



# OSTIS-2012

(Open Semantic Technologies for Intelligent Systems)

УДК 004.896

## ОБРАБОТКА КОНТЕКСТНЫХ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

А.Г.Киселёва, Г.Д.Киселёв

*Национальный технический университет Украины «Киевский политехнический институт»  
г. Киев, Украина*

[kiseleva\\_anna@ukr.net](mailto:kiseleva_anna@ukr.net)

[kiselev\\_gd@bigmir.net](mailto:kiselev_gd@bigmir.net)

В статье описана библиотека программ позволяющая повысить достоверности исходных данных в контекстных приложениях. Приводится анализ применения различных методов фильтрации и сглаживания временных рядов. Оценивалась эффективность методов экстраполяции и линейного предсказания.

**Ключевые слова:** контекстно-зависимые приложения, контекст, сглаживание, предсказание.

### ВВЕДЕНИЕ

Контекстно-зависимыми приложениями (КЗП) называются интеллектуальные системы, в которых обеспечивается эффективный интерфейс с пользователем и с другими программными системами за счет использования контекстных данных, получаемых от разных информационных источников. Под контекстными данными или контекстом понимают информационные потоки (временные ряды), идущие от датчиков (сенсоров), сканирующих окружающее пользователя пространство (оборудование, программное обеспечение, физическое пространство и социальная среда). На физическом уровне входные и выходные информационные потоки переносятся электрическими сигналами.

Сложность решения задач обработки физических сигналов поступающих от различных датчиков обусловлена тем, что для регистрации информационных потоков поступающих от пользователя и снимаемых с датчиков, помимо полезного сигнала, фиксируются электрические и магнитные сигналы от посторонних источников. Такие помехи, неизбежно возникающие в реальных условиях, рассматриваются как аддитивный шумовой компонент, искажающий полезный сигнал. В общем случае, создатели КЗП не имеют данных о характеристиках шумовой компоненты, которые могут изменяться в зависимости от типа

сенсоров и условий эксплуатации системы. В связи с этим, актуальной является задача использования средств выделения полезного сигнала на фоне помехи, используя различные алгоритмы фильтрации и сглаживания сигналов, которые в минимальной степени искажают форму информативных фрагментов физических сигналов. В настоящей статье исследуется один из возможных подходов к решению задачи выбора эффективных алгоритмов фильтрации и сглаживания, а так же прогнозирования входных сигналов в распределенных интеллектуальных системах, в которых основным потребителем и источником контекстной информации является современные устройства, например, многофункциональный мобильный телефон. Для таких систем на алгоритмы подавления шумов и прогнозирования накладываются требования по минимизации потребляемых вычислительных ресурсов и времени вычислений.

Оцифрованные входные сигналы от контекстных источников входных сигналов, в частности, от датчиков мобильного телефона, описываются в виде временных рядов, членами которых являются вещественные числа. Поэтому, используемые алгоритмы прогнозирования, сглаживания и фильтрации сигналов должны эффективно работать с временными рядами. Краткосрочный прогноз значений членов временного ряда используется не только для плавного изменения выходных характеристик устройств, но и для

контроля достоверности контекстообразующих временных рядов.

### Предмет исследования

Авторами статьи были разработаны и протестированы программы фильтрации и сглаживания временных рядов для следующих известных методов:

1. Метод простого скользящего среднего (ПСС) (SMA — SimpleMovingAverage);
2. Метод двойного скользящего среднего (ДСС) (DMA — DoubleMovingAverage)
3. Метод экспоненциального сглаживания (ЭС) (ES — ExponentialSmoothing);
4. Метод двойного экспоненциального сглаживания (ДЭС) (DES - DoubleExponentialSmoothing);
5. Фильтра Кальмана(КФ)(FilterKalmana - FK)

Методы скользящего среднего и экспоненциального сглаживания строятся на предположении, что при определении средних значений в некотором числовом ряду случайные отклонения погашаются. При сглаживании этими методами фактические значения членов временного ряда заменяются средними значениями, которые определяют срединную точку периода скользящего [2].

При реализации метода простого скользящего среднего (ПСС) используется формула:

$$s_t = \frac{1}{k} \sum_{n=0}^{k-1} x_{t-n} = s_{t-1} + \frac{x_t - x_{t-k}}{k}, \quad \text{где } k -$$

порядок скользящего среднего или количество предыдущих значений временного ряда, для которых вычисляется среднее значение;  $x_t$  — предыдущее зашумленное значение временного ряда;  $s_t$  — сглаженное значение очередного элемента временного ряда.

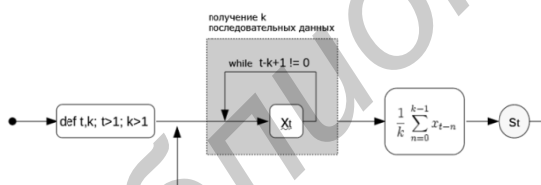


Рисунок1 –функциональная схема алгоритма вычисления простого скользящего среднего

В методе двойного скользящего среднего (ДСС) сглаженное значение элементов временного ряда вычисляется в два этапа. На первом этапе рассчитывается первичное сглаженное значение очередного элемента временного ряда. На втором этапе это значение уточняется, при этом, каждое зашумленное значение суммируется с рассчитанным на первом этапе сглаженным значением.(рис. 2).

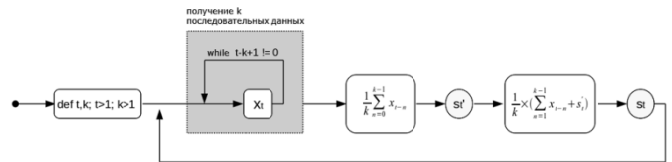


Рисунок2 – функциональная схема алгоритма вычисления двойного скользящего среднего

Скользящее среднее не даст точного сглаживания, если временные ряды монотонно возрастают или убывают. Этот метод лучше подойдет для ряда с небольшими случайными отклонениями данных от некоторого постоянного или медленно изменяющегося значения.

В методе простого экспоненциального сглаживания (ПЭС) весовые коэффициенты предыдущих наблюдаемых значений временного ряда увеличиваются по мере приближения к текущему (по времени) элементу в соответствии с формулой[3]:

$$s_t = \alpha x_t + (1-\alpha)s_{t-1}, \quad \text{где } \alpha - \text{ параметр сглаживания.}$$

Выбор параметра сглаживания  $\alpha$  представляет собой достаточно сложную проблему. Чем ближе параметр сглаживания к единице, тем больше влияние последних наблюдений и тем больше скорость убывания весов предыдущих наблюдений. Однако, если высокочастотная (шумовая) компонента ряда имеет достаточно большую дисперсию, не следует использовать большие значения параметра сглаживания из-за ухудшения качества прогноза.

Метод экспоненциального сглаживания не дает удовлетворительных результатов, если значения временного ряда монотонно возрастают или убывают. В таких случаях может быть применен метод экспоненциального сглаживания с учетом тренда. В частности, модель Хольта [4], в которой используется два уравнения – первое для сглаживания элементов ряда и второе для сглаживание тренда:

$$s_t = \alpha \gamma_t + (1-\alpha)(s_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1, \\ b_t = \gamma(s_t - s_{t-1}) + (1-\gamma)b_{t-1} \quad 0 \leq \gamma \leq 1,$$

где  $S_t$ — сглаженное значение прогнозируемого элемента для времени  $t$ ;  $b_t$  - оценка прироста тренда, показывающая возможное возрастание или убывание значений за одно приращение времени;  $\alpha, \gamma$ - параметры сглаживания ( $0 \leq \alpha \leq 1; 0 \leq \gamma \leq 1$ );  $k$  - количество периодов времени, на которые на которые производится прогноз.

Параметры сглаживания  $\alpha$  и  $\gamma$  выбираются. Чем больше значения параметров, тем большему сглаживанию подвергаются данные. При  $\alpha=\gamma$  имеет место особый случай, поскольку в одинаковой мере производится сглаживание текущего значения и тренда. Такой вариант называется двойным экспоненциальным сглаживанием (ДЭС) Брауна [5].

Фильтр Кальмана (ФК), для расчета сглаженного значения очередного элемента

временного ряда, использует его зашумленное значение, которое пересчитывается с использованием уточняемого для каждого последующего значения временного ряда параметра ошибки [6]. Достоинства ФК:

- имеет рекуррентную формулу;
- легко обобщается для многомерных и нестационарных временных рядов;
- легко модифицируется на случай ненулевых математического ожидания шума коррелированных шумов.

Уравнения ФК имеют вид:

$$\text{начальные условия: } P_{k,k} = P_{k,k-1}(I - K_k); \quad x_{1,1} = z_1,$$

$$x_{2,1} = z_2, \quad P_{1,1} = R$$

$$\text{экстраполяция: } \hat{x}_{k,k-1} = \hat{x}_{k-1,k-1} + q$$

$$P_{k,k-1} = P_{k-1,k-1} + Q$$

$$\text{фильтрация: } \hat{x}_{k,k} = \hat{x}_{k,k-1} + K_k(z_k - \hat{x}_{k,k-1});$$

$$K_k = P_{k,k}(R - P_{k,k}); \quad P_{k,k} = P_{k,k-1}(I - K_k);$$

где  $\hat{x}_{k,k}$  - оценка вектора состояний по  $k$  измерениям в  $k$ -й момент времени (оценка фильтрации);  $P_{k,k}$  - ковариационная матрица ошибки оценки фильтрации;  $\hat{x}_{k,k-1}$  - экстраполированная оценка вектора состояний на  $k$ -й момент времени по  $k-1$  измерениям;  $P_{k,k-1}$  - ковариационная матрица ошибки оценки вектора экстраполяции;  $K_k$  - коэффициент усиления ФК;  $q$  - оценка среднего значения шума состояния;  $Q$  - оценка дисперсии шума состояния;  $R$  - оценка дисперсии шума измерения.

Для оценки применимости в контекстно-зависимых приложениях мобильного телефона предсказаний значений членов временного ряда полученных с применением методов линейного предсказания и экстраполяции, авторами статьи были разработаны и протестированы алгоритмы и программы для следующих известных методов:

1. Метода линейного предсказания (ЛП);
2. Метода линейной экстраполяции (ЛЭ);
3. Метода полиномиальной экстраполяции (ПЭ);
4. Метода сплайн экстраполяции (СЭ).

Основная задача при построении модели линейного предсказания (в частности модели авторегрессии) состоит в том, чтобы правильно отыскать параметры этой модели [7]. Для этой цели используется система уравнений Юла-Уолкера.

Каждый элемент выборки ЛП умножается на соответствующий весовой коэффициент. Величина зависимости  $x_{n+1}$  члена временного ряда от предыдущих  $n$  членов называется коэффициентом автокорреляции.

Экстраполяцией называют особый вид аппроксимации, при котором функция аппроксимируется вне заданного интервал [8].

### 1. Метод линейной экстраполяции

В данном методе, искомое значение вне интервала  $[x_0, x_n]$  ищется с помощью уравнения прямой,

построенной на 2-х точках лежащих на границах интервала:  $f(x) = y_n + (x - x_n)/(x_0 - x_n)(y_0 - y_n)$ .

2. Метод полиномиальной экстраполяции  
Полиномиальная экстраполяция [4] – метод, в котором вместо уравнения прямой линии на наборе из  $N$  точек  $[x_0, x_n]$  строится полиномиальная функция  $P$ , которая проходит через эти точки:  $y_i = P(x_i)$ , для всех  $1 \leq i \leq N$  и вычисляется значение функции в точке  $x_{n+1}$ .



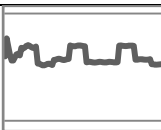
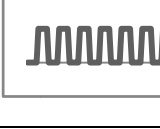
### 3. метод сплайн экстраполяции

В данном методе экстраполяция выполняется на основе сплайнов [9,10], с помощью которых строится результирующий полином, проходящий через все точки входного ряда.

## Тестирование методов обработки контекстных данных

Для проверки эффективности работы алгоритмов фильтрации и сглаживания входных сигналов была создана тестовая среда, в которой использованы тестовые временные ряды, показанные в таблице 1. Созданная тестовая среда использовалась для отладки и анализа разработанных авторами программ, для рассмотренных в работе методов.

Таблица 1. Тестовые временные ряды

Номер теста	Тестовый временной ряд	Графическая интерпретация
1	$y = \text{tresh}$	
2	$y = \text{tresh} * \sin(3.14 * i / N)$	
3	$y = \text{tresh} + 2 * \sin(i/10) * \cos(i/10) + 2 * \sin(i)/i$	
4	$y = 10 * ([1/\text{tresh} * x] / 10 - [1/\text{tresh} * x] / 10)$	

Для тестирования программ фильтрации и сглаживания временных рядов был выбран генератор шума на основе источника случайных чисел с нормальным (Гаусовский) распределением с заданным математическим ожиданием  $\mu$  и дисперсией  $\sigma^2$  для компоненты шума. В блоке сравнения тестовой среды сравнивается сглаженный зашумленный временной ряд с исходным тестовым временным рядом. Результаты сравнения характеризуют качество алгоритмов сглаживания.

## Результаты тестирования

На рисунке 3 показаны результаты тестирования фильтра Кальмана для зашумлённого тестового временного ряда полученной с акселерометра мобильного телефона (тест номер 3). Дисперсия шума (сигма) равна 0.5. Из-за небольшого уровня зашумлённости и особенностей тестового ряда (имеет резкие скачки), фильтр Кальмана сглаживает не только шум, но и саму функцию, из-за чего сильно сгладилась её форма и появился эффект запаздывания. Такой же результат показал и метод простого скользящего среднего. Отфильтрованная функция сохраняет периодичность исходной тестовой функции.



Рисунок 3 - Результаты тестирования фильтра Кальмана сгенератором шума с дисперсией 0.5

На рисунке 4 показан тот же тестовый временной ряд, как и в первом случае, при этом дисперсия шума (сигма) равна 5. Как видно из рисунка, уровень шума практически больше уровня значений тестового временного ряда. Поскольку процент сглаживания высок, форма исходного тестового ряда практически не различима на уровне шумов, так же не сохраняется периодичность повторения и уровни значений его амплитуды. Таким образом, для сложных, сильно зашумленных временных рядов алгоритмы сглаживания/фильтрации показывают плохой, зачастую неприемлемый, результат.

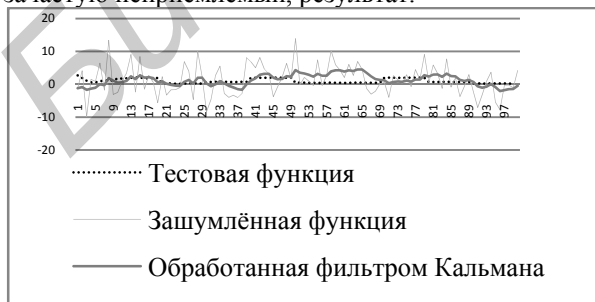


Рисунок 4 - Результаты тестирования на генераторе шума с дисперсией 5

Результаты тестирования программ для всех исследованных методов фильтрации и сглаживания временных рядов приведены в таблице 2 на примере тестов 1-4. Программы тестировались на компьютере с процессором с тактовой частотой 2.26ГГц. Длина тестового временного ряда 1000 элементов.

Таблица 2. Среднеквадратическое отклонение

Дисперсия шума Sigma=0.5				
Метод сглаживания	Тест1	Тест2	Тест3	Тест4
ФК	0,158017	0,349812	0,467901	0,325037
ПСС	0,132832	0,309358	0,788688	0,888607
ДСС	0,0767471	0,295795	0,696336	0,334083
ПЭС	0,0796013	0,299081	0,75302	0,889265
ДЭС	0,0763433	0,296688	0,785822	0,889241
Дисперсия шума Sigma=5				
Метод сглаживания	Тест1	Тест2	Тест3	Тест4
ФК	1,57463	1,56321	1,61398	1,65909
ПСС	1,2332	1,06842	1,09635	1,19313
ДСС	0,76762	0,785053	0,646097	1,10649
ПЭС	0,75865	0,696137	0,979228	0,897497
ДЭС	0,76054	0,67862	0,769288	0,846703

Кроме того, в процессе тестирования оценивалось ресурсопотребление исследуемых программ. Испытания проводились на компьютере с процессором с тактовой частотой 2.26ГГц. Контролировалось время работы программы соответствующего алгоритма и объем используемой памяти для временных рядов длиной 1000 членов. Полученные результаты приведены в таблице 3.

Таблица 3. Ресурсопотребление программ фильтрации и сглаживания временных рядов

	ФК	ПСС	ДСС	ПЭС	ДЭС
Время обработки временного ряда(мс)	1.157	2.181	2.865	3.021	5.543
Важествовано памяти(Б)	540	454	596	468	954

Тестирование программ для всех методов сглаживания с дисперсией шума равной 0,5 приведено на рисунке 5.

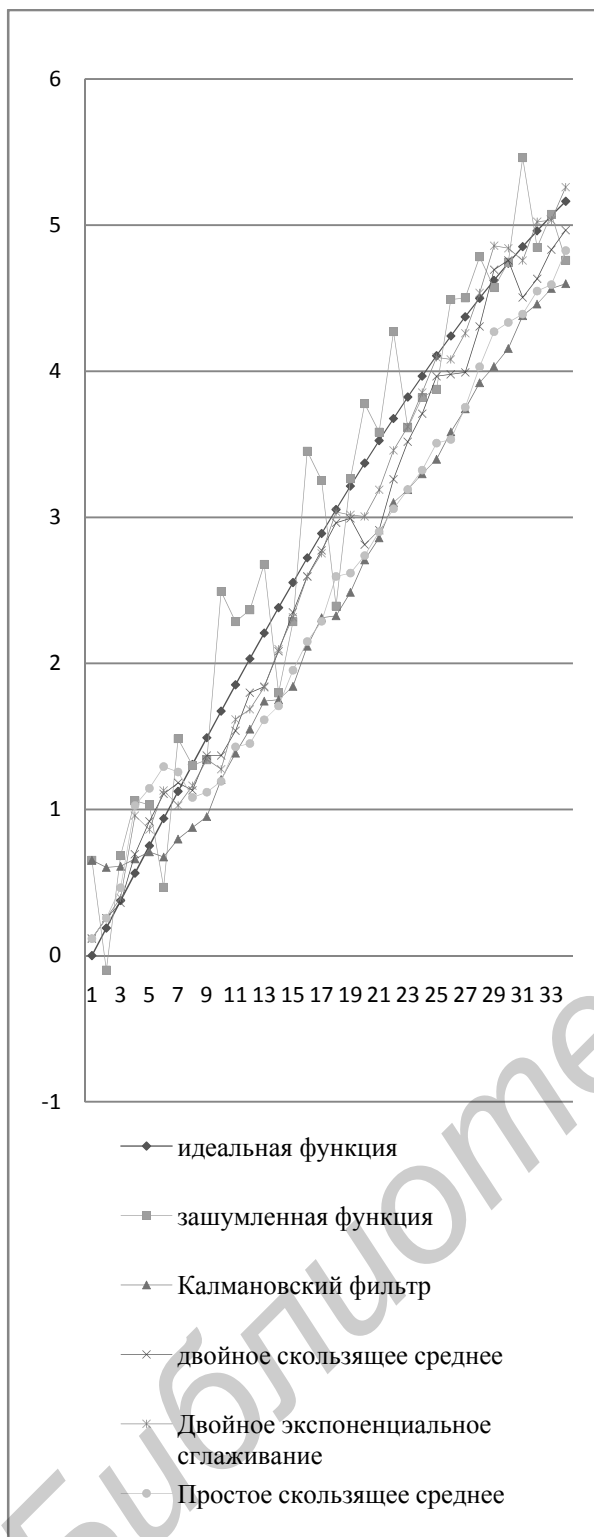


Рисунок 5 – Сглаживание исходного временного ряда с применением методов ПСС, ДСС, ДЭС, КФ

Из рисунка видно, что при гладких входных функциях методы ДСС и ДЭС дают наиболее приемлемый результат, но при этом они наиболее ресурсоёмкие.

Результаты тестирования программ реализующих исследуемые методы предсказания временных рядов приведены в таблице 4. Качество прогнозирования оценивается по величине среднеквадратической ошибки прогноза (табл.4).

В первом разделе таблицы приведены результаты прогнозирования для обучающей выборки равной двум, во втором - десяти.

Таблица 4. Среднеквадратическая ошибка прогноза

Количество элементов в обучающей выборке P=2				
Метод предсказания	тест1	тест2	тест3	тест4
ЛП	0,3333	0,6382	0,422482	2,50000
ЛЭ	0,0000	0,0033	2,524410	7,50000
ПЭ	1,0000	0,0047	3,524410	7,50000
СЭ	0,0000	0,0047	2,524410	7,50000
Количество элементов в обучающей выборке P=10				
Метод предсказания	тест1	тест2	тест3	тест4
ЛП	0,0909	0,6204	0,366496	1,66667
ЛЭ	0,0000	0,0204	3,040490	-
СЭ	0,0000	0,0204	0,197941	9,46078

## ЗАКЛЮЧЕНИЕ

В контекстно-зависимых приложениях, как правило, необходимо использовать предсказания возможных значений членов временного ряда снимаемого с датчиков телефона. Использование для этих целей метода линейного предсказания ограничено тем, что предсказание возможно только для одного следующего члена временного ряда. Размер выборки ЛП сильно влияет на качество предсказания. При малой выборке, могут пропускаться долгосрочные тренды временного ряда.

Методы экстраполяции более эффективны для краткосрочных прогнозов, при которых требуется предсказание нескольких членов временного ряда. Они достаточно просты и потребляют мало ресурсов процессора мобильного телефона. Линейная экстраполяция применима для задач, в которых временной ряд изменяется без резких скачков, практически линейно. Данный метод имеет высокое быстродействие, но низкую точность на нелинейных участках входного ряда. Экстраполяция сплайнами более гибкая, чем полиномиальная, т.к. при добавлении нового члена временного ряда в интервал экстраполяции не требуется построение полинома на всём интервале, а лишь на ближайших к искомой точках.

Контекстно-зависимые приложения, работающие с зашумленными исходными данными, формируют ошибочные результаты. Фильтрация или сглаживание соответствующих временных рядов является обязательным этапом, повышающим достоверность исходных данных. Проведенные эксперименты подтвердили эффективность применения рассмотренных методов подавления случайной шумовой компоненты для реальных сигналов, снимаемых с датчиков мобильных устройств. Созданная библиотека программ сглаживания и фильтрации, а так же прогнозирования временных рядов позволяет создавать новые приложения для решения контекстно-зависимых задач.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [Oviatt, 2003] Oviatt S. L. Multimodal Interfaces/Oviatt S. L. // The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Jacko J. and Sears A. (Eds.). Mahwah, NJ: Lawrence Erlbaum Assoc. 2003. P. 286–304.
- [Kendall, 1983] Kendall M. G., Kendall's advanced theory of statistics/ Kendall, M. G., Stuart, A., Ord, J. K. // Vol. 3, Hodder Arnold, London
- [Лукашин, 2003] Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов / Лукашин Ю. П. // М. : Финансы и статистика, 2003. – 415 с.
- [Gardner, 1980] Gardner Jr. Forecasting with Exponential Smoothing: Some Guidelines for Model Selection/ Gardner Jr., Everette S. and David G. Dannenbring // In Decision Sciences, 11., 370-383, 1980.
- [Bowerman, 1993] Bowerman, Forecasting and Time Series: An Applied Approach/Bowerman, Bruce J. and Richard T. O'Connell. // Duxbury Thomson Learning, 1993.
- [Згуровский, 1995] Згуровский М. З. Аналитические методы калмановской фильтрации для систем с априорной неопределённостью / М. З. Згуровский, В. Н. Подладчиков – К. : Наукова думка, 1995. – 283 с.
- [Грэй, 1980] Грэй А. Х. Линейное предсказание речи/Грэй А. Х. // М. связь.-1980.-25 стр.
- [Казаков, 1987] Казаков И.Е. Методы оптимизации стохастических систем/Казаков И.Е., Гладков Д.И. // М. 1987.- 304
- [Brezinski, 1991] Brezinski C. Extrapolation Methods. Theory and Practice/ Brezinski C. Redivo Zaglia M. // North-Holland. - 1991.
- [Роджерс, 2001] Роджерс Д. Математические основы машинной графики/Роджерс Д., Адамс Дж. // М.: Мир, 2001. — ISBN 5-03-002143-4

## CONTEXT DATA PROCESSING FOR INTELLIGENT SYSTEM

Kyselova A.G, Kyselov G.D.

National Technical University of Ukraine  
"Kiev Polytechnic Institute"

[kiselev\\_gd@bigmir.net](mailto:kiselev_gd@bigmir.net)

[kiseleva\\_anna@ukr.net](mailto:kiseleva_anna@ukr.net)

## INTRODUCTION

Context-aware system is employed in a variety of contemporary information systems in order to enhance the naturalness, flexibility and convenience of user interface. Context is a mode of communication according to human senses (sight, touch, hearing, smell, and taste) or type of computer input devices (input devices equivalent to human senses: cameras (sight), haptic sensors (touch), microphones (hearing), olfactory (smell) and input devices that do not map directly to human senses: keyboard, mouse, writing tablet, motion input and others). A multimodal interface provides several distinct tools for input and output of data. Data fusion input data is much researched in the field of sensor networks where the primary motivation is to reduce communication costs by the integration of similar data sources.

**Keywords** – context-aware system, context, smoothing, prediction.

## RAW DATA-PROCESSING PACKAGE

The raw data-processing package to evaluate the reliability of data information for sensors environment. That it consist of algorithms that smoothed data because the data information influences by environment status may be imperfect. An often-used technique in industry is "smoothing". This technique, when properly applied, reveals more clearly the underlying trend, seasonal and cyclic components.

There are three distinct groups of smoothing methods

- Averaging Methods
- Exponential Smoothing Methods
- Kalman filter

And prediction method:

- Linear prediction
- Spline extrapolation
- Polynomial extrapolation
- Line extrapolation

## CONCLUSION

The goal of testing is to estimate algorithm's quality, trace functionality issue, and estimate resources consumption and estimate of difference between the data after the algorithm with original data.

The result of testing is choose algorithm in which general restrictions typical for mobile environments shall be kept in mind:

- low memory consumption
- error tolerance for the input data
- speed