ВР X , x_{\max} – максимальное значение ВР X .

Далее возможно две ситуации при приведении ВР к унифицированному ВР. Первая – это при n и m не равны, тогда необходимо провести сжатие или расширение ВР Y . Вторая ситуация при $n=m$, в таком случае считаем, что ВР Y приведен к унифицированному виду.

Рассмотрим два ВР Y и Z различных длин. Для расчета значений унифицированного ВР z_p необходимо определить между какими значениями ВР Y y_t и y_{t+1} оно находится, поэтому необходимо ввести соответствие между координатами ВР t и p :

$$t = \frac{p * n}{m} \text{ DIV } 1, \quad (2.2)$$

где $a \text{ DIV } b$ – операция целочисленного деления числа a на b .

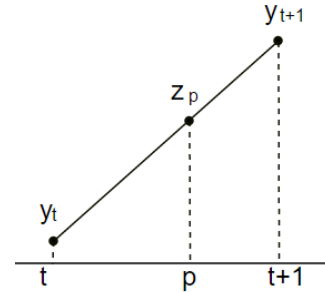


Рисунок 1 – определение значения z_p

В данном случае (2.2) применяется для получения целой части значения a . Далее определяем значение z_p , опираясь на подобие треугольников с общей гипотенузой (y_t, y_{t+1}) и противолежащим углом в 90 градусов (рис. 1). Так как точки p и t относятся к разным координатным, то нам необходимо найти p' – положение точки p в координатах ВР Y . Этом можно сделать по следующей формуле:

$$p' = \frac{p * n}{m}. \quad (2.3)$$

Таким образом взяв дробную часть от p' мы получим расстояние от точки t до p' или же коэффициент отношения треугольников. Теперь мы можем получить значения z_p :

$$z_p = (y_{t+1} - y_t) * (p' \text{ MOD } 1) + y_t \quad (2.4)$$

Где $a \text{ MOD } b$ – операция получения остатка деления числа a на b . Таким образом по формуле (2.4) можно получить ВР Z вне зависимости от того, сжимаем мы ВР или расширяем.

3. Пример обработки ВР методикой F-преобразований и алгоритмом унификации ВР

Рассмотрим ВР ts11 (ВР изображен на рисунке 2) длиной в 909 значений из базы данных, предложенной в рамках конкурса International Time Series Forecasting Competition [ITSFC]. В рамках эксперимента будем сжимать данный ВР до 100 значений и расширять до 2000 методом F-преобразований и алгоритмом унификации ВР, чтобы можно было их сравнить.

Для того чтобы сократить ВР ts11 при помощи прямого F-преобразования с коэффициентом сжатия $c=9$. В результате преобразования получаем ВР из 101 значения, поэтому для получения ВР нужной нам длины мы отсекаем лишнее последнее значение

(результат изображен на рисунке 3). Для расширения ВР необходимо применить обратное F-преобразование. Однако мы сталкиваемся с проблемой выбора коэффициента, если $c=2$, то ВР получится длиной 1818 (результат изображен на рисунке 4) т.е. нам будет не хватать 182 значений, если $c=3$, то получится ВР длиной 2727 значений, а отсекать 727 значений, которые составляют больше четверти временного ряда нельзя.

Далее преобразуем заданный ВР ts11 при помощи алгоритма унификации с параметрами $k=50$, $m=100$. В результате мы получаем, как и при применении F-преобразования, сглаженный ВР без всечастотных колебаний, однако его длина составляет 100 значений, поэтому нет необходимости отсекать какие-либо значения. Результат отображен на рисунке 5.

При преобразовании ВР ts11 при помощи алгоритма унификации с параметрами $k=50$, $m=2000$ у нас получился ВР длиной в 2000 значений изображенный на рисунке 6. Данный ВР получился не сглаженный и сохранил все высокочастотные колебания.

Выводы. При использовании метода F-преобразований для сжатия и расширения следует учитывать длину полученного ВР и требуемую длину, чтобы количество отсекаемых значений не влияло на основную тенденцию ВР.

Использование предложенного алгоритма не имеет недостатка с отсечением значений, подобно методу F-преобразований. При расширении ВР данный алгоритм может иметь небольшую погрешность, которая не повлияет ни на общую тенденцию ВР ни на высокочастотные шумы.

Если длина требуемого временного ряда делится нацело на длину заданного временного ряда, то результаты расширения ВР по обоим методикам будут идентичны.

При сжатии ВР результаты методик сильно отличаются друг от друга из-за того, что метод F-преобразований предназначен для сглаживания ВР, а предложенный метод берет соответствующие значения ВР без сглаживания.

4. Заключение

Выводы. При использовании метода F-преобразований для сжатия и расширения следует учитывать длину полученного ВР и требуемую длину, чтобы количество отсекаемых значений не влияло на основную тенденцию ВР.

Использование предложенного алгоритма не имеет недостатка с отсечением значений, подобно методу F-преобразований. При расширении ВР данный алгоритм может иметь небольшую погрешность, которая не повлияет ни на общую тенденцию ВР ни на высокочастотные шумы.

Если длина требуемого временного ряда делится нацело на длину заданного временного ряда, то

результаты расширения ВР по обоим методикам будут идентичны.

При сжатии ВР результаты методик сильно отличаются друг от друга из-за того, что метод F-преобразований предназначен для сглаживания ВР, а предложенный метод берет соответствующие значения ВР без сглаживания.

По результатам эксперимента можно сделать вывод, что новый алгоритм унификации лучше выполняет поставленную задачу, по сравнению с методом F-преобразований. Главное его преимущество заключается в том, что он может преобразовывать заданный ВР во ВР любой указанной длины.

Библиографический список

[Перфильева 2003] Перфильева И. Нечеткое преобразование: применительно к проблеме роста рифов. В кн: Демикко Р. Клир GJ и др. редакторы. Нечеткая логика в геологии. Амстердам: Академическая пресса: 2003, 275-300.

[Novak 2010] Novak, V., Stepnicka, M. Dvorak, A., Perfilieva, I. and Pavliska, V. Analysis of seasonal time series using fuzzy approach, International Journal of General Systems, 2010 39, pp.305-328.

[Perfilieva 2010] Perfilieva I., Yarushkina N., Afanaseva T. Relaxed Discrete F-Transform and its Application to the Time Series Analysis / // Da Ruanetal (Eds.): Computational Intelligence. Foundations and Applications (Proc.of the 9th Int. FLINS Conf.). P. 249 -255, World Scientific, Emei, Chengdu, China, 2-4 August, 2010.

[ITSFC] International Time Series Forecasting Competition [Электронный ресурс]. URL: <http://irafm.osu.cz/cif/main.php> (дата обращения: 16.03.2015).

TIME SERIES REDUCTION ALGORITHM TO THE SPECIFIED FORMAT

Sapunkov A.A. Afanasjeva T.V.

Ulyanovsk State Technical University

sapaslks@gmail.com

tv.afanasjeva@gmail.com

The paper describes a method of unification of the time series, in other words, bringing the different length of time series to a time series of the same length for later processing or analysis. It describes the most popular method for solving this problem, the method of F-transformation, and describes the new developed method. At the end of the two methods are compared, indicate advantages and disadvantages.

Key words: time series, the unification of the length of time series, F-transform, processing time series.

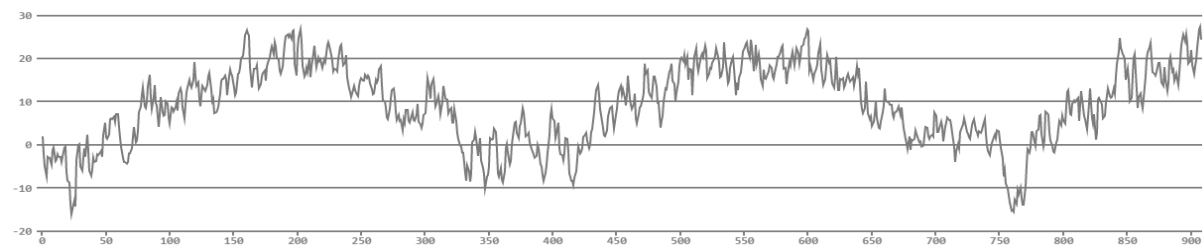


Рисунок 2 – BP ts11

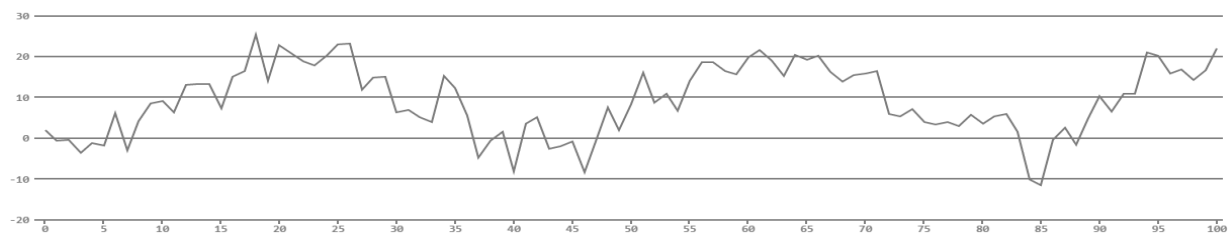


Рисунок 3 – BP ts11 сжатый методом F-преобразований

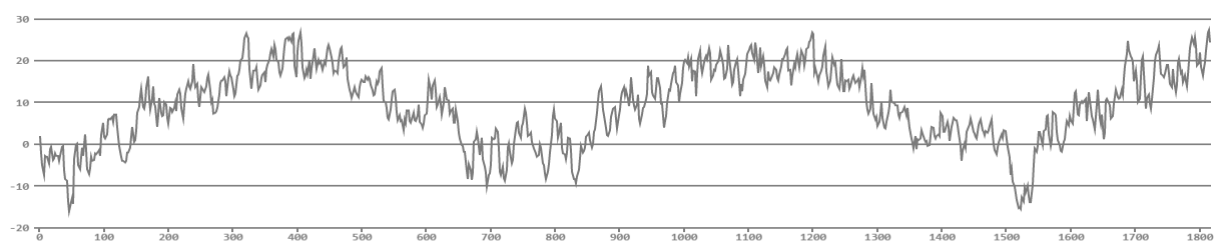


Рисунок 4 – BP ts11 расширенный методом F-преобразований

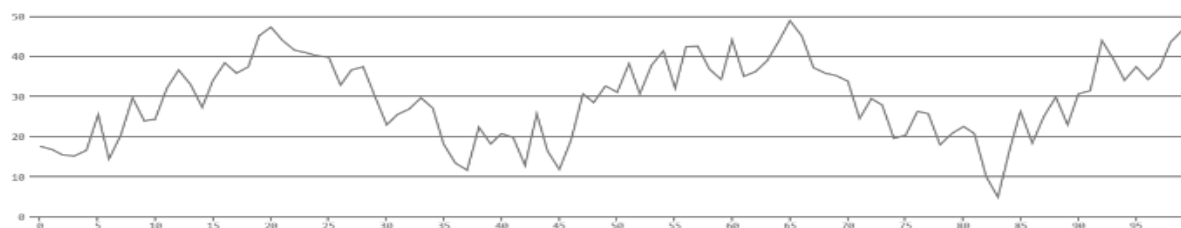


Рисунок 5 – BP ts11 сжатый алгоритмом унификации BP

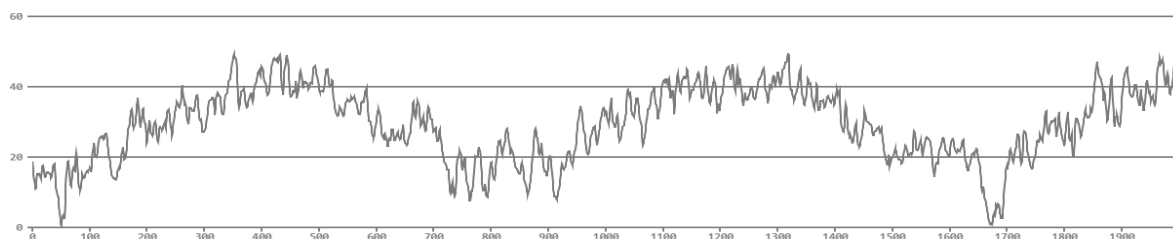


Рисунок 6 – BP ts11 расширенный алгоритмом унификации BP