

УДК 621.379.54; 621.372.54.037.372

Рогоза В. С., д.т.н.; Сергеев-Горчинский А. А., асп.

(Нац. технический университет Украины «КПИ». +380 (68) 810 04 28. alexey.horchynskyi@yahoo.com)

ВЫБОР МЕРЫ ПОДОБИЯ ЦИФРОВЫХ СИГНАЛОВ ДЛЯ СТОХАСТИЧЕСКОГО РАСЧЕТА ОПТИМАЛЬНЫХ ПАРАМЕТРОВ ЦИФРОВОГО ФИЛЬТРА

Рогоза В. С., Сергеев-Горчинский О. О. **Вибір міри подібності цифрових сигналів для стохастичного розрахунку оптимальних параметрів цифрового фільтру.** Досліджуються питання розрахунку оптимального порядку фільтру нижніх частот, який призначений для виділення низькочастотної складової стаціонарного сигналу, представленого набором дискретних зашумлених значень. Для порівняння сигналів були вибрані міри: коефіцієнт кореляції Пірсона, відстань Евкліда, косинусна міра, середньоквадратична помилка. В процесі досліджень була створена програмна бібліотека для генерування стаціонарних сигналів, їх модифікування і згладжування. Були виконані експерименти по генеруванню сигналів з різними характеристиками. Знайдені значення міри подібності для початкового і згладженого сигналів. В результаті була вибрана міра, яка дозволяє підвищити вірогідність стохастичного розрахунку оптимального порядку фільтрації стаціонарних сигналів. Стаття завершується висновками по результатах, отриманих при порівнянні міри подібності цифрових сигналів.

Ключові слова: цифровий сигнал, порівняння, міра, цифровий фільтр, подібність, коефіцієнт кореляції Пірсона, відстань Евкліда, косинусна міра, середньоквадратична помилка

Рогоза В. С., Сергеев-Горчинский А. А. **Выбор меры подобия цифровых сигналов для стохастического расчета оптимальных параметров цифрового фильтра.** Исследуются вопросы расчета оптимального порядка фильтра нижних частот, который предназначен для выделения низкочастотной составляющей стационарного сигнала, представленного набором дискретных зашумленных значений. Для сравнения сигналов были выбраны меры: коэффициент корреляции Пирсона, расстояние Евклида, косинусная мера, среднеквадратичная ошибка. В процессе исследований была создана программная библиотека для генерирования стационарных сигналов, их модифицирования и сглаживания. Были выполнены эксперименты по генерированию сигналов с различными характеристиками. Найденны значения мер подобия для исходного и сглаженного сигналов. В результате была выбрана мера, которая позволяет повысить вероятность стохастического расчета оптимального порядка фильтрации стационарных сигналов. Статья завершается выводами по результатам, полученным при сравнении мер подобия цифровых сигналов.

Ключевые слова: цифровой сигнал, сравнение, мера, цифровой фильтр, подобие, коэффициент корреляции Пирсона, расстояние Евклида, косинусная мера, среднеквадратичная ошибка

Rohoza V. S., Serheyev-Horchyns'kyu O. O. **Selection of similarity measure of the digital signals for a stochastic calculation of the digital filter optimum parameters.** The matters of calculating the optimal order of the low-pass filter designed to isolate the low-frequency component of the non-time-varying signal represented by a set of noisy signal discrete values are studied. To compare the signals such measures as Pearson correlation factor, Euclidean distance, cosine measure and mean-squared error were selected. During the study it was created the software library for generating the non-time-varying signals, their modification and smoothing. The experiments were performed to generate signals with different characteristics. The similarity measure values for the source and smoothed signals were found. In consequence it was chosen the measure that allows to increase the certainty of stochastic calculation of the optimal order of non-time-varying signals filtering. The article ends with the conclusions using the results obtained by comparison of the digital signal similarity measures.

Keywords: digital signal, comparison, measure, digital filter, similarity, Pearson correlation factor, Euclidean distance, cosine measure, mean-squared error

1. Введение. В различных технических задачах требуется сравнение данных, которые в общем случае могут иметь различное представление. Одной из моделей представления данных является представление в виде временных рядов. В цифровой обработке сигналов при проектировании адаптивных фильтров, осуществляющих подстройку по образцовым сигналам, требуется расчет меры подобия временных рядов, составленных из дискретных значений сигналов [1].

В статье [2] расчет параметров фильтра осуществлялся при сравнении сглаженного и аппроксимированного временных рядов. Для оценки подобия временных рядов было использовано значение среднеквадратичной ошибки (СКО). Следует отметить, что в прикладных методах анализа данных используются и другие численные показатели подобия временных рядов, которые называются «мерами подобия» [3] либо «мерами близости» [4].

Чтобы проверить целесообразность использования СКО для оценки оптимальности фильтрации [2] была создана программная библиотека для цифровой обработки сигналов.

Для оценки подобия были сгенерированы исходные сигналы различной формы и их модифицированные копии, полученные путем добавления к исходным сигналам случайных составляющих (шума) и последующего их сглаживания.

Информация о модулях, вошедших в разработанную программную библиотеку, приведена в разделе «Создание тестовой среды».

В разделе «Выбор мер подобия» описаны выбранные математические отношения, которые определяют меру подобия двух временных рядов.

В разделе «Сравнение мер подобия» описаны эксперименты, которые были проведены с целью выбора меры, позволяющей повысить вероятность стохастического расчета оптимального порядка фильтрации стационарных сигналов.

В разделе «Выводы» авторы обобщают результаты, полученные в разделе «Сравнение мер подобия».

2. Создание тестовой среды. В процессе исследований была создана библиотека программ для генерирования цифровых сигналов с различными характеристиками и обработки сгенерированных сигналов. Структура библиотеки изображена на Рис. 1.



Рис. 1. Структура библиотеки для цифровой обработки сигналов

Функции основных модулей библиотеки программ следующие. «Модуль генерирования сигнала» предназначен для генерирования стационарных сигналов и включает генерирование сигнала и шумовой составляющей.

«Модуль генерирования сигнала по функции» реализует программное генерирование сигналов следующих форм: синусоидального, прямоугольного, пилообразного, треугольного. Программная реализация (программный метод) предоставляет возможность настройки амплитуды, частоты и фазы сигнала.

«Модуль генерирования шумовой составляющей» добавлен для приближения формы сигнала к форме возникающей в реальных (естественных) условиях передачи. В модуле реализован метод генерирования случайных чисел с нормальным распределением и возможностью задания значений математического ожидания и среднеквадратичного отклонения.

Для минимизации шума был создан «Модуль нерекурсивной фильтрации сигнала». В модуле реализован программный метод фильтрации простым скользящим средним (ПСС), который используют для фильтрации шума с нормальным распределением [5]. В фильтре ПСС обработанное значение $y(t)$ в момент времени t определяется соотношением [2]:

$$y(t) = \frac{1}{n} \cdot \sum_{i=1}^n x(t-i), \quad (1)$$

где n – порядок фильтрации (число предыдущих исходных значений сигнала, принятых для расчета выходного сигнала в текущей временной точке); $x(t-i)$ – значение исходного сигнала в момент (точка) времени дискретизации $t-i$; $y(t)$ – сигнал на выходе фильтра в текущий момент времени.

В компоненте «Модуль оценки фильтрации» были реализованы программные методы оценки подобия сигналов (меры подобия): коэффициент корреляции Пирсона, расстояние Евклида, косинусная мера, среднеквадратичная ошибка. Перечисленные методы были объединены в «Модуль оценки подобия сигналов». Для дополнительной оценки значений числовых последовательностей был реализован «Модуль статистической оценки».

Созданная программная библиотека была использована для формирования зашумленных сигналов и их последующей цифровой обработки. Цель выполненных экспериментов была направлена на определение меры подобия сигналов, которая позволяла бы повысить вероятность стохастического расчета оптимального порядка фильтрации стационарных сигналов.

3. Выбор мер подобия. Как было отмечено во введении, при решении технических задач часто возникает необходимость сравнения временных рядов с целью определения степени их близости. Такая близость может быть оценена на основе использования тех или иных мер подобия анализируемых сигналов. Например, к задачам, при решении которых используются упомянутые меры близости относятся задачи адаптивной фильтрации, классификации по образцу, обучения по образцу и некоторые другие.

В публикациях, касающихся обсуждаемой в данной статье тематики, предложены различные математические соотношения для расчета меры подобия числовых последовательностей, которые можно разделить на следующие группы: меры расстояния, угловые меры, корреляционные меры.

Для каждой из трех упомянутых групп мер существует множество вариантов расчета, из которых авторами были выбраны следующие [6, 7]:

- коэффициент корреляции Пирсона (Pearson Correlation Coefficient);
- расстояние Евклида (Euclidean Distance);
- косинусная мера (Cosine Similarity);
- среднеквадратичная ошибка (Mean Square Error).

Математические соотношения для перечисленных мер приведены в Табл. 1. Все из перечисленных выше мер подобия были реализованы программно.

Расчет математического ожидания для последовательностей значений $x(t)$, $y(t)$ может быть выполнен с использованием следующих соотношений:

$$\bar{X} = \frac{1}{n} \cdot \sum_{t=1}^n x(t), \quad (2)$$

$$\bar{Y} = \frac{1}{n} \cdot \sum_{t=1}^n y(t), \quad (3)$$

где t – момент (точка) времени дискретизации;

n – общее количество точек дискретизации сигналов X , Y .

С помощью разработанной программной библиотеки был сгенерирован тестовый зашумленный синусоидальный сигнал. В Табл. 2 приведены характеристики тестового сигнала.

Математические соотношения для расчета мер подобия двух числовых последовательностей (временных рядов)

Табл. 1

Коэффициент корреляции Пирсона	$C(X, Y) = \frac{\sum_{t=1}^n (x(t) - \bar{X}) \cdot (y(t) - \bar{Y})}{\sqrt{\sum_{t=1}^n (x(t) - \bar{X})^2} \cdot \sqrt{\sum_{t=1}^n (y(t) - \bar{Y})^2}}$	(4)
Расстояние Евклида	$D(X, Y) = \sqrt{\sum_{t=1}^n (x(t) - y(t))^2}$	(5)
Косинусная мера	$S(X, Y) = \frac{\sum_{t=1}^n x(t) \cdot y(t)}{\sqrt{\sum_{t=1}^n x(t)^2} \cdot \sqrt{\sum_{t=1}^n y(t)^2}}$	(6)
Среднеквадратичная ошибка	$E(X, Y) = \frac{1}{n} \cdot \sum_{t=1}^n (x(t) - y(t))^2$	(7)

Характеристики тестового сигнала

Табл. 2

Форма сигнала	синусоида
Амплитуда сигнала	3
Частота сигнала	0.1
Распределение шумовой составляющей	нормальное
Среднеквадратичное отклонение	5
Частота дискретизации	0.025
Продолжительность регистрации сигнала	20 с

Отношение сигнал-шум для сгенерированного сигнала было равным 7.809 дБ. Сгенерированный сигнал был сглажен с помощью фильтра ПСС для порядка фильтра n принимающего значение от 1 до $N = 800$, где N – выбранное в экспериментах максимальное количество отсчетов (моментов времени дискретизации) зашумленного сигнала. Были рассчитаны значения мер подобия для двух временных рядов (проведено N сравнений): временных рядов исходного незашумленного сигнала и временных рядов зашумленного сглаженного сигнала (при n от 1 до $N = 800$).

Поскольку значения косинусной меры и значения коэффициентов Пирсона принимают значения на отрезке $[-1, 1]$, а значения расстояния Евклида и среднеквадратичной ошибки лежат вне отрезка $[-1, 1]$, для приведения значений всех мер к одному диапазону значений, к мерам расстояния Евклида и среднеквадратичной ошибки было применено деление на максимальное значение (из возможных) для каждой меры подобия.

Значения нормированных мер подобия изображены на Рис. 2. По оси абсцисс отложены значения порядков фильтрации n , по оси ординат отложены значения мер подобия (рассчитанные при сравнении исходного незашумленного и зашумленного сглаженного временных рядов).

Поскольку среди мер, анализируемых в экспериментах, было использовано расстояние Евклида, для которого оптимальным является наименьшее значение, то за основу было принято следующее общее условие оптимума: для выбранных мер подобия оптимальным является минимальное возможное значение.

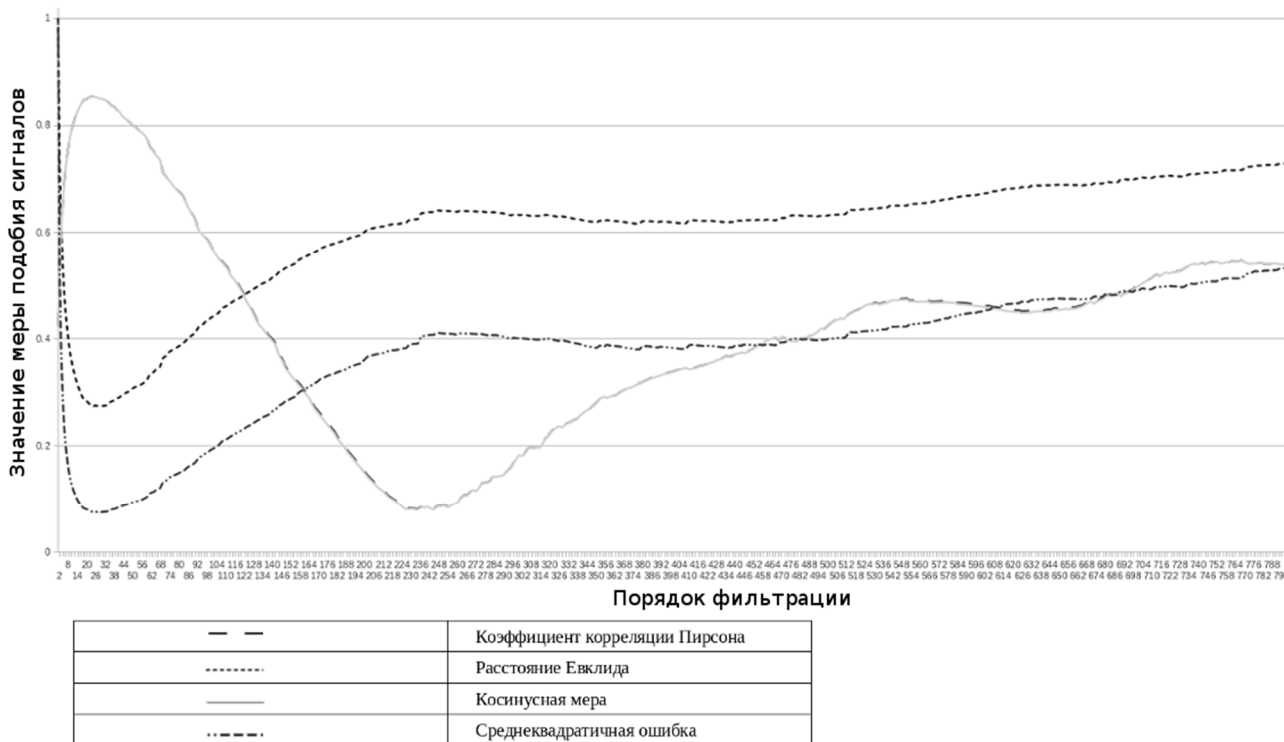


Рис. 2. Значения мер подобия сигналов (ось ординат) для разных порядков фильтрации (n от 1 до $N = 800$) (ось абсцисс)

Для того, чтобы привести все меры к одному условию оптимума, значение меры корреляции Пирсона (допустимые значения на отрезке $[-1, 1]$) и значение косинусной меры (допустимые значения на отрезке $[-1, 1]$) вычитались из единицы.

После приведения мер к общему условию оптимума и выполнению необходимых расчетов были построены графики, отображающие нормированные значения мер подобия сигналов (см. Рис. 3).

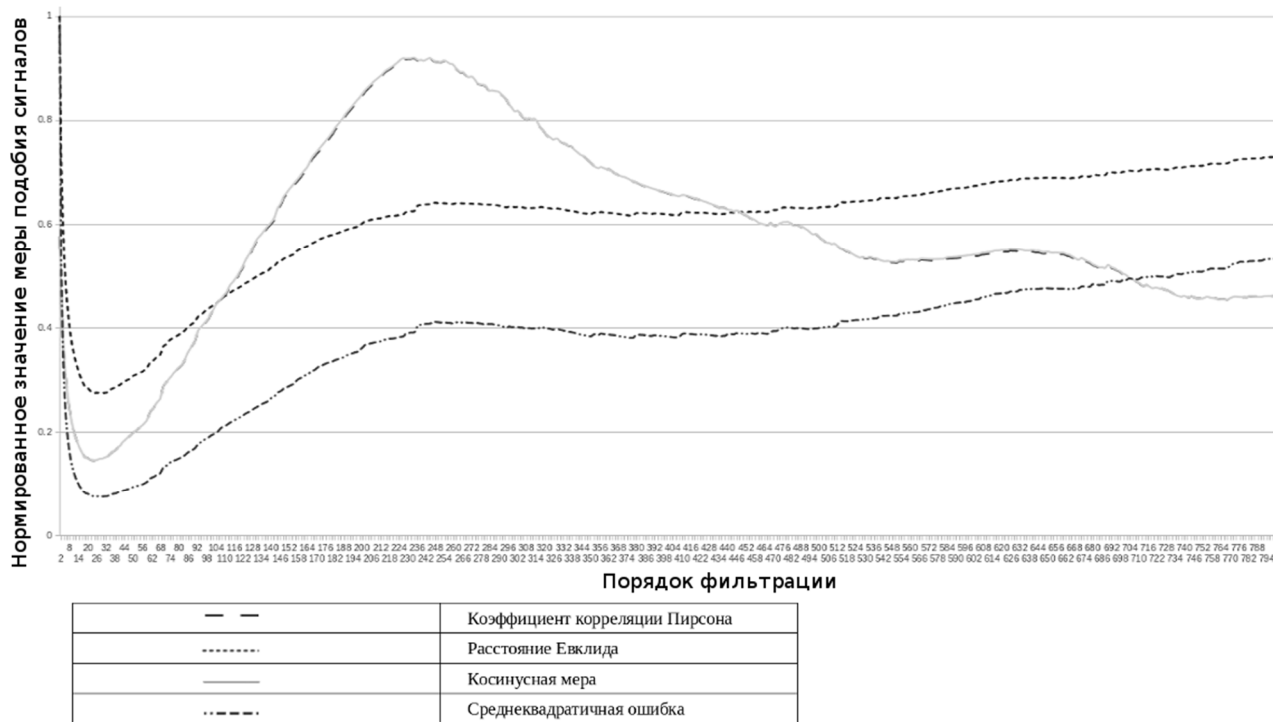


Рис. 3. Значения мер подобия сигналов (ось ординат) для разных порядков фильтрации (n от 1 до $N = 800$) (ось абсцисс)

По оси абсцисс отложены значения порядков фильтрации n , по оси ординат отложены нормированные значения мер. Из Рис. 3 видно, что значения выбранных мер подобия имеют по одному глобальному оптимуму, т. е. могут быть рассмотрены в качестве целевых функций.

На Рис. 4 изображены гистограммы, которые построены для количества порядков фильтрации из Рис. 3. По оси абсцисс отложены значения нормированных мер подобия $[0, 1]$, а по оси ординат отложены количества порядков фильтрации, при которых было получено значение меры.

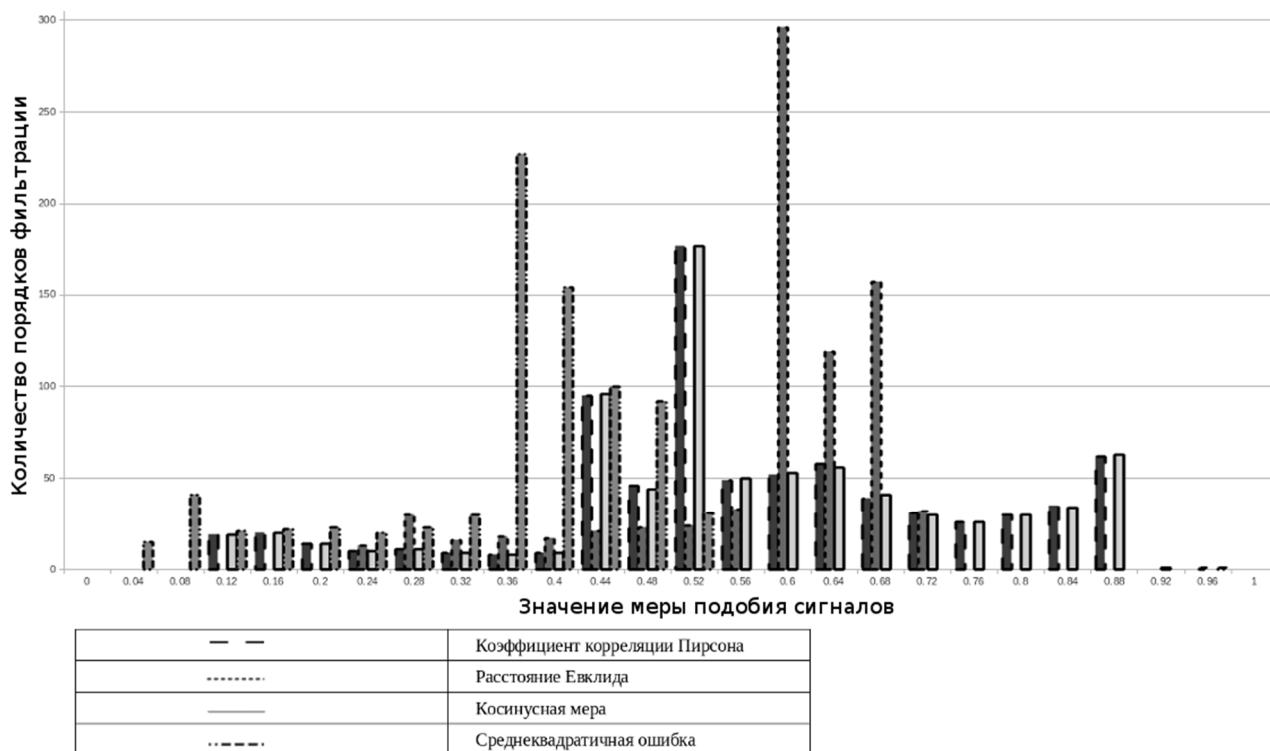


Рис. 4. Количество порядков фильтрации (n от 1 до $N = 800$) (ось ординат) с равным значением мер подобия (ось абсцисс)

Из Рис. 4 видно, что отношение оптимальных и неоптимальных порядков фильтрации ПСС варьируется для разных мер подобия.

Для оценки распределения последовательности значений в теории вероятностей используется мера эксцесса, определяемая вторым из приведенных ниже соотношений:

$$\bar{M} = \frac{1}{N} \cdot \sum_{n=1}^N M(n), \quad (8) \quad \varepsilon = \frac{N \cdot \sum_{n=1}^N (M(n) - \bar{M})^4}{\left(\sum_{n=1}^N (M(n) - \bar{M})^2 \right)^2}, \quad (9)$$

где N – общее количество моментов времени дискретизации исходного (сгенерированного) сигнала; $M(n)$ – значение меры подобия для порядка фильтрации n .

Эксцесс позволяет измерить остроту пика распределения значений ряда, а в нашем случае значений мер подобия для порядков фильтрации от 1 до $N = 800$.

В [2] для расчета оптимальных параметров цифрового фильтра были выделены два лучших алгоритма: рой частиц и генетический алгоритм. В обоих алгоритмах поиск начинался с генерирования заданного количества начальных значений n (порядка фильтрации) и m (порядка аппроксимации). В контексте текущей статьи может быть выбрано 800 значений n (значений порядка фильтрации), каждое с вероятностью 0.00125.

Как отмечено выше, значение эксцесса позволяет численно определить остроту пика распределения значений порядков фильтрации. Можно сделать вывод, что при увеличении значения эксцесса будет увеличиваться количество порядков фильтрации с одинаковым

значением меры подобия (их вероятность будет увеличиваться), при этом вероятность нахождения глобального оптимума будет уменьшаться.

Приведем далее результаты вычисления значений эксцесса для сигналов разной формы и разных уровней шумовой составляющей.

4. Сравнение мер подобия. Выше мы предложили использовать значение эксцесса для оценки значений мер подобия при разных порядках n фильтрации. Рассмотрим этот вопрос более внимательно. В результате выполненных нами экспериментов были сгенерированы сигналы имеющие шумовую составляющую с различным значением среднеквадратичного отклонения σ .

На Рис. 5 приведены значения эксцесса мер подобия сигналов: исходного незашумленного и зашумленного сглаженного (для значений σ , выбранных в пределах от 1 до 50). По оси абсцисс отложены значения σ , по оси ординат отложены значения эксцесса мер подобия для разных порядков фильтрации от 1 до $N = 800$.

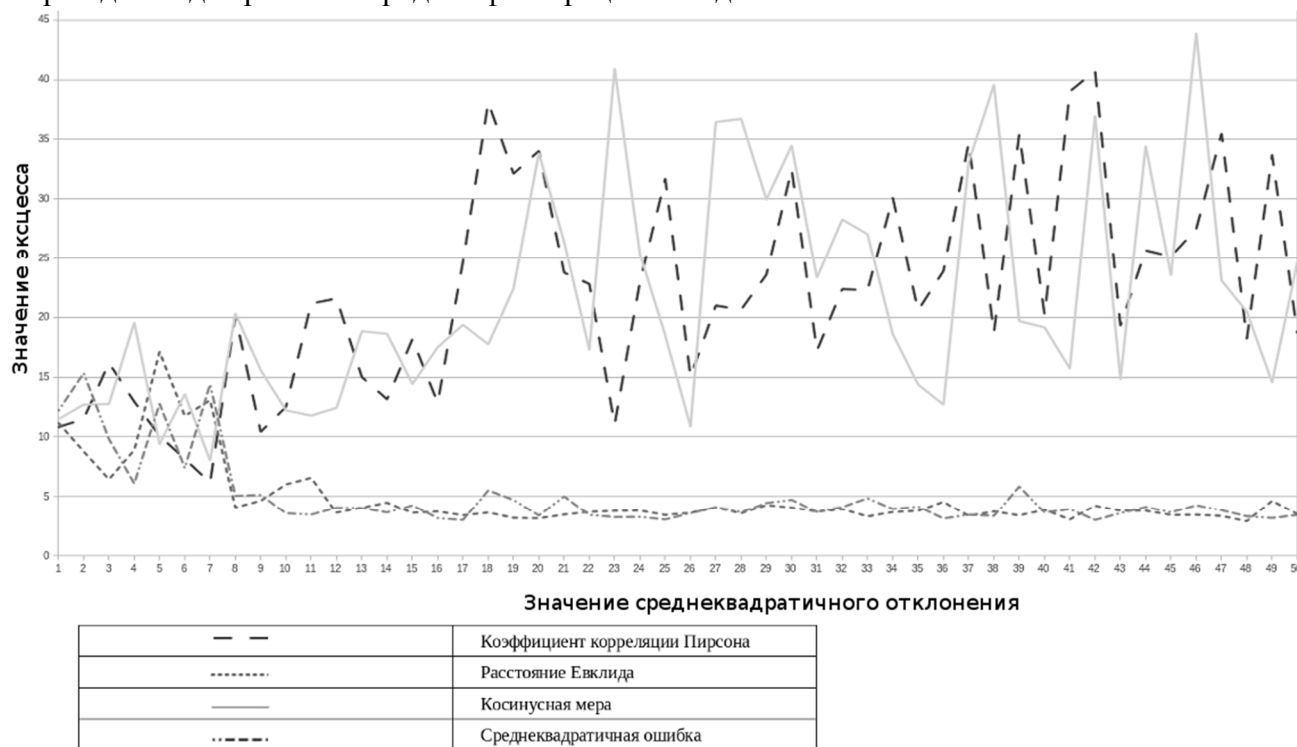


Рис. 5. Значения эксцесса (ось ординат) для мер подобия сигналов: исходного незашумленного и зашумленного сглаженного (ось абсцисс)

В целях полноты экспериментов количество форм тестовых сигналов было расширено до четырех: синусоидального, прямоугольного, треугольного и пилообразного. Для всех этих форм сигналов в Табл. 3 приведены значения математического ожидания эксцесса для мер подобия сигналов: исходного незашумленного и зашумленного сглаженного ($\sigma = 1 \dots 50$).

Значения математического ожидания эксцесса для мер подобия сигналов

Табл. 3

Мера подобия	Форма сигнала			
	Синусоидальная	Прямоугольная	Треугольная	Пилообразная
Коэффициент корреляции Пирсона	25.73	22.69	28.71	26.51
Расстояние Евклида	3.98	4.18	4.11	4.01
Косинусная мера	26.16	25.03	27.34	28.31
Среднеквадратичная ошибка	3.77	4.03	3.73	3.85

Оценив средние значения эксцесса для мер подобия, приведенные в Табл. 3 и на Рис. 5, можно сделать следующие выводы:

– Среднеквадратичная ошибка как мера подобия сигналов имеет наименьшее среднее значение эксцесса, следовательно позволяет повысить вероятность стохастического расчета оптимального порядка фильтрации.

– По среднему значению эксцесса рассмотренные меры подобия можно разделить на две группы:

* *первая группа*: расстояние Евклида и среднеквадратичная ошибка;

* *вторая группа*: коэффициент корреляции Пирсона и косинусная мера.

5. Выводы. Сравнение сигналов требуется при решении различных технических задач, в том числе при проектировании адаптивных цифровых фильтров. Формальной оценкой степени близости сигналов может служить мера подобия, определенная тем или иным способом. От выбора меры подобия зависит то, насколько оптимальным будет решение вопроса о порядке фильтра.

Для того, чтобы исследовать данный вопрос, была создана тестовая среда (программная библиотека), включающая в себя модули, генерирующие сигналы разной формы, и модули анализирующие эти сигналы. Основными компонентами библиотеки являются: модуль генерирования сигнала (модуль генерирования сигнала по функции, модуль генерирования шумовой составляющей), модуль нерекурсивной фильтрации, модуль оценки качества фильтрации (оценки подобия сигналов и статистической оценки).

Для оценки изменения значений мер подобия для разных сигналов с помощью разработанной библиотеки были сгенерированы сигналы следующих форм: синусоидальный, прямоугольный, треугольный и пилообразный. К сгенерированным значениям была добавлена случайная составляющая (шум), после чего полученный сигнал был сглажен с помощью нерекурсивного фильтра. Было показано, что применяя к зашумленному сигналу фильтрацию с разным порядком, можно найти значение порядка, при котором мера подобия принимает наименьшее значение (при этих условиях порядок фильтрации можно считать оптимальным).

Для сравнения мер подобия предложено использовать эксцесс. В результате оценки значений эксцесса при различных уровнях шума было установлено, что: 1) рассмотренные меры по среднему значению эксцесса можно разделить на две группы; 2) мера среднеквадратичной ошибки имеет наименьшее значение эксцесса. Кроме того, из результатов экспериментов следует, что среднеквадратичная ошибка позволяет повысить вероятность стохастического расчета оптимального порядка фильтрации стационарных сигналов.

Реализованная библиотека может быть использована для экспериментальной обработки цифровых сигналов в различных технических задачах.

Литература

1. Сергиенко А. Цифровая обработка сигналов / А. Сергиенко. – [3-е изд.]. – Санкт-Петербург : изд. «БХВ-Петербург», 2011. – С. 593-595.
2. Rogoza V., Sergeev A., The Comparison of the Stochastic Algorithms for the Filter Parameters Calculation, Springer, Advances in Systems Science, Vol. 240, Switzerland, 2014. – 241-250 pp.
3. Cha S., Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions, M3AS, Issue 4, Vol. 1, Singapore, 2007. — 300-307 pp.
4. Загоруйко Н. Прикладные методы анализа данных и знаний / Н. Загоруйко, – Новосибирск : издательство Института математики, 1999. – С. 195-199.
5. Wang D., Ronsin J., Veronique H., Compared performances of morphological, median type and running mean filters, SPIE, Vol. 1818, 1992. – 384-391 pp.
6. Xu R., Wunsch D., Survey of Clustering Algorithms, IEEE Transactions on Neural Networks, Vol. 16, No. 3, 2005. – 645-678 pp.
7. Окунь Я. Факторный анализ / Я. Окунь. Москва : изд. «Статистика», 1974. – С. 52-57.
8. Смирнов Н. В. Курс теории вероятностей и математической статистики / Н. В. Смирнов, И. В. Дунин-Барковский. Москва : издательство «Наука», 1969. – С. 85-94.