

6. БЛАГОДАРНОСТИ

Автор благодарит организацию INTAS за поддержку согласно гранта INTAS-OPEN-97-0606.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. http://www.honeywell.com/sensing/prodinfo/temperature/catalog/c15_93.pdf
2. Самсонов Г.В., Киц А.И., Кюздени О.А., "Датчики для измерения температуры в промышленности", Наукова думка, Киев, 1972
3. www.fluke.com/products/data_acquisition/hydra/home.asp?SID=7&AGID=0&PID=5308
4. Iyengar S.S. Distributed Sensor Network - Introduction to the Special Section // Transaction on Systems, Man, and Cybernetics. – 1991. - Vol. 21, No.5. - P. 1027-1031.
5. Brignell E.J. Digital compensation of sensors/ Scientific Instruments. – 1987. - Vol. 20, No 9. - P. 1097-1102.
6. Jayasimha D.N., Iyengar S.S., Kashyap R.L. Information Integration and Synchronization in Distributed Sensor Networks // Transaction on Systems, Man, and Cybernetics. – 1991. - Vol. 21, No.5. - P. 1032-1043.
7. K.B. Lee, R.D. Schneeman. Distributed Measurement and Control Based on the IEEE 1451 Smart Transducer Interface Standards // Proceedings 16th IEEE Instrumentation and Measurement Technology Conference (IMTC/99). – Venice (Italy). – 1999. - P. 608-613.
8. Kazakos D. Asymptotic Error Probability Expressions for Multihypothesis Testing Using Multisensor Data // Transaction on Systems, Man, and Cybernetics. – 1991. - Vol. 21, No.5. - P. 1101-1114.
9. C.Alippi, A.Ferrero, V.Piuri. Artificial Intelligence for Instruments & Applications/ IEEE Instrumentation and Measurement Magazine. – June 1998. – P. 9-17.
10. A.Sachenko, V.Tymchyshyn. Low Cost Intelligent Module for Distributed Sensor Network/ Proc. of 4th Symposium Low Cost Automation "LCA'95". - Buenos Aires (Argentina). - 1995. - P.197-202.
11. Golovko V., Grandinetti L., Kochan V., Laopoulos T., Sachenko A., Turchenko V. Sensor Signal Processing Using Neural Networks/ Proc. IEEE Region 8 Intern. Conf. Africon'99. - Cape Town (South Africa). – 1999. - P. 339-344.
12. A.Sachenko, V.Kochan, V.Turchenko. Intelligent Distributed Sensor Network // Proceedings of 15th IEEE Instrumentation and Measurement Technology Conference (IMTC/98). – St. Paul (USA). – 1998. - P. 60-66.
13. V.Golovko, L.Grandinetti, V.Kochan, T.Laopoulos, A.Sachenko, V.Turchenko, V.Tymchyshyn. Approach of an Intelligent Sensing Instrumentation Structure Development // Proc. IEEE Intern. Workshop on Intelligent Signal Processing (WISP'99). – Budapest (Hungary). – 1999. – P. 336-341.
14. Головки В.А. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями. – Брест: Изд. БПИ, 1999. – 264 с.
15. V.Kochan, V.Tymchyshyn. Precision ADC module for IBM PC // Proc. 41.Internationales Wissenschaftliches Kolloquium. - Ilmenau (Germany). - 1996. - P. 668-672.
16. A.Sachenko, V.Kochan, V.Turchenko, T.Laopoulos, V.Golovko. Using Neural Networks for Decreasing ADC Error/ Proc. IMEKO TC-4 Intern. Workshop of ADC Modelling and Testing (IWADC/99). – Bordeaux (France). – 1999. - P.78-81.
17. A.Sachenko, V.Kochan, V.Turchenko, V.Tymchyshyn, N.Vasylyuk. Intelligent Nodes for Distributed Sensor Network/ Proceedings 16th IEEE Instrumentation and Measurement Technology Conference (IMTC/99). – Venice (Italy). – 1999. - P.1479-1484.

УДК 681.3

Саченко А.А., Кочан В.В., Коваль В.С.

СТАТИСТИЧЕСКИЕ МЕТОДЫ СЛИЯНИЯ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ ДИСТРИБУТИВНЫХ СЕНСОРНЫХ СЕТЯХ

В последнее время получили значительное распространение интеллектуальные дистрибутивные сенсорные сети [2,17,18]. Результаты измерений всегда содержит погрешности, как систематические, так и случайные [3,6], поэтому действительные значения измерительных сигналов, в особенности, при обработке сигналов сенсоров, характеризуют с учетом интервалов неопределенности. Шумы, которые влияют на процесс измерения, имеют разную структуру, причем значительный удельный вес занимают погрешности случайной природы [3,4,6]. В отличие от других, случайные шумы не компенсируются калибровкой или поверкой [6], поэтому они вызывают наибольший интерес при статистической обработке информации. Методы статистической обработки данных, кроме определения случайной погрешности измеряемых данных, разрешают их усреднить и на основе него найти как более точный результат, так и его погрешность [6]. В таких условиях актуальной задачей есть слияние сенсорных данных

[2,17,18].

Слияние данных есть подходом к обработке данных, а под задачей слияния данных (data fusion), понимают технологию, благодаря которой большое количество различных данных может быть объединена в точное единое значение [10,12,13].

Среди известных методов сенсорного слияния данных можно выделить методы размытой логики [12], размытые меры с треугольной и Гаусовской формой распределения погрешностей сигнала [13], Байесовская статистика [7], искусственные нейронные сети [1], многовариантный анализ [11].

Для решения задачи сенсорного слияния данных, в методах статистической обработки информации используется множество алгоритмов [2,9,14,15,18], что позволяет осуществить такое объединение в точечной оценке (фиксированный момент времени). Тем не менее, эти алгоритмы имеют существенный недостаток, который состоит в сложности прогнозирования поведения исследуемого объекта. В работах [2,18]

Саченко Анатолий Алексеевич. Д.т.н., профессор, зав. каф. "Информационно-вычислительных систем и управления", директор Института компьютерных информационных технологий Тернопольской академии народного хозяйства.

Кочан Владимир Владимирович. К.т.н., доцент каф. Специализированных компьютерных систем Тернопольской академии народного хозяйства.

Коваль Василий Сергеевич. Преподаватель каф. "Информационно-вычислительных систем и управления" института компьютерных информационных технологий Тернопольской академии народного хозяйства

показано, что алгоритм, предложенный D.Dolev (DOL) [9] является правильным лишь в том случае, когда значение наибольшего и (или) наименьшего элемента выборки показов сенсоров являются грубыми погрешностями. Алгоритм S.Mahaneу и F.Schneider (MS) [15] в своей работе, в отличие от предыдущего, использует интервалы допустимых значений показов сенсоров (входных данных), а также не проводит отсева грубых отклонений от среднего значения показов сенсоров (аномальных значений). Поэтому, как показано в [2,18], результат выполнения данного алгоритма содержит большие погрешности, чем в предыдущем алгоритме. Смешанный алгоритм оптимальных областей R.B.Brooks и S.S.Iyengar (BI) [8] использует интервалы допустимых значений показов сенсоров и учитывает проблему “Византийского согласования” [9]. Работа алгоритма базируется на средней взвешенной величине. То есть показы считаются тем правильнее, чем с большим количеством пересекаются его значения с интервалами показов других сенсоров. Предложенный в [2,18] алгоритм среднеквадратических весовых коэффициентов (MSWF) учитывает недостатки и преимущества вышеприведенных алгоритмов: осуществляет отсев аномальных значений показов сенсоров, использует взвешенную оценку входных показов, а также учитывает проблему “Византийского согласования”, что позволяет обрабатывать информацию с наименьшей погрешностью. Однако алгоритм MSWF требует более сложной обработки данных.

Дальнейшее усовершенствование алгоритмов слияния данных возможно за счет учета особенностей дальнейшей обработки сигналов. Если далее предусмотрено временное объединение данных, то есть возможность одновременно резко улучшить эффективность слияния. Критерием эффективности можно считать значение среднеквадратической погрешности [8]. Для слияния данных с их временным объединением предлагается алгоритм, базирующийся на известном методе наименьших квадратов (МНК) [3].

АЛГОРИТМ СЛИЯНИЯ ДАННЫХ НА ОСНОВЕ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Для анализа предложенного алгоритма, который базируется на МНК, примем, что существует n показов сенсоров – значений y_i полученных в разные моменты времени i . Согласно регрессионного анализа, МНК состоит в том, чтобы зная показы сенсора, провести такую линию регрессии, чтобы сумма квадратов отклонений Δ_i^2 между показаниями сенсора y_i и значениями проведенной прямой \bar{y}_i была минимальной (1).

$$\min \rightarrow \sum_{i=1}^n \Delta_i^2 = \sum_{i=1}^n (y_i - \bar{y}_i)^2, \quad (1)$$

где y_i – значение показаний сенсора; \bar{y}_i – полученное значение функции в момент времени – i , которое рассчитывается согласно (2); n – объем выборки показов сенсора.

$$\bar{y}_i = a + b \cdot x_i, \quad (2)$$

где a, b – искомые константы модели (коэффициенты модели), x_i – переменная времени в момент i .

Коэффициенты модели a, b рассчитываются согласно (3) и (4) [5]

$$a = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad (3)$$

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}. \quad (4)$$

На основе регрессии, для показов каждого сенсора, формируется результирующее значение показов модели, как усреднение существующих моделей регрессии (5)

$$y^{REZ} = \frac{\sum_{j=1}^m \bar{y}_{ij}}{m}, \quad (5)$$

где y_i^{REZ} – результирующее значение в i -й момент времени, \bar{y}_{ij} – значение j – регрессии в момент времени i .

Заметим, что уравнение прямой есть одним из возможного набора кривых для выравнивания выборок, поэтому она не всегда будет самой лучшей [3,4,5]. Её часто выбирают через простоту аппроксимации. В данном случае большое количество уравнений кривых, после некоторых преобразований представляют линейной зависимостью (процесс линеаризации). При этом, МНК и процедуры подбора прямой регрессии полностью переносятся на случай линеаризации [5]. При решении алгоритма, возникает задача в выборе уравнения кривой. Решением такой задачи может стать предложенный в [5] подход, когда формируется базовый набор кривых, которые сводятся к линейному тренду. На следующем шаге формируются регрессии за МНК для каждой кривой базового набора, после чего осуществляется расчет критерия выбора регрессии. К критериям выбора регрессии [5] относят: коэффициент детерминации (6)

$$\max \rightarrow r^2 = \frac{\sigma_{per}^2}{\sigma_{заг}^2} = \frac{a \sum_{i=1}^n y_i + b \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n y_i^2}{n}}{\sum_{i=1}^n y_i - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n}}, \quad (6)$$

который отображает часть общей дисперсии, которая объясняет регрессию; среднеквадратическую погрешность

$$\min \rightarrow \sigma = \frac{1}{n} \sum_{i=1}^n e_i^2, \quad (7)$$

где e_i^2 – погрешность в i -й момент времени; и средне абсолютную процентную погрешность.

Таблица 1

Показы сигналов, которые принимают сенсоры (равномерный закон распределения)

Действительный сигнал	x=	0	1	2	3	4	5	6	7	8	9	10
	y=	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
Показания, которые принимают сенсоры	y1=	-0,59440	2,84600	0,63510	-4,56820	-1,02580	1,57450	1,23110	-3,02350	1,52250	6,42350	2,34680
	y2=	-2,41430	3,02654	-2,58150	0,39960	-0,46610	-0,27570	4,91387	3,06090	0,96720	5,62390	7,32460
	y3=	-5,62350	-1,01070	-2,23020	1,21130	3,56480	-3,26590	2,75599	0,47480	-2,53150	1,63070	1,31090
	y4=	-0,78190	-2,65480	-6,23150	-0,60810	-0,26870	0,23630	1,37780	4,52360	1,48590	2,00110	-1,32450
	y5=	2,65420	-2,74610	0,80810	0,71120	4,25360	4,86950	1,23000	1,93090	-1,86520	1,25170	4,97796
	y6=	-2,56280	-2,62080	3,25340	-5,06235	1,46540	-0,62150	5,62563	-0,16610	2,02390	1,91250	2,86600

Таблица 2

Результирующие показания работы алгоритмов и их абсолютные отклонения (равномерный закон распределения)

x=	0	1	2	3	4	5	6	7	8	9	10
y=	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
МНК	-1,47480	-1,04294	-0,61108	-0,17921	0,25265	0,68451	1,11637	1,54823	1,98010	2,41196	2,84382
Отклонения	-0,52520	-0,45706	-0,38892	-0,32079	-0,25265	-0,18451	-0,11637	-0,04823	0,01990	0,08804	0,15618
D.Dolev	-1,58835	-0,86008	-0,84243	0,76036	1,03635	0,22840	2,56969	1,32580	0,52760	2,79205	2,87744
Отклонения	-0,41165	-0,63993	-0,15758	-1,26036	-1,03635	0,27160	-1,56969	0,17420	1,47240	-0,29205	0,12256
Mahaney	-1,55378	-0,52664	-1,05797	0,95290	1,19606	0,41953	2,85573	1,11038	0,30298	3,14057	2,91831
Отклонения	-0,44622	-0,97336	0,05797	-1,45290	-1,19606	0,08047	-1,85573	0,38962	1,69702	-0,64057	0,08169
Brooks	-1,48465	0,14022	-1,44891	1,33798	1,51548	0,80180	3,42781	0,67953	-0,14625	3,83760	3,00050
Отклонения	-0,51535	-1,64022	0,44891	-1,83798	-1,51548	-0,30180	-2,42781	0,82047	2,14625	-1,33760	-0,00050
MSWF	-1,51991	-1,09398	-1,24025	0,52560	1,38344	-0,27479	2,74715	0,99583	0,97786	2,23645	2,86140

$$\min \rightarrow \Delta_{abc} = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{y_i} \cdot 100, \quad (8)$$

При выборе критерия выходят из максимального коэффициента детерминации, минимального коэффициента средне-квадратичной погрешности и минимального значения средней абсолютной процентной погрешности.

Следовательно, использование таких критериев позволяет выбрать и использовать наилучшую кривую для построения регрессии за МНК в предложенном алгоритме [5].

СРАВНИТЕЛЬНАЯ ОЦЕНКА АЛГОРИТМОВ СЛИЯНИЯ ДАННЫХ

Рассмотренные алгоритмы исследовались на данных, имитирующих результаты измерения по шести каналам. Данные объединяются на интервале [0,10]. Действительные значения результатов описываются прямой $y=0.5 \cdot x-2$. Данные зашумлены погрешностями с равномерным (таблица 1) и нормальным законом распределения (таблица 3). Максимальная погрешность разброса равна двум относительно действительных значений результатов измерения. Кроме погрешности разброса по равномерному и нормальному законам, смоделированы два аномальных значения результатов измерения (робастности), которые могут возникать по разным причинам (неисправностями устройства, большой амплитудой шума и т.п.). Результаты измерения по каждому каналу, а также значения действительного сигнала представлены в таблицах 1, 3. В таблицах 2, 4 отображены результаты работы вышеупомя-

нутых алгоритмов и абсолютные отклонения полученных результатов от действительного сигнала (соответственно при равномерном и нормальном законах распределения). Графическое отображение результатов работы алгоритмов показано на рисунках 1, 2. На основании данных, представленных в таблицах 2, 4, можно сделать вывод, что применение алгоритма с использованием метода наименьших квадратов показало наименьшее отклонение от действительного сигнала в сравнении с известными алгоритмами. При равномерном законе распределения, среднеквадратичная погрешность на интервале значений от [0,10] для алгоритма МНК составляет лишь 0,080471, тогда как за алгоритмом DOL она составляет 0,736602, за алгоритмом MS – 1,053659, за смешанным алгоритмом BI – 1,989669, за алгоритмом MSWF – 0,769026. Такие результаты свидетельствуют о том, что предложенный алгоритм в 9 раз лучше за алгоритм DOL и MSWF, в 13 раз лучший за MS алгоритм и в 24 раза лучшие за BI. При нормальном законе распределения, среднеквадратичная погрешность (таблица 4) за алгоритмом МНК составляет 0,006166, за DOL – 0,95723, за MS – 0,302545, за BI – 2,291698, за алгоритмом MSWF – 0,526526.

ВЫВОДЫ

Предложен алгоритм слияния сенсорных данных (при допустимом их временном слиянии) на основе метода наименьших квадратов, который позволяет обрабатывать данные в среднем в 14 раз лучше в сравнении с известными алгоритмами при равномерном законе распределения и в 205 раз при нормальном законе распределения.

Показы сигналов, которые принимают сенсоры (нормальный закон распределения)

Действительный сигнал	x=	0	1	2	3	4	5	6	7	8	9	10
	y=	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
Показания, которые принимают сенсоры	y1=	-1,14935	-2,56825	3,26530	4,25630	-0,73190	0,78402	1,29778	-1,95360	3,00618	3,81727	-0,95621
	y2=	-6,32640	-1,63190	-0,29223	-1,51302	-0,79263	0,45765	5,23160	0,88164	-1,32600	2,99811	3,14834
	y3=	-3,14285	4,32265	-1,12950	-1,58035	-4,23560	3,36250	1,68890	0,88930	5,52300	3,47532	3,99598
	y4=	-2,48850	-1,31876	-0,86839	3,63120	0,15310	-3,23560	0,63240	0,77164	1,80480	7,21340	4,12032
	y5=	-5,69820	3,52346	-1,52862	0,34472	-2,56480	0,83317	0,00527	-2,31230	2,49640	1,96227	2,34753
	y6=	-3,28810	-2,66152	-3,65482	0,57853	-0,93838	-0,47790	4,53260	2,24420	0,99920	0,13450	0,02310

Таблица 4

Результирующие показания работы алгоритмов и их абсолютные отклонения (нормальный закон распределения)

x=	0	1	2	3	4	5	6	7	8	9	10
y=	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
МНК	-1,93234	-1,45378	-0,97523	-0,49668	-0,01813	0,46042	0,93898	1,41753	1,89608	2,37463	2,85318
Отклонения	-0,06766	-0,04622	-0,02477	-0,00332	0,01813	0,03958	0,06102	0,08247	0,10392	0,12537	0,14682
D.Dolev	-3,65441	-0,49885	-0,95469	0,76036	-1,25693	0,55853	2,03792	0,14724	2,07663	3,06324	2,37874
Отклонения	1,65441	-1,00115	-0,04532	-1,26036	1,25693	-0,05853	-1,03792	1,35276	-0,07663	-0,56324	0,62127
Mahaney	-3,68223	-0,05571	-0,70096	0,95290	-1,51837	0,39351	2,23142	0,08681	2,08392	3,26681	2,11318
Отклонения	1,68223	-1,44429	-0,29904	-1,45290	1,51837	0,10649	-1,23142	1,41319	-0,08392	-0,76681	0,88682
Brooks	-3,73788	0,83057	-0,19351	1,33798	-2,04125	0,06345	2,61844	-0,03405	2,09850	3,67395	1,58206
Отклонения	1,73788	-2,33057	-0,80649	-1,83798	2,04125	0,43655	-1,61844	1,53405	-0,09850	-1,17395	1,41795
MSWF	-3,25518	-1,21600	-1,13038	0,52560	-0,79465	0,68145	1,63906	0,68393	2,03620	3,47478	2,38600
Отклонения	1,25518	-0,28400	0,13038	-1,02560	0,79465	-0,18145	-0,63906	0,81607	-0,03620	-0,97478	0,61400

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- Головкин В.А., Нейроинтеллект: теория и применение. Книга 1. Организация и обучение нейронных сетей с прямыми и обратными связями, Брест Изд.БПИ, 1999, 264с.
- Коваль В.С., Саченко А.О., Турченко В.О., Алгоритмы объединения данных в интеллектуальных дистрибутивных сенсорных сетях, 9-th NATIONAL SCIENTIFIC SYMPOSIUM with International Participation "METROLOGY AND METROLOGY ASSURANCE'99", Sept. 14-16, 1999, Sozopol, Bulgaria, pp.183-191 (with ISBN 954-438-229-2).
- Кузин Л.Т., Основы кибернетики, Т.1. Математические основы кибернетики. Учеб. пособие для студентов вузов. М. "Энергия", 1973, 504с.
- Львовский Е.Н., Статистические методы построения эмпирических формул: Учеб. пособие для вузов. – 2-у изд., перераб. и доп. – М.: Высш. шк., 1988, - 239с.
- Льюис К.Д., Методы прогнозирования экономических показателей, Пер. с англ. И предисл. Е.З.Демиденко. – М.: Финансы и статистика, 1986. – 133с.
- Новицкий П.В., Зограф И.А., Оценка погрешностей результатов измерений, - Л.: Энергоатомиздат, 1985, -248с.
- Vox G. and Tiao G., "Bayesian Inference in Statistical Analysis", Addison – Wesley, California, 1973
- Brooks R., Iyengar S., Robust Distributed Computing and Sensing Algorithm, Computer, June 1996, pp. 53-60.
- Dolev D., The Byzantine Generals Strike Again, J. Algorithms, 1982, pp.14-30
- Gavin Kally, Data fusion: from primary metrology to process measurement, Proceeding of the 16 IEEE Instrumentation and Measurement Technology Conference, Venice, Italy, May 24-26, 1999.
- Johnsson R.A. and Wichern D.W., "Applied Multivariate Statistical Analysis, Prentice-Hall International Editions, New Jersey, 1992.
- Karlson B., Fuzzy Measures for Sensor Data Fusion in industrial Recycling, Measurement Science and Technology, in press.
- Karlson B., Jarrhed J. and Wide B., A Fusion Toolbox for Sensor Data Fusion in Industrial Recycling, IEEE Instrumentation and Measurement Technology Conference, Venice, Italy, May 24-26, 1999.
- Lamport S., Shustak R., and Pease M., The Byzantine Generals Problem, ACM Trans. Program. Lang. Syst., July 1982, pp. 382-401.
- Mahaney S., Schneider F., Reaching Approximate Agreement in the Presence of Faults, J. ACM, July 1986, pp 516.
- Marzullo K. "Tolerating Failures of Continuous-Valued Sensors," ACM Trans. Computer Systems, Nov. 1990, pp. 284-304.
- Sachenko A., Kochan V., Turchenko V. "Intelligent distributed sensor network", Proceeding of the 16 IEEE Instrumentation and Measurement Technology Conference, Venice, Italy, St, Paul, USA – May 18-20, 1998.
- Sachenko A., Kochan V., Turchenko V., Koval V., Data Fusion Algorithm in Intelligent Distributed Sensor Networks, - International Conference on Neural Networks and Artificial Intelligence ICNNAI'99, 12-15 October 1999, Brest, Belarus.

