Анализ временных рядов и выявление процессов с размытой периодичностью

**Человек:** Предметом исследования является методика оценки шумовой компоненты во временных рядах и ее удаление, выделение тренда и колебаний c различными периодами, вводится понятие Т-ε и Т-h-ε почти периодов для конечных рядов. В основу анализа положено требование гладкости функции, представляющей исходные данные и имеющей производные до четвертого порядка включительно и выделение почти периодов на основе функций типа Альтера – Джонсона. Отдельно выделяется тренд длины периодов, выявленных в данных ряда колебаний. Алгоритм решения задачи основан на минимизации отклонений рассчитываемых значений от гладкой функции при условии соответствия отклонений от исходных данных уровню шума. Для выявления колебательной составляющей и тренда почти периодов используется модифицированная функция Альтера – Джонсона. Предлагаемая методика и алгоритмы оценки и устранения шума в данных позволяют обоснованно определить уровень шума в данных, удалить из данных шумовую компоненту, выявить почти периоды в данных в смысле введенных в статье определений, выделить в данных трендовую и колебательную составляющие, выявить, при необходимости, тренд изменения почти периодов.

**Key words:** шум, фильтрация шума, временной ряд, тренд, почти период, периодические функции, спектр сигнала, декомпозиция данных, численное моделирование, анализ временного ряда

=================================

**FastText\_KMeans\_Clean:** В нашем случае рассматривается почти периодичность на конечном интервале, поэтому далее будем рассматривать ее в несколько ином смысле. Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов. В частности выделенная шумовая компонента не является белым шумом: прослеживаются короткопериодические колебания и меняется амплитуда колебаний.

**Key words part:** 0.7142857142857143

=================================

**FastText\_KMeans\_Raw/:** Статистическая информация, данные, отражающие результаты экспериментальных исследований в различных областях можно описать в виде временных рядов или последовательностей данных. В нашем случае рассматривается почти периодичность на конечном интервале, поэтому далее будем рассматривать ее в несколько ином смысле. Для функции Джонсона (T-ε ) характерно выделение кратных периодов с постепенным ростом величины ε , для ее модификации (T- h-ε, h= kT ) характерно выделение кратных периодов менее выражено и наблюдается уменьшение величины ε . При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов.

**Key words part:** 0.6785714285714286

=================================

**FastText\_PageRank\_Clean/:** Здесь будем придерживаться следующего определения. В первую очередь речь идет о быстром преобразовании. В последнем случае . Под нормой удобнее всего использовать просто . · Выделение короткопериодических колебаний. 5. Результаты численного моделирования. Соответствующие данные представлены на рисунках ниже. · удалить из данных шумовую компоненту;.

**Key words part:** 0.4642857142857143

=================================

**FastText\_PageRank\_Raw/:** Здесь будем придерживаться следующего определения. В первую очередь речь идет о быстром преобразовании. В последнем случае . · Выделение короткопериодических колебаний. 5. Результаты численного моделирования. Соответствующие данные представлены на рисунках ниже. · удалить из данных шумовую компоненту;. · найти T -ε и Т- h -ε почти периоды;.

**Key words part:** 0.5

=================================

**Mixed\_ML\_TR/:** Функцию f( x) будем называть периодической на интервале [a, b ] с периодом T>0, если для любого x, x+ T [a, b ] f( x)= f( x+ T). В определенном отношении понятие периодичности можно обобщить, введя почти периодические функции. Функцию f( x) будем называть T- h-ε почти периодической с периодом T>0 и константами T> h> 0, ε> 0, если для любого x, существует такое, что |f( x)- f( x+ T+ τ)|< ε. Рассмотрим некоторые методы, позволяющие выделить подобные периодические составляющие в эмпирических данных. Для T- h-ε почти периодической функции можно в качестве аналога функции Альтера – Джонсона использовать ее модификацию, преобразующую исходный временной ряд в спектр периодов φ h( T). Кроме того, преобразование Фурье дает, вообще говоря, посторонние максимумы для несинусоидальных колебаний, а дискретное быстрое преобразования Фурье небезразлично к некратности исследуемого интервала значений функции ее периоду (или почти периоду). которое справедливо для любой четырежды дифференцируемой функции. Удаление короткопериодических колебаний с периодом T можно проводить на основе трактовки короткопериодических колебаний, как шума. Таким образом, предлагаемая методика и алгоритмы оценки и устранения шума в данных, выявления T -ε и Т- h -ε почти периодов, позволяют:.

**Key words part:** 0.75

=================================

**MultiLingual\_KMeans/:** В определенном отношении понятие периодичности можно обобщить, введя почти периодические функции. Функцию f( x) будем называть T- h-ε почти периодической с периодом T>0 и константами T> h> 0, ε> 0, если для любого x, существует такое, что |f( x)- f( x+ T+ τ)|< ε. Рассмотрим некоторые методы, позволяющие выделить подобные периодические составляющие в эмпирических данных. Кроме того, преобразование Фурье дает, вообще говоря, посторонние максимумы для несинусоидальных колебаний, а дискретное быстрое преобразования Фурье небезразлично к некратности исследуемого интервала значений функции ее периоду (или почти периоду). которое справедливо для любой четырежды дифференцируемой функции. Удаление короткопериодических колебаний с периодом T можно проводить на основе трактовки короткопериодических колебаний, как шума. Таким образом, предлагаемая методика и алгоритмы оценки и устранения шума в данных, выявления T -ε и Т- h -ε почти периодов, позволяют:.

**Key words part:** 0.5714285714285714

=================================

**Multilingual\_PageRank/:** Наиболее популярны методы, основанные на преобразовании Фурье, для которого разработано большое количество программных средств [4]. которое справедливо для любой четырежды дифференцируемой функции. В первом случае коэффициенты a, b находятся методом наименьших квадратов. При целых ( T-1)/2 последний член в скобках исчезает. 5. Результаты численного моделирования. В качестве примера использования предлагаемой методики рассмотрим просчеты на данных о количестве солнечных пятен по годам (числа Вольфа) с 1700 по 2016 год [12]. 11-летние почти периодические колебания существенно изменяются по амплитуде. · удалить из данных шумовую компоненту;.

**Key words part:** 0.5357142857142857

=================================

**RuBERT\_KMeans\_Without\_ST/:** Функцию f( x) будем называть T-ε почти периодической на интервале [a, b ] с периодом T >0 и константой ε>0 , если для любого x, |f( x)- f( x+ T)|< ε. Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. Альтернативой является дискретное преобразование Фурье с шагом по выбранному диапазону периодов, однако последнее значительно медленнее и, кроме того, неточно для короткопериодических (высокочастотных) колебаний. График колебаний периода рассматриваемой выше последовательности приведен на рисунки 3. Возникает задача устранения, по возможности, такого шума. При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов. Поскольку при устранении шума функцию f( x) можно считать дифференцируемой, возможно численное нахождение ее производной с приемлемой точностью. Таким образом, предлагаемая методика и алгоритмы оценки и устранения шума в данных, выявления T -ε и Т- h -ε почти периодов, позволяют:.

**Key words part:** 0.8214285714285714

=================================

**RuBERT\_KMeans\_With\_ST/:** Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. График колебаний периода рассматриваемой выше последовательности приведен на рисунки 3. Значения, сопоставляемые элементам получающегося ряда, содержат и ошибки измерения и, в общем случае, подвержены случайным внешним воздействиям. При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов. Последняя задача о нахождении коэффициентов b, k сводится к нахождению линейной зависимости методом наименьших квадратов, после чего аналогично находится коэффициентов a, что позволяет легко выделить экспоненциальный тренд.

**Key words part:** 0.6428571428571429

=================================

**RUBERT\_page\_rank\_Without\_ST/:** Под нормой удобнее всего использовать просто . Первое, очевидно, точнее выявляет высокочастотные (короткопериодические) колебания, второе - низкочастотные (длиннопериодические) колебания. 3. Анализ временных рядов и выявление процессов с размытой периодичностью. Полученные данные являются, вообще говоря, предметом для дальнейшего анализа. · проводить независимый анализ выделенных трендовых данных и зависимостей, носящих колебательный характер, и выявлять в них аналитические и дифференциальные зависимости.

**Key words part:** 0.5714285714285714

=================================

**RUBERT\_page\_rank\_With\_ST/:** Для этого ведем предварительно несколько определений. В первую очередь речь идет о быстром преобразовании. В последнем случае . Под нормой удобнее всего использовать просто . Соответствующие данные представлены на рисунках ниже.

**Key words part:** 0.3928571428571429

=================================

**RUSBERT\_KMeans\_Without\_ST/:** В определенном отношении понятие периодичности можно обобщить, введя почти периодические функции. Здесь будем придерживаться следующего определения. Для T- h-ε почти периодической функции можно в качестве аналога функции Альтера – Джонсона использовать ее модификацию, преобразующую исходный временной ряд в спектр периодов φ h( T). При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов. Величина шума в точке i определяется из.

**Key words part:** 0.8214285714285714

=================================

**RUSBERT\_KMeans\_With\_ST/:** В нашем случае речь может идти только о дискретном преобразовании на ограниченном интервале. В частности для введенной выше T-ε почти периодической функции можно задать функцию Альтера – Джонсона [5,6], преобразующую исходный временной ряд в спектр периодов φ( T). 4. Алгоритм выявления шума, почти периодических процессов и долговременного тренда. При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов. Значения функции при этом определяются из. При целых ( T-1)/2 последний член в скобках исчезает. · выявить, при необходимости, тренд изменения почти периодов;.

**Key words part:** 0.8214285714285714

=================================

**RUSBERT\_page\_rank\_Without\_ST/:** Здесь будем придерживаться следующего определения. Под нормой удобнее всего использовать просто . При таком подходе можно использовать следующую схему:. Соответствующие данные представлены на рисунках ниже. Полученные данные являются, вообще говоря, предметом для дальнейшего анализа.

**Key words part:** 0.42857142857142855

=================================

**RUSBERT\_page\_rank\_With\_ST/:** Здесь будем придерживаться следующего определения. Анализ и обработка зашумленных данных вызывает значительные трудности. При таком подходе можно использовать следующую схему:. 5. Результаты численного моделирования. Соответствующие данные представлены на рисунках ниже.

**Key words part:** 0.5

=================================

**Simple\_PageRank/:** Альтернативой является дискретное преобразование Фурье с шагом по выбранному диапазону периодов, однако последнее значительно медленнее и, кроме того, неточно для короткопериодических (высокочастотных) колебаний. В частности для введенной выше T-ε почти периодической функции можно задать функцию Альтера – Джонсона [5,6], преобразующую исходный временной ряд в спектр периодов φ( T). Для T- h-ε почти периодической функции можно в качестве аналога функции Альтера – Джонсона использовать ее модификацию, преобразующую исходный временной ряд в спектр периодов φ h( T). Кроме того, преобразование Фурье дает, вообще говоря, посторонние максимумы для несинусоидальных колебаний, а дискретное быстрое преобразования Фурье небезразлично к некратности исследуемого интервала значений функции ее периоду (или почти периоду). Для слабо осциллирующих функций, то есть таких, для которых период осцилляции значительно больше шага исследуемого временного, выражение в правой части равенства (4) пренебрежимо мало. На рисунке 5 представлены соответственно шум, 11-летних почти периодические колебания, длинные (94-106 лет) почти периодические колебания и долговременный тренд.

**Key words part:** 0.75

=================================

**TextRank/:** Функцию f( x) будем называть периодической на интервале [a, b ] с периодом T>0, если для любого x, x+ T [a, b ] f( x)= f( x+ T). Функцию f( x) будем называть T-ε почти периодической на интервале [a, b ] с периодом T >0 и константой ε>0 , если для любого x, |f( x)- f( x+ T)|< ε. Функцию f( x) будем называть T- h-ε почти периодической с периодом T>0 и константами T> h> 0, ε> 0, если для любого x, существует такое, что |f( x)- f( x+ T+ τ)|< ε. Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. В частности для введенной выше T-ε почти периодической функции можно задать функцию Альтера – Джонсона [5,6], преобразующую исходный временной ряд в спектр периодов φ( T). Для T- h-ε почти периодической функции можно в качестве аналога функции Альтера – Джонсона использовать ее модификацию, преобразующую исходный временной ряд в спектр периодов φ h( T).

**Key words part:** 0.6785714285714286

=================================

**TF-IDF\_KMeans/:** Функцию f( x) будем называть T-ε почти периодической на интервале [a, b ] с периодом T >0 и константой ε>0 , если для любого x, |f( x)- f( x+ T)|< ε. Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. В этих спектрах точкам максимумов соответствуют периоды колебаний анализируемых процессов. В последнем случае . 3. Анализ временных рядов и выявление процессов с размытой периодичностью. Возникает задача устранения, по возможности, такого шума. Удаление короткопериодических колебаний с периодом T можно проводить на основе трактовки короткопериодических колебаний, как шума. · выделить в данных трендовую и колебательную составляющие;.

**Key words part:** 0.7857142857142857

=================================

**Текст:** Статистическая информация, данные, отражающие результаты экспериментальных исследований в различных областях можно описать в виде временных рядов или последовательностей данных. Во многих случаях подобные ряды в той или иной мере содержат повторяющиеся последовательности данных. Наличие таких повторов говорит о присутствии циклических участков в этих процессах. В то же время нельзя говорить о наличии в них строгой периодической составляющей.. Обычно почти периодические функции понимают в смысле определения X. Бора [1, 2, 3].. Функция f( x) , непрерывная в интервале (-∞,∞), называется почти периодической функцией, если для любого ε>0 существует относительно плотное множество U ε - почти периодов этой функции. То есть, почти периодическая функция, если для каждого ε>0 существует такое L= L(ε) , что в каждом интервале длины L найдется хотя бы одно число τ , для которого .. 2. Почти периодические функции, их определение и выделение почти периодов. В нашем случае рассматривается почти периодичность на конечном интервале, поэтому далее будем рассматривать ее в несколько ином смысле.. Для этого ведем предварительно несколько определений.. Функцию f( x) будем называть периодической на интервале [a, b ] с периодом T>0, если для любого x, x+ T [a, b ] f( x)= f( x+ T).. В определенном отношении понятие периодичности можно обобщить, введя почти периодические функции. Здесь будем придерживаться следующего определения.. Функцию f( x) будем называть T-ε почти периодической на интервале [a, b ] с периодом T >0 и константой ε>0 , если для любого x, |f( x)- f( x+ T)|< ε.. Введем теперь еще одно понятие, описывающее наличие цикличности в поведении функций.. Функцию f( x) будем называть T- h-ε почти периодической с периодом T>0 и константами T> h> 0, ε> 0, если для любого x, существует такое, что |f( x)- f( x+ T+ τ)|< ε.. Рассмотрим некоторые методы, позволяющие выделить подобные периодические составляющие в эмпирических данных.. Наиболее популярны методы, основанные на преобразовании Фурье, для которого разработано большое количество программных средств [4]. В нашем случае речь может идти только о дискретном преобразовании на ограниченном интервале. В первую очередь речь идет о быстром преобразовании. Недостатком такого преобразования (шаг идет по частотам) является его неточность для длиннопериодических колебаний, то есть таких, что период колебаний имеет порядок интервала, на котором задаются данные. Альтернативой является дискретное преобразование Фурье с шагом по выбранному диапазону периодов, однако последнее значительно медленнее и, кроме того, неточно для короткопериодических (высокочастотных) колебаний. В этих спектрах точкам максимумов соответствуют периоды колебаний анализируемых процессов.. Другую группу методов представляют методы прямого анализа периодичности, не предполагающего синусоидальности самих колебаний.. В частности для введенной выше T-ε почти периодической функции можно задать функцию Альтера – Джонсона [5,6], преобразующую исходный временной ряд в спектр периодов φ( T). (1). В ряде случаев удобнее использовать нормированный спектр в виде. (2). В последнем случае .. Под нормой удобнее всего использовать просто .. Для T- h-ε почти периодической функции можно в качестве аналога функции Альтера – Джонсона использовать ее модификацию, преобразующую исходный временной ряд в спектр периодов φ h( T). . или. (3). В подобных спектрах минимумам соответствуют периоды колебаний анализируемых процессов.. На рисунках 1 и 2 представлены результаты выявления колебательных характеристик функции . Функция задана таблицей с шагом 0,1.. График на рисунке 1 иллюстрирует результаты анализа спектра функции при постоянном шаге по частотам (быстрое дискретное преобразование Фурье) и при постоянном шаге по периодам. Первое, очевидно, точнее выявляет высокочастотные (короткопериодические) колебания, второе - низкочастотные (длиннопериодические) колебания. Кроме того, преобразование Фурье дает, вообще говоря, посторонние максимумы для несинусоидальных колебаний, а дискретное быстрое преобразования Фурье небезразлично к некратности исследуемого интервала значений функции ее периоду (или почти периоду).. . Рисунок 1. . На рисунке представлены значения функции Альтера – Джонсона и ее модификации, значение h , для которой задается величиной h= T/20 .. . Рисунок 2. . Для функции Джонсона (T-ε ) характерно выделение кратных периодов с постепенным ростом величины ε , для ее модификации (T- h-ε, h= kT ) характерно выделение кратных периодов менее выражено и наблюдается уменьшение величины ε . В общем случае модификация T- h-ε плохо выделяет длинные периоды при наличии более коротких. Последнее показывает границы применимости различных схем выделения периодической составляющей во временных рядах.. При использовании схемы T- h-ε в процессе расчета для i точки вычисляются величины li , определяющие значения, «подозреваемые в качестве периода». Последовательность значений li , точнее ее тренд, показывают зависимость (или отсутствие зависимости) величины периода от времени.. График колебаний периода рассматриваемой выше последовательности приведен на рисунки 3.. . Рисунок 3. . . 3. Анализ временных рядов и выявление процессов с размытой периодичностью. Рассмотрим некоторые проблемы анализа временных рядов. Элементами таких рядов являются пары, задающие момент наступления события (значение аргумента) и соответствующий ему результат (значение функции). Последовательность событий может измеряться как с постоянным, так и с переменным шагом. Значения, сопоставляемые элементам получающегося ряда, содержат и ошибки измерения и, в общем случае, подвержены случайным внешним воздействиям. В дальнейшем такого рода ошибки измерений и результаты внешних воздействий будем трактовать, как шум.. Анализ и обработка зашумленных данных вызывает значительные трудности. Возникает задача устранения, по возможности, такого шума. Для его устранения используются различные методы сглаживания, такие как, методы скользящей средней, экспоненциального сглаживания и др. [7,8,9,10]. Другой подход может быть реализован на основе представления исходных данных, как суммы гладкой функции, представляющей анализируемый процесс и шума. Требование гладкости при этом оказывается достаточно сильным, что позволяет не только оценить уровень шума, но и выделить такую гладкую функцию [11].. 4. Алгоритм выявления шума, почти периодических процессов и долговременного тренда. При анализе данных, представленных временным рядом важно выделить такие характеристики описываемого им процесса, как апериодический тренд и набор сопровождающих процесс колебательных подпроцессов.. При таком подходе можно использовать следующую схему:. · Удаление шума.. · Выделение короткопериодических колебаний.. · Удаление короткопериодических колебаний на основе их рассмотрения как шума.. · Выделение колебаний с более длинным периодом с последовательным их удалением до получения апериодической составляющей.. . Удаление шума осуществляется согласно [11], исходя из соотношения. , (4). которое справедливо для любой четырежды дифференцируемой функции.. Для слабо осциллирующих функций, то есть таких, для которых период осцилляции значительно больше шага исследуемого временного, выражение в правой части равенства (4) пренебрежимо мало. В этих условиях. (5). среднеквадратичное значение шума σ находится в соответствии с (4, 5) и составит. (6). Значения функции при этом определяются из. (7). Величину минимального почти периода в смысле введенных выше определений будем определять исходя из соотношений (2,3). Для повышения точности определения почти периодов целесообразно удалить из данных заведомо апериодическую часть. В качестве такой части (тренда) наиболее естественно принять либо функцию , либо функцию . В первом случае коэффициенты a, b находятся методом наименьших квадратов. Во втором для нахождения коэффициентов можно использовать производную выделенной после удаления шума функции f( x). Поскольку при устранении шума функцию f( x) можно считать дифференцируемой, возможно численное нахождение ее производной с приемлемой точностью. Если f'( x) на всем интервале значений сохраняет знак, то гипотеза о наличии экспоненциального тренда допустима. Исходя из того, что , получаем, что в полулогарифмических координатах , где α= 1, если f'( x)>0, либо α= -1 в противном случае. Последняя задача о нахождении коэффициентов b, k сводится к нахождению линейной зависимости методом наименьших квадратов, после чего аналогично находится коэффициентов a, что позволяет легко выделить экспоненциальный тренд.. Удаление короткопериодических колебаний с периодом T можно проводить на основе трактовки короткопериодических колебаний, как шума. Величина шума в точке i определяется из. . При целых ( T-1)/2 последний член в скобках исчезает.. Таким образом, задача об устранении коротких колебаний полностью сводится к задаче (7).. Последние шаги алгоритма повторяются до устранения периодических составляющих.. В результате применения алгоритма исходный временной ряд приводится к сумме вида. (8). Здесь yi, xi, si – значения исходного ряда, аргумента и шума соответственно; - значения апериодической части тренда и его очередных колебательных составляющих.. Такое представление позволяет провести содержательный анализ процесса и выявить составляющие его подпроцессы.. 5. Результаты численного моделирования. В качестве примера использования предлагаемой методики рассмотрим просчеты на данных о количестве солнечных пятен по годам (числа Вольфа) с 1700 по 2016 год [12]. Соответствующие данные представлены на рисунках ниже.. . Рисунок 4. . На рисунке 4 представлены соответственно исходные данные, данные с удалением шума, трендовые данные с удалением 11-летних циклов, долговременный тренд с удалением длинных циклов (94-106 лет).. . . Рисунок 5. . На рисунке 5 представлены соответственно шум, 11-летних почти периодические колебания, длинные (94-106 лет) почти периодические колебания и долговременный тренд.. Полученные данные являются, вообще говоря, предметом для дальнейшего анализа. В частности выделенная шумовая компонента не является белым шумом: прослеживаются короткопериодические колебания и меняется амплитуда колебаний. 11-летние почти периодические колебания существенно изменяются по амплитуде. Периоды длинных колебаний значительно меняются во времени. Долговременный тренд, скорее всего, является фрагментов низкочастотных колебаний, период которых не определяется из-за малости самого временного интервала, на котором проводились наблюдения.. . 6. Выводы. Таким образом, предлагаемая методика и алгоритмы оценки и устранения шума в данных, выявления T -ε и Т- h -ε почти периодов, позволяют:. · обоснованно определить уровень шума в данных;. · удалить из данных шумовую компоненту;. · найти T -ε и Т- h -ε почти периоды;. · выявить, при необходимости, тренд изменения почти периодов;. · выделить в данных трендовую и колебательную составляющие;. · проводить независимый анализ выделенных трендовых данных и зависимостей, носящих колебательный характер, и выявлять в них аналитические и дифференциальные зависимости..