

**Библиографические ссылки**

1. Вольдек А. И. Индукционные магнитогидро-динамические машины с жидкометаллическим рабочим телом. Л. : Энергия. Ленингр. отд-ние, 1970.  
 2. Курейчик В. М. Интеллектуальные системы. М. : Физматлит, 2005.

3. Буль О. Б. Методы расчета магнитных систем электрических аппаратов. М. : Академия, 2006.  
 4. Батищев Д. И. Генетические алгоритмы решения экстремальных задач : учеб. пособие / под ред. Я. Е. Львовича. Воронеж, 1995.

Таблица 2

**Решение, найденное ГА и уточненное локальным алгоритмом**

Наименование параметра (независимой переменной) оптимизации	Обозначение в программной системе	Решение ГА	Решение, улучшенное локальным поиском	Известное решение
Сила тока в первой катушке, А	I1	23,5	25	25
Сила тока во второй катушке, А	I2	25	25	25
Сила тока в третьей катушке, А	I3	23,5	25	25
Фаза во второй катушке, град	fi2	94	97	105
Фаза в третьей катушке, град	fi3	124	146	165
Частота питающего напряжения, Гц	f	21	20	20
Значение целевой функции $W(X)$	–	0,001 952 29	0,001 585 8	0,001 606 8

S. S. Bezhitskiy, E. A. Golovenko, V. A. Goremykin, M. V. Pervuhin

**SOLVING PROBLEM OF OPTIMAL CHOICE OF PLAN LINEAR INDUCTION MACHINES POWER SUPPLY CHARACTERISTICS USING GENETIC ALGORITHM WITH LOCAL SEARCH**

*The approach for choice of optimal variant of power supply characteristics of plan linear induction machines (LIM) is described. Optimization of LIM structure is based on usage of evolution algorithm for global optimization with local search for solution improvement. The useful optimization function features for optimal choice of a plan linear induction machines structure.*

*Keywords: parametric optimization, computational modeling, traveling magnetic field, linear induction machines.*

© Бежитский С. С., Головенко Е. А., Горемыкин В. А., Первухин М. В., 2010

УДК 004.045

А. А. Еремеевский, В. Х. Ханов

**МЕТОД ПОИСКА ЗАДАННОГО ШАБЛОНА ВО ВРЕМЕННОМ РЯДУ\***

*Рассматривается проблема поиска заданного шаблона в существующем временном ряду. Проводится сравнение скорости разработанной модели поиска при различных длинах временного ряда и шаблонов поиска.*

*Ключевые слова: обработка временных рядов, поиск шаблона во временном ряду, индекс поиска.*

Обработка временных рядов в настоящее время является актуальным инструментом анализа информации, накапливаемой в процессе контроля и управления технологическими процессами, мониторинга окружающей среды, исследования экономических показателей.

Одной из наиболее распространенных задач является поиск заданного шаблона в существующем временном

ряду. Суть этой задачи сводится к тому, что имеется шаблон в виде временного ряда фиксированной длины, а в исходном ряду необходимо найти фрагменты, соответствующие шаблону с заданной точностью.

Для прямолинейного решения данной задачи необходимо прежде всего установить окно выборки из исходного ряда, равное длине шаблона, а затем установить окно

\*Работа выполнена в рамках реализации ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 гг.», ПСП № 1032 от 27 мая 2010 г.

на начало временного ряда и, смещая на один элемент, выбирать из него фрагменты, после чего каждый фрагмент следует сравнить с шаблоном. Очевидно, что такое решение весьма трудоемко и требует значительного времени. Поэтому необходимо найти способы оптимизации данного процесса.

Оптимизация этой задачи впервые была предложена группой исследователей университета штата Мэриленд (США) [1]. Они разработали так называемый *ST*-индекс. В качестве меры, определяющей точность соответствия шаблона и фрагмента ряда, взято евклидово расстояние. Этот выбор основан на том, что дискретное преобразование Фурье (ДПФ) не изменяет расстояния между рядами. Это следует из теоремы Парсевалля:

$$\sum_{i=0}^{n-1} |x_i|^2 = \sum_{F=0}^{n-1} |x_F|^2,$$

где  $x_i$  – исходный ряд;  $x_F$  – дискретное преобразование Фурье. А уже из свойства линейности ДПФ вытекает, что сохраняется и евклидово расстояние между рядами:

$$D(\vec{x}, \vec{y}) = D(\vec{X}, \vec{Y}),$$

где  $D(\vec{x}, \vec{y})$  – евклидово расстояние между шаблоном и фрагментом ряда;  $D(\vec{X}, \vec{Y})$  – евклидово расстояние между ДПФ шаблона и фрагмента ряда.

Разработчики *ST*-индекса опирались на ряд работ по исследованию спектра временных рядов различных процессов, в том числе поведения основных биржевых показателей. Как известно, наибольшей энергией обладают несколько первых гармоник ряда Фурье. Таким образом, если рассчитать евклидово расстояние на основе только первых  $n$  коэффициентов ДПФ, то оно будет мало отличаться от действительного расстояния между рядами:

$$D(S, Q) = \left( \sum_{i=1}^l (S[i] - Q[i])^2 \right)^{\frac{1}{2}} \geq \left( \sum_{i=1}^n (S[i] - Q[i])^2 \right)^{\frac{1}{2}}, n < l,$$

где  $D(S, Q)$  – евклидово расстояние между ДПФ шаблона и фрагмента ряда;  $l$  – длина шаблона и фрагмента ряда;  $S$  – ДПФ шаблона;  $Q$  – ДПФ фрагмента ряда.

Таким образом, сохранив критерий точности соответствия шаблона и фрагмента ряда и вычисляя евклидово расстояние по первым  $n$  коэффициентам фурье-образов шаблона и фрагмента, мы гарантированно не пропустим искомый фрагмент ряда. Конечно, существует возможность ошибочной выборки такого фрагмента, полное расстояние между которым и шаблоном будет больше установленного. Для окончательного ответа необходимо вычислить полное евклидово расстояние между отобранными фрагментами и шаблоном.

Для снижения вычислительной сложности применяют быстрое дискретное преобразование Фурье (БДПФ), при этом длина шаблона должна быть кратна  $2^n$ , где  $n$  – любое целое положительное число больше 1.

При практическом использовании *ST*-индекса возникает несколько проблем:

– *ST*-индекс зависит от длины шаблона, в результате для каждого поиска необходимо заново перестраивать

индекс. При этом необходимое количество преобразований Фурье для временного ряда выражается формулой:

$$N = L - l + 1,$$

где  $L$  – длина временного ряда;  $l$  – длина шаблона. Следовательно, при достаточно большой длине временного ряда необходимое количество преобразований Фурье также будет возрастать, что делает процесс поиска более ресурсозатратным;

– количество операций для каждого БДПФ вычисляется по формуле [2]:

$$N * (\log N),$$

где  $N$  – количество значений исходного ряда. Таким образом, количество операций каждого БДПФ линейно зависит от количества значений для преобразования. Как следствие, при возрастающей длине шаблона поиска будет возрастать и количество операций для решения поставленной задачи;

– при возрастании длины шаблона возрастает и количество гармоник для каждого БДПФ [2]:

$$n = 2^m, N = 1 + 2^{m-1},$$

где  $n$  – исходная длина ряда;  $N$  – количество гармоник БДПФ. Как было сказано выше, для определения точного соответствия шаблона участку временного ряда необходимо вычислить полное евклидово расстояние, при этом трудоемкость данного вычисления напрямую зависит от количества элементов шаблона, так как при увеличивающейся длине шаблона увеличивается и количество гармоник БДПФ. Поэтому при увеличении длины шаблона трудоемкость вычисления полного евклидова расстояния также возрастает.

На основании вышеизложенного встает вопрос о создании постоянного индекса временного ряда, не зависящего от длины шаблона поиска. Данную задачу удалось решить путем дальнейшей оптимизации *ST*-индекса.

Временной ряд  $X_N$  изначально разбивается на  $K = N/4$  частей и для каждой части создается постоянный индекс  $Q_K$  из трех значений гармоники БДПФ. При изменении временного ряда индексация производится только для вновь добавленных значений, что позволяет не перестраивать индекс, но при этом всегда поддерживать его актуальность.

Сам шаблон поиска также разбивается на  $k = n/4$  частей и для каждой части производится БДПФ с образованием индексов  $S_k$ . Далее производится вычисление евклидова расстояния для каждого индекса  $S_k$ :

$$\left( \sum_{i=1}^3 (S_k[i] - Q_k[i])^2 \right)^{\frac{1}{2}},$$

$$k = 1, 2, \dots, n/4, K = 1, 2, \dots, N/4.$$

Совпадение индекса  $S_k$  с индексом  $Q_K$  (нулевое евклидово расстояние) означает совпадение части шаблона с частью временного ряда. Очевидно, что дальнейшее сравнение можно производить, опираясь на точки, где это совпадение произошло. Таким образом, если для индекса  $S_k$  найдено  $L$  соответствий, то для индекса  $S_{k+1}$  (следующий фрагмент шаблона поиска) будет вычислено только  $L$  евклидовых расстояний с индексом временного ряда  $Q_{k+1}$  (следующий фрагмент временного ряда). При этом

если индекс  $S_{k+1}$  не совпал с индексом  $Q_{k+1}$ , то весь шаблон начиная от точки  $K$  является несовпадающим с заданным фрагментом, благодаря чему каждый последующий шаг отсеивает значительную часть вычислений. Количество же необходимых преобразований Фурье зависит только от длины шаблона и вычисляется по формуле

$$N = n/4.$$

При этом обеспечивается требуемая точность при любой длине шаблона, так как индексы содержат не больше трех гармоник и для них всегда вычисляется полное евклидово расстояние.

Для оценки эффективности исследования было произведено сравнение скорости разработанной авторами модели поиска с  $ST$ -индексом при различных длинах временного ряда и шаблонов поиска (рис. 1, 2). Это сравнение позволяет сделать вывод, что разработанный метод поиска заданного шаблона во временном ряду эффективен при возрастающей длине шаблона и временного ряда. При этом длина шаблона практически не оказывает влияния на скорость поиска.

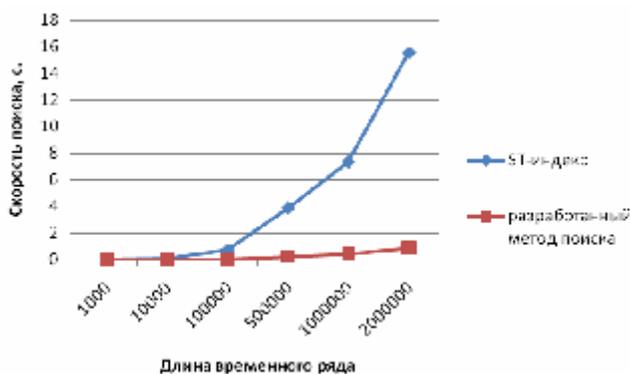


Рис. 1. Сравнение скорости поиска  $ST$ -индекса с индексом разработанного метода при шаблоне постоянной длины в зависимости от длины временного ряда

Рассматриваемый метод был апробирован при первичном анализе телеметрической информации, принимаемой с российских космических аппаратов (КА) на зарубежные наземные командно-измерительные системы.

Отечественные КА имеют формат телеметрических кадров, существенно отличающийся от международных стандартов телеметрии. Вследствие этого зарубежная аппаратура не может в режиме реального времени осуществлять покадровую синхронизацию поступающей с КА телеметрии, выделять и обрабатывать кадры и передавать очищенные данные на последующие этапы обработки. Обычно зарубежная аппаратура при приеме телеметрии с отечественных КА осуществляет полную запись всего потока информации на магнитные носители и последующую обработку программными методами. При подобной постобработке важно создать механизм быстрого доступа к требуемым кадрам телеметрии. Метод поиска заданного шаблона во временном ряду позволил создать упорядоченный индекс для быстрого перехода к нужным фрагментам принятой телеметрии.

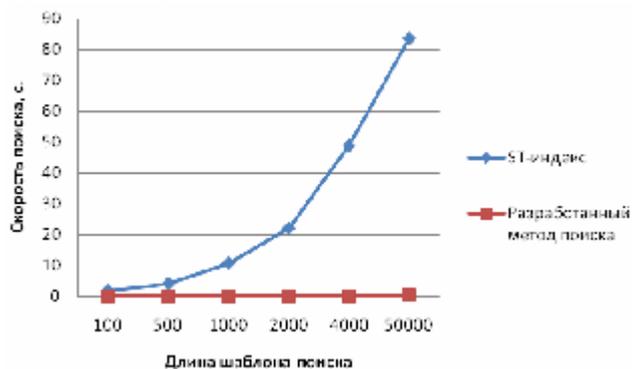


Рис. 2. Сравнение скорости поиска  $ST$ -индекса с индексом разработанного метода при постоянной длине временного ряда в зависимости от длины шаблона

### Библиографические ссылки

1. Faloutsos C., Ranganathan M., Manolopoulos Y. Fast Subsequence Matching in Time-Series Databases // Proc. of the ACM SIGMOD Conf. on Management of Data. Minneapolis, Minn., 1994. P. 419–429.
2. Сергиенко А. Б. Цифровая обработка сигналов. СПб. : Питер, 2002.

A. A. Eremeevski, V. Kh. Khanov

## METHOD OF SEARCH OF THE SET TEMPLATE IN TIME SERIES

The problem of search of the set template in existing time series is considered. Comparison of speed of the developed model of search is made at various lengths time series and search templates.

Keywords: processing of time series, template search in time series, search index.

© Еремеевский А. А., Ханов В. Х., 2010