

Жгун Т. В., Чижова Е. А.

**КОМПЛЕКСНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА СИСТЕМЫ КАК
РЕШЕНИЕ ЗАДАЧИ ВЫДЕЛЕНИЯ ПОЛЕЗНОГО СИГНАЛА ИЗ
ЗАШУМЛЕННЫХ ДАННЫХ**

Татьяна Валентиновна Жгун

Кандидат физико-математических наук, доцент

Tatyana.Zhgun@novsu.ru

Екатерина Александровна Чижова

студент

katja001002@gmail.com

ФГАОУ ВО «Новгородский государственный университет имени Ярослава

Мудрого», Россия, Великий Новгород

**INTEGRATED INDICATORS OF THE QUALITY OF THE SYSTEM AS
THE SOLUTION TO THE PROBLEM OF SEPARATION OF USEFUL
SIGNAL FROM NOISY DATA**

Tatyana Valentinovna Zhgun

Ekaterina Alexandrovna Chizhova

Novgorod State University a. Yaroslav the Wise, Russia, Veliky Novgorod

***Аннотация.** Рассматривается устойчивая модификация метода главных компонент, использующая количественную оценку шума входных данных — отношение сигнал/шум, при определении структуры главных компонент и главных факторов для разных наблюдений.*

***Abstract.** A stable modification of the principal component method is considered, which uses a quantitative estimation of the input data noise – the signal-to-noise ratio in determining the structure of the principal components and the principal factors for different observations.*

Ключевые слова: интегральные индикаторы, ошибки данных, метод главных компонент, отношение сигнал/шум

Keywords: integral indicators, data errors, principal component method, signal-to-noise ratio.

Задача обнаружение сигналов с заданными допустимыми вероятностями ошибочных решений, обусловленных помехами, в той или иной форме присуща многим информационным системам. Эта задача является одной из основных проблем во многих технических системах: в радиолокации, астрономии, оптической связи, навигации, телевизионной автоматике и пр. и решение её обеспечивается аппаратным образом. При исследовании физических явлений, технических систем и при решении проблем управления социально-экономическими системам возникает проблема определение комплексного показателя сложных систем. При описании стохастических динамических систем в задачах гидродинамики, магнитной гидродинамики, астрофизике, физике плазмы, радиофизике интегральные величины, характеризующие такие системы, являются их основными характеристиками. Например, все законы сохранения в механике и электродинамике сплошных сред записываются для интегральных величин. Интегральные характеристики описывают динамику в целом, позволяя отвлечься от побочных эффектов, связанных со случайностью показателей, искаженных помехами и являются ключом к пониманию структурообразования в стохастических динамических системах [1].

Построение интегрального индикатора вводит отношения порядка на многомерном множестве объектов и позволяет сравнивать качество объектов. Международные организации, аналитические центры и социальные науки на рубеже тысячелетий существенно увеличили количество применяемых интегральных индикаторов для измерения разнообразных латентных характеристик социально-экономических систем: социального капитала, человеческого развития, качества жизни, качества управления и др. Согласно данным ООН, к 2011 году существовало 290 индексов, разработанных для ранжирования или

комплексной оценки стран [2, 3]. Обсуждение плюсов и минусов составных показателей приводится в работе [3].

Статистические данные, описывающие социальные системы в разные моменты времени, не только отражают динамику рассматриваемого качества системы (т.е. изменение полезного сигнала), но и содержат случайные компоненты (шум). Отсутствие полной информации о свойствах полезных сигналов и помех является существенной особенностью задачи построение интегральной характеристики качества социальной системы на основании статистических данных, что дает основание отнести рассматриваемую проблему к классу задач выделения сигналов в условиях априорной неопределенности.

Рассмотрим построение интегральной оценки системы из m объектов, для которой известны таблицы описаний объектов для ряда наблюдений — матрицы размерностью $m \times n$ $A^t = \{a_{ij}^t\}_{i,j=1}^{n,m}$, $t = 1, \dots, T$. Элемент матрицы a_{ij}^t — значение j -го показателя i -го объекта, вектор $a_i^t = (a_{i1}^t, \dots, a_{in}^t)$ — описание i -го объекта в момент t . Для каждого момента t вектор интегральных показателей имеет вид

$$(1) \quad q^t = A^t \cdot w^t, \quad q_i^t = \sum_{j=1:n} w_j^t \cdot a_{ij}^t.$$

где $q^t = \langle q^t_1, q^t_2, \dots, q^t_m \rangle^T$ — вектор интегральных индикаторов момента t , $w^t = \langle w^t_1, w^t_2, \dots, w^t_m \rangle^T$ — вектор весов показателей для момента t , A^t — матрица преобразованных данных для момента t . Численные характеристики системы предварительно подвергнуты унификации — приведению значений переменных на отрезок $[0, 1]$ по принципу: «чем больше, тем лучше». Значимость регистрируемых показателей для функционирования системы в общем случае неизвестна. Для решения задачи управления требуется дать мотивированную интегральную оценку каждого наблюдаемого объекта на всем промежутке наблюдений. Для построения искомого интегрального индикатора ка-

чества системы требуется найти веса показателей w^t для каждого момента времени, адекватно отражающие свойства рассматриваемой системы. Т. е. определяемые весовые коэффициенты должны отражать **структуру** оцениваемой системы.

Одним из наиболее простых методов анализа структуры исследуемой системы является метод главных компонент (МГК). Пространство главных компонент оптимально для моделирования внутренней структуры данных. Среди преимуществ метода главных компонент для анализа потока данных (рентгеновских изображений) авторы [4] отмечают высокую эффективность метода для задачи фильтрации шумов и для поиска наиболее характерных особенностей в данных. Успешное применение МГК для описания структуры систем самого различного вида позволяет предположить, что и для описания социальных систем метод будет тоже давать адекватные результаты.

Принципиальным отличием определения композитных индексов для социально-экономических систем является неопределенность качества используемых данных, в отличие от характеристик технических и физических систем, для которых заранее известна погрешность измерения. Тем не менее, именно статистические данные, содержащие неустранимые погрешности, в настоящее время представляют собой наилучшие оценки имеющихся реальных величин в социальных системах [3].

Получение точных характеристик объекта (и весов композитных индексов в том числе) на основании однократного измерения, неизбежно содержащего неизвестную погрешность, не представляется возможным. Однако по серии таких измерений вычисление неизвестной характеристики вполне вероятно. Такую задачу, в частности, с успехом решает астрофотометрия, определяющая основные числовые параметры астрономических объектов не по однократному наблюдению (изображению), а по **серии** зашумленных изображений.

Задача отыскания наилучшего из способов распознавания сигнала при наличии помех является одной из основных и сложных проблем в измерительной технике, радиолокации, астрономии, оптической связи, локации, навигации, телевизионной автоматике и многих других областях науки и техники. Идеальный прием сигналов в условиях воздействия шума и помех основывается на простых и глубоких идеях, изложенных в наиболее последовательной и ясной форме Ф. М. Вудвордом [5]. Следуя этой идее, задачей идеального устройства, на вход которого поступает смесь сигнала с шумом, является полное разрушение ненужной информации, содержащейся в смеси, и сохранение полезной информации о тех параметрах сигнала, которые представляют интерес для пользователя системы. В любой коммуникационной системе наиболее важным параметром, характеризующим уровень помех, является отношение уровня сигнала к уровню шума (ОСШ). Это величина наиболее полно описывает качество воспроизведения сигнала в телевизионных системах, в системах мобильной связи, в астрофотометрии. Приемлемое качество воспроизводимого сигнала определяется задаваемым стандартом пороговым значением отношением сигнал/шум.

Используя основные идеи, лежащие в основе астрофотометрии, можно рассматривать построение интегральной характеристики изменения качества сложной системы как решение задачи выделения полезного сигнала в многомерном массиве зашумленных данных по серии наблюдений в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум. Алгоритм приводится в [6]. Определяемым выходным сигналом в этом случае являются весовые коэффициенты линейной свертки (1). Эта задача аналогична задаче восстановления цифровых изображений, искаженных гауссовским шумом. Метод главных компонент (МГК) позволяет выделить структуру в зашумленном массиве данных и, в частности, с успехом применяется для шумоподавления при распознавании изображений.

МГК для разных наблюдений должен описывать неизменную структуру системы. Следовательно, свободные от влияний ошибок значения собственных чисел и собственных векторов будут тем сигналом, который нужно по имеющимся реализациям выделить из зашумленных данных и который характеризует структуру системы. Простейшим способом фильтрации ошибок является усреднение значений, которое работает на основе предположения об абсолютно случайной природе шума и дает стабильные результаты для технических систем. Именно усреднение значений используется в астрофотографии для подавления шумов в изображении. Для собственных чисел выделяемый сигнал даст усредненное значение по всем наблюдениям — эмпирические собственными числа. Но собственные векторы в МГК определяются с точностью до направления и среднее значение факторных нагрузок переменных зависит от выбранного направления. На основе вычисленных собственных векторов (упорядоченных в порядке убывания собственных чисел), нужно согласовать их направления и распознать случайные и неслучайные компоненты этих векторов.

Таблица 1 — Определение эмпирической главной компоненты

Согласование знаков 1-ой, 2-ой и 6-ой переменной									
Год	Переменные								
	1	2	3	4	5	6	7	8	9
2012	0.19	-0.11	-0.36	-0.16	0.22	-0.43	-0.11	0.74	-0.01
2013	0.19	-0.06	-0.12	0.15	-0.10	-0.70	0.33	0.56	0.01
2014	0.22	-0.02	0.25	0.39	0.25	-0.42	0.42	-0.49	-0.28
2015	0.43	-0.09	0.02	0.36	0.23	-0.67	0.41	0.03	-0.12
2016	0.30	-0.04	0.09	0.35	0.31	-0.43	0.43	-0.42	-0.37
Среднее, m	0.27	-0.06	-0.02	0.22	0.18	-0.53	0.30	0.09	-0.15
Среднеквадр. отклонение, s	0.10	0.04	0.23	0.23	0.16	0.14	0.23	0.56	0.17
ОСШ	2.63	1.73	0.11	0.94	1.14	3.84	1.27	0.15	0.90
Сумма ОСШ по строке									12.72
Сумма действующих ОСШ									6.47
Согласование знаков 3-ей переменной									
Год	Переменные								
	1	2	3	4	5	6	7	8	9
2012	-0.19	0.11	0.36	0.16	-0.22	0.43	0.11	-0.74	0.01
2013	-0.19	0.06	0.12	-0.15	0.10	0.70	-0.33	-0.56	-0.01
2014	0.22	-0.02	0.25	0.39	0.25	-0.42	0.42	-0.49	-0.28
2015	0.43	-0.09	0.02	0.36	0.23	-0.67	0.41	0.03	-0.12
2016	0.30	-0.04	0.09	0.35	0.31	-0.43	0.43	-0.42	-0.37
Среднее, m	0.11	0.00	0.17	0.22	0.13	-0.08	0.21	-0.44	-0.15

Среднеквадр. отклонение, s	0.29	0.08	0.14	0.23	0.21	0.60	0.33	0.29	0.17
ОСШ	0.40	0.04	1.23	0.98	0.63	0.13	0.63	1.52	0.93
Сумма ОСШ по строке									6.48
Сумма действующих ОСШ									0

Наличием неслучайного (значимого) вклада переменной в структуру главных компонент будем считать не большую величину факторной нагрузки после вращения, а **инвариантность** факторной нагрузки при возмущениях, признаком которой будет вычисляемое отношение сигнал/шум, определяемое отношением среднего значения факторной нагрузки (сигнал) к среднеквадратичному отклонению (шум). Если это отношение выше порогового значения, то такую переменную считаем неслучайной. Иначе переменная характеризует шумовую составляющую сигнала и не участвует в дальнейшем рассмотрении. Критерием выбора направления собственных векторов будет максимизация уровня сигнала — суммы вычисленных значений ОСШ у значимых переменных. Значимые переменные определенной таким образом эмпирической главной компоненты (ЭГК), как и в факторном анализе, будут участвовать в дальнейшем рассмотрении, а незначимые переменные обнуляются. В таблице 1 приведен фрагмент определения значимых нагрузок ЭГК по пяти наблюдениям. Первая и шестая переменные оказались значимыми (вычисленное значение ОСШ выше выбранного порогового значения 2,2). Значение нагрузок у этих переменных без учета нормировки по величине эмпирического собственного числа будут 0,27 и $-0,53$ соответственно.

В работе рассмотрено решение задачи построения латентной интегральной характеристики изменения качества системы для ряда наблюдений. Задача относится к классу задач выделения полезного сигнала из массива данных в условиях априорной неопределенности и решается на основании количественной характеристики уровня ошибок в данных — отношения сигнал/шум. Предложен метод построения интегральных характеристик с определением неслучайных составляющих главных компонент, характеризующих структуру рассматриваемой системы. Алгоритм был использован для получения инте-

гральных индикаторов качества жизни населения субъектов Российской Федерации за 2007–2016 гг. и показал хорошее качество полученных результатов [7]. Предлагаемая методика может быть использована для вычисления интегральных оценок изменения качества плохо формализуемых систем.

Список литературы

1. Кляцкин, В. И. Интегральные характеристики – ключ к пониманию структурообразования в стохастических динамических системах / В. И. Кляцкин // Успехи физических наук. – 2011. – Т. 181. – № 5. – С. 457–482.
2. Bandura, R. Composite Indicators and Rankings: Inventory 2011 [Electronic resource] / Bandura R. – Working Paper. Abbreviated version. – 2011. – 257 p. – URL: <http://nebula.wsimg.com/beb0f5b0e9b6f0c298ae72e7519b22a4> (дата обращения: 01.11.2018)
3. Nardo, M. Tools for composite indicators building [Electronic resource] / Nardo M., Saisana M., Saltelli A., Tarantola S. Joint. – Research Centre, Ispra, Italy, 2005. – P. 134.
4. Теслюк, А. Б. Применение метода главных компонент к анализу дифракционных изображений биомолекулярных объектов / А. Б. Теслюк, Р. А. Сенин, В. А. Ильин // Математическая биология и биоинформатика. – 2013. – Т. 8. – № 2. – С. 708–715.
5. Вудворд, Ф. М. Теория вероятностей и теория информации с применениями в радиолокации / Ф. М. Вудворд. – Москва : Советское радио, 1955. – 128 с.
6. Жгун, Т. В. Алгоритм построения интегрального индикатора качества сложной системы для ряда последовательных наблюдений / Т. В. Жгун // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. – 2019. – Т. 6. – № 1. – С. 5–25.
7. Жгун, Т. В. Построение интегральной характеристики качества жизни субъектов Российской Федерации с помощью метода главных компонент / Т. В. Жгун // Экономические и социальные перемены: факты, тенденции, прогноз. – 2017. – Т. 10. – № 2. – С. 214–235.