

АЛГОРИТМ ПОСТРОЕНИЯ ИНТЕГРАЛЬНОГО ИНДИКАТОРА КАЧЕСТВА СЛОЖНОЙ СИСТЕМЫ ДЛЯ РЯДА ПОСЛЕДОВАТЕЛЬНЫХ НАБЛЮДЕНИЙ

© 2017 г. Т.В. Жгун

Новгородский государственный университет имени Ярослава Мудрого
(173003 Великий Новгород, ул. Большая Санкт-Петербургская, д. 41),

E-mail: Tatyana.Zhgyn@novsu.ru

Поступила в редакцию: 26.10.2016

В статье предлагается алгоритм построения интегральной характеристики изменения качества системы на основании регистрируемых измерений, который обеспечивает решение задачи выделения сигнала в многомерном массиве данных в условиях априорной неопределенности о свойствах сигнала на основании задаваемого отношения сигнал/шум. Построение латентной интегральной характеристики изменения качества системы на основе статистических показателей для ряда последовательных наблюдений производится на основе метода главных компонент с учетом наличия шума в измеряемых данных (ОСП-алгоритм). В отличие от классического метода главных компонент, где информативность вычисленной интегральной характеристики задается априорно и обеспечивается выбором числа главных компонент, в предлагаемом алгоритме информативность решения оценивается апостериорно на основании дисперсионного критерия и выбранного параметра отношения сигнал/шум. С помощью предложенного алгоритма построены интегральные индикаторы качества жизни субъектов Российской Федерации за 2007–2014 годы.

Ключевые слова: качество системы управления, интегральная характеристика качества, изменение характеристики качества, интегральные индикаторы качества жизни, шум измеряемых данных, отношение сигнал/шум, метод главных компонент, информативность метода главных компонент.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Жгун Т.В. Алгоритм построения интегрального индикатора качества сложной системы для ряда последовательных наблюдений // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6, № 1. С. 5–25. DOI: 10.14529/cmse170101.

Введение

Интегральные характеристики качества системы используются при решении задач управления для совершенствования механизмов управления на основании сопоставления интегральных индикаторов рассматриваемых объектов и рейтингов объектов, определяемых их интегральными характеристиками, и являются показателем степени достижения цели управления. Согласно данным ООН, к 2011 году существовало 290 индексов, разработанных для ранжирования или комплексной оценки стран в соответствии с некими экономическими, политическими, социальными или экологическими критериями. Обзор феноменального роста количества применяемых для комплексной оценки стран сводных индексов, проведенного Бандура [13–15], показал, что только 26 из них (9%) было сформировано до 1991 года, 132 (46%) — с 1991 года по 2006 год, а в период 2007–2011 годы было сформировано 130 сводных индексов.

Стремительное увеличение числа композитных индексов является явным признаком их важности в разработке государственной политики и оперативной значимости в макроэкономике в целом. Все основные международные организации, такие как Организация экономического сотрудничества и развития (OECD), Европейский Союз, Всемирный экономический форум или Международный Валютный Фонд конструируют составные

показатели в самых различных областях [15, 23]. Общей целью большинства из этих показателей является ранжирование стран и их сравнительный анализ по некоторой агрегированной мере [15, 16, 26, 27]. Возможность использовать в политических целях единый показатель, обобщающий сложные, плохо формализуемые процессы (качество жизни, финансовую устойчивость, рыночную политику и т. д.) является сильным искушением для заинтересованных сторон. Следовательно, улучшение методик построения таких показателей является очень важным вопросом исследования и с теоретической, и с практической точек зрения.

Создатели композитных индексов считают, что сводная статистика действительно адекватно описывает реальность, и такое выделение итога является чрезвычайно полезным в обеспечении интереса средств массовой информации и внимания политиков. С точки зрения противников составных характеристик, достаточно сформировать соответствующий набор показателей, и не имеет смысла идти далее, к получению интегральной характеристики. Их основными возражениями является произвольный характер определения весов, с помощью которых переменные объединяются в единую характеристику. Обсуждение плюсов и минусов составных показателей приводится в работах [16, 23].

Организация экономического сотрудничества и развития (*OECD*) и развития ведет непрерывную работу по совершенствованию методик построения композитных индексов [23–26]. В 2008 году *OECD* совместно с Объединенным исследовательским центром (*Joint Research Centre European Commission*) с подготовлен справочник [17], явившийся итогом многолетних исследований в этой области [23–26, 29], где для широкого круга заинтересованных лиц изложен набор технических принципов, которые должны помочь исследователям в формировании составных показателей. Основным методом агрегации данных авторы выбирают линейную свертку показателей, а основным инструментом построения сводных показателей — факторный анализ.

Для построения качественного интегрального показателя необходима, во-первых, тщательная теоретическая проработка теоретических аспектов измеряемого явления, ибо «то, что плохо определено, вероятно, будет плохо измерено» [23], во-вторых, качественный выбор данных, так как качество составных индикаторов в значительной степени зависит от качества базовых показателей, и в-третьих, адекватный инструмент для работы с многомерными данными.

Факторный анализ впервые был использован для объединения показателей в единый индекс при разработке индекса здоровья в работе Хайтауэра в 1978 году [19]. При вычислении Показателей социально-экономического статуса (*Socio-Economic Status Indices, SES*) метод главных компонент был принят в качестве стандарта метода построения, где вычисляемый индекс определяла проекция на первую главную компоненту [22, 30]. Эту же методику использовали Линдман и Селин при создании Индекса экологической устойчивости [21], Сомарриба и Пена при измерении качества жизни в Европе [28]. Среди отечественных исследований следует отметить работы С.А. Айвазяна [1, 2] по определению индекса качества жизни населения.

Однако первая главная компонента хорошо приближает моделируемую ситуацию, если максимальное собственное число ковариационной матрицы дает вклад не менее 70% в сумму всех собственных чисел. Такое соотношение выполняется, если рассматривается небольшое количество признаков (не более пяти), и одно из свойств системы явно доминирует над остальными. При описании социально-экономических систем число пе-

ременных значительно более пяти, а структура системы не допускает простого приближения. В качестве выхода из этой ситуации рассматривается согласно [2] понижение порога информативности до 55%, и деление исходной системы на подсистемы, описываемые меньшим количеством переменных. В рассмотренных выше исследованиях [19–22, 28, 30] вклад наибольшего собственного числа колебался от 13% до 38%, за исключением работы [28], рассматривавшей модельный пример, где эта цифра составила 56%. Авторы следовали рекомендациям [28], утверждавшим, что первая главная компонента дает удовлетворительные весовые показатели даже в тех случаях, когда наибольшее собственное число вносит небольшой вклад в сумму всех собственных чисел.

Исследователи Организации экономического сотрудничества и развития придерживаются иной точки зрения. Для формирования составного индикатора используют факторный анализ, где метод главных компонент применяется исключительно для извлечения факторов с тем, чтобы количество извлеченных факторов объясняло бы более 50% общей дисперсии. Значение композитного индекса в этом случае определяют только значимые нагрузки выбранных главных факторов после вращения. Де-факто именно эта методика становится стандартом при вычислении композитных индексов [20, 23, 24, 29]. Хотя авторы [23] обращают внимание, что различные методы извлечения основных компонент и различные способы вращения подразумевают различные значимые переменные, а значит, разные веса переменных при вычислении композита, следовательно, и различные значения вычисляемого интегрального индикатора. Кроме того, факторный анализ предполагает наличие достаточной корреляции между исходными переменными, что в некотором смысле противоречит идее полного описания исследуемого явления совокупностью независимых величин.

Следует также отметить еще одно обстоятельство. Методика определения весовых коэффициентов с помощью факторного анализа не может применяться для сравнения характеристик описываемых объектов в динамике, так как даже при фиксированных методах извлечения факторов и способа вращения, факторный анализ для разных наблюдений системы структуру главных факторов (значимые нагрузки) определяет по-разному, что делает бессмысленным межвременные сравнения [10].

Недостаточное удовлетворение исследователей качеством построенных композитных показателей основывается не только на методических проблемах построения, но и на недостаточном качестве используемых данных, так как фиксируемые статистическими организациями данные неизбежно содержат погрешности измерения. Тем не менее, именно статистические данные, содержащие неустраняемые погрешности, в настоящее время представляют собой наилучшие оценки имеющихся реальных величин в социальных системах [18, 23].

Получение точных характеристик объекта 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. Используя основные идеи, лежащие в основе астрофотометрии, будем рассматривать построение интегральной характеристики изменения качества сложной системы как решение задачи выделения полезного сигнала по серии наблюдений, содержащей описание неизвестного параметра (в

многомерном массиве зашумленных данных) в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум.

Целью статьи является разработка методики построения интегральной характеристики изменения качества сложной системы на основании серии регистрируемых измерений с помощью метода главных компонент. В разделе 1 представлена постановка задачи вычисления интегральной характеристики изменения качества системы. В разделе 2 обсуждается вопрос определения информативности решения поставленной задачи при наличии шума в измеряемом сигнале. Раздел 3 посвящен алгоритму построения интегральной характеристики изменения качества системы на основании задаваемого отношения сигнал/шум. В последнем разделе с помощью предложенного алгоритма построены интегральные индикаторы изменения качества жизни субъектов Российской Федерации за 2007–2014 годы. В заключении делается вывод, что предлагаемая методика может быть использована для вычисления интегральных оценок изменения качества плохо формализуемых систем.

1. Постановка задачи вычисления интегральной характеристики изменения качества системы

Рассмотрим построение интегральной оценки системы из m объектов, для которой известны таблицы из n описаний объектов для ряда наблюдений $t = 1, 2, \dots$ — матрицы $A^t = \{a_{ij}\}_{i,j=1}^{n,m}$. Для каждого момента t вектор интегральных показателей имеет вид

$$q^t = A^t \cdot w^t, \quad (1)$$

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

предобработанных данных для момента t . Для построения интегрального индикатора качества системы требуется найти веса показателей w^t для каждого момента времени.

Будем рассматривать оцениваемый объект как сложную (не поддающуюся удовлетворительной формализации), большую (число состояний выше современных вычислительных возможностей) систему. Такими системами являются как каждый биологический объект, так и любая социальная система. Система доступна для наблюдения, и известно конечное, достаточно большое число регистрируемых с некоторой точностью численных характеристик этой системы в разные моменты времени. Значимость регистрируемых показателей для функционирования системы в общем случае неизвестна. Для решения задачи управления требуется дать мотивированную оценку каждого наблюдаемого объекта на всем промежутке наблюдений, т.е. вычислить в динамике интегральную характеристику качества системы. Численные характеристики системы предварительно подвергнуты унификации — приведению значений переменных на отрезок $[0, 1]$ по принципу: «чем больше, тем лучше».

Любое измерение, в том числе и статистическое, неизбежно связано с точностью измерительного прибора, поэтому результат измерения неизбежно содержит неустранимую ошибку. Построение интегральной характеристики системы можно рассматривать как задачу выделения полезного сигнала на фоне шумов в условиях отсутствия априор-

ной информации о статистических характеристиках аддитивного шума и полезного сигнала при наличии единственной реализации измеряемого процесса. Эта задача аналогична задаче восстановления цифровых изображений, искаженных белым гауссовским шумом. Метод главных компонент (МГК) позволяет выделить структуру в многомерном массиве данных и с успехом применяется для распознавания изображений и для шумоподавления.

Количественные характеристики конкретной системы, функционально связанные с ее структурными особенностями и условиями ее работы, зависят от отношения сигнала к шуму. Это отношение часто используется для количественной оценки эффективности различения сигналов в опико-электронных, телевизионных автоматических системах, в средствах контроля и диагностики.

ОСШ — отношение сигнал/шум (англ. *signal-to-noise ratio*, сокр. *SNR*) представляет собой отношение сигнала (а если точнее, суммы сигнала и шума) к шуму. Величина может быть вычислена либо как безразмерное отношение амплитуды сигнала к амплитуде шума $SNR = A_s / A_n$, либо в децибелах $SNR (dB) = 20 \cdot \log_{10} (A_s / A_n)$. Эта величина наиболее полно описывает качество воспроизведения сигнала в телевизионных системах, в системах мобильной связи, в астрофотометрии.

Выбор порогового значения отношения полезного сигнала к шуму фона, позволяющего различить сигнал на фоне шума, обоснован в [8]. Современные технические системы (и человеческий глаз) уверенно выделяют сигнал из шума, если уровень ОСШ системе составляет около 7 дБ (в безразмерных единицах — 2,2). Такое пороговое значение используется в фотометрии слабых объектов: при регистрации сигнала от тусклых звезд необходимо, чтобы отношение сигнал/шум превышало 2,2. Телевизионные системы могут восстановить из шума неповрежденную картинку, если отношение уровня сигнала к уровню шумов превосходит пороговое значение (SNR_{thr}) около 7дБ. Именно это значение $SNR_{thr} = 2,2$ и будет использоваться далее.

Статистические данные неизбежно содержат погрешность измерения. Любой результат, полученный на основании этих данных, будет содержать неустранимую ошибку. Переход к другому моменту времени означает изменение данных, которое при неизменной структуре системы вызвано как изменением ситуации, так и случайными ошибками. Метод главных компонент на основании различных для разных моментов значений собственных векторов и собственных значений описывает неизменную структуру системы. Следовательно, именно значения собственных чисел и собственных векторов будут тем неизменяющимся сигналом, который нужно распознать, т.е. по имеющимся реализациям выделить сигнал из зашумленных данных. Предположение о том, что при вариации входных данных у собственных чисел имеется общая тенденция, иллюстрирует рис.1, где представлены значения упорядоченных по убыванию собственных чисел для разных наблюдений. В среднем значении рассматриваемых величин хорошо просматривается тенденция (сигнал) и случайное отклонение от нее. Именно усреднение значений используется в астрофотографии для подавления шумов. Усреднение работает на основе предположения об абсолютно случайной природе шума. Соответственно, случайные отклонения от истинных данных будут последовательно снижаться по мере усреднения возрастающего числа наблюдений.

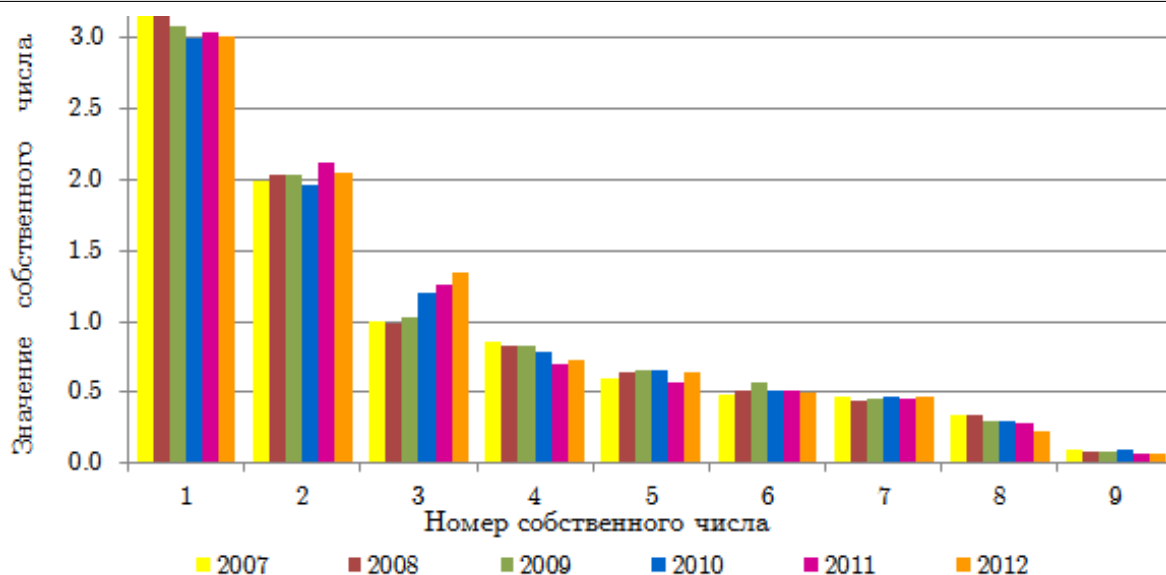


Рис. 1. Собственные значения ковариационной матрицы переменных для различных моментов наблюдения

Собственные векторы в методе главных компонент определяются с точностью до направления, в отличие от собственных чисел, определяемых однозначно. Среднее значение факторных нагрузок переменных зависит от выбранного направления и не может однозначно характеризовать сигнал. Следовательно, на основе вычисленных для различных наблюдений собственных векторов (упорядоченных в порядке убывания собственных чисел) нужно распознать случайные и неслучайные компоненты этих векторов и определить значения координат неслучайных переменных этих векторов.

Наличием неслучайного (т.е. 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 величина отношения сигнал/шум, определяемая средними значениями переменной (сигнал) и среднеквадратичным отклонением (шум). Если это отношение выше порогового значения, такую переменную считаем неслучайной — действующей. При невыполнении такого соотношения переменная характеризует шумовую составляющую сигнала и не участвует в дальнейшем рассмотрении. Для проверки этого условия нужно согласовать направления собственных векторов на основании некоторого критерия. После согласования направлений главных компонент и определения значимых переменных они, как и в факторном анализе, будут участвовать в дальнейшем рассмотрении, а незначимые переменные игнорируются (обнуляются). Модифицированные по такому принципу главные компоненты будем далее именовать эмпирическими главными компонентами (ЭГК).

2. Информативность интегральной характеристики изменения качества системы

В алгоритмах вычисления интегральной характеристики по ОСШ-методу [8–11] используется традиционное для МГК понятие информативности, которое определяет число главных компонент l , используемых для вычисления интегральной характеристики.

$$\gamma_{\sigma} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_l}{\lambda_1 + \lambda_2 + \dots + \lambda_n} \geq \theta. \quad (2)$$

Однако размерность пространства признаков в задачах вычисления интегральной характеристики качества сложной системы не является чересчур большой, и вычислительных проблем при определении собственных чисел и векторов не возникает. А качественное описание структуры системы требует либо всех главных компонент, либо их достаточно большого количества. Может оказаться, что ценная для конкретной задачи информация содержится как раз в направлениях с меньшей дисперсией. Например, при создании цифровой модели рельефа, которая строится по оцифрованным снимкам, искомым рельеф дают восьмая и девятая главные компоненты, а главные компоненты 12 и 13 в методе «Гусеница» свидетельствует о наличии в анализируемых данных периодичности с дробным периодом [5, 6].

Подходы к оценке числа главных компонент по необходимой доле объясненной дисперсии формально применимы всегда, однако неявно они предполагают, что нет разделения на «сигнал» и «шум», и любая заранее заданная точность имеет смысл. При разделии данных на полезный сигнал и шум задаваемая точность теряет смысл и требуется переопределить понятие информативности. Аналогично дисперсионной информативности согласно (2) можно определить SNR -информативность для выбранного числа эмпирических главных компонент N :

$$\gamma_{SNR} = \frac{S_{11} + S_{12} + \dots + S_{1N}}{S_{21} + S_{22} + \dots + S_{2N}}, \quad (3)$$

где S_{1k} — сумма величин ОСШ у действующих переменных k -ой ЭГК, S_{2k} — сумма ОСШ всех переменных k -ой ЭГК. Эта величина будет апостериорной оценкой (сверху) SNR -информативности. В отличие от дисперсионной информативности, SNR -информативность не может достигать 100% по логике построения. Информативность выбранной системы признаков определяется дисперсионной и SNR -информативностью:

$$\gamma = \gamma_{\sigma} \cdot \gamma_{SNR}. \quad (4)$$

3. Алгоритм построения интегральной характеристики изменения качества системы как решение задачи выделения полезного сигнала на фоне шумов

Приведем алгоритм построения интегральной характеристики изменения качества системы для ряда наблюдений на основании регистрируемых измерений, который обеспечивает решение задачи выделения сигнала в многомерном массиве данных в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум. Основным параметром алгоритма является пороговое значение отношения полезного сигнала к шуму фона. По умолчанию используется пороговое значение $SNR_{thr.} = 2,2$, применяемое в современных технических системах и позволяющее различить слабый сигнал на фоне шума. Возможно использование несколько более высокого порогового значения ОСШ, характеризующего именно слабый сигнал — до трех единиц. Если структура системы позволяет из содержательных соображений выделить подсистемы рассматриваемой системы, то переменные, описывающие систему, могут быть объединены в блоки и интегральная характеристика всей системы вычисляется через характеристики подсистем. Входными данными алгоритма являются:

- таблицы из n описаний m объектов для ряда наблюдений $t = 1, \dots, T$ — матрицы $A^t = \{a_{ij}^t\}_{i,j=1}^{n,m}$. Элемент матрицы a_{ij}^t — значение j -го показателя i -го объекта, вектор $a_i^t = (a_{i1}^t, \dots, a_{im}^t)$ — описание i -го объекта в момент t ;
- пороговое значение отношения сигнал/шум SNR_{thr} .

Алгоритм включает следующие шаги:

1 шаг. Подготовка данных

- обеспечение однородности выборки и унификация показателей.

Отсутствующие значения для отдельных наблюдений восстанавливаются интерполяцией по имеющимся данным. Устраняются те переменные, которые связаны друг с другом известной функциональной зависимостью (например, из двух переменных — количество мужчин и женщин в процентах — остается одна). Для прочих экспериментальных данных строгая коллинеарность практически не наблюдается, и переменные могут участвовать в решении задачи даже при высоких значениях парной корреляции.

Если исходный показатель связан с анализируемым интегральным свойством качества монотонной зависимостью, то при унифицировании исходных показателей (приведении их на отрезок $[0, 1]$ по принципу «чем больше, тем лучше») переменные x_{ij} для каждого момента наблюдения преобразовываются по правилу:

$$a_{ij} = s_j + (-1)^{s_j} \cdot \frac{x_{ij} - m_j}{M_j - m_j},$$

где $s_j = 0$, если оптимальное значение j -го показателя максимально и $s_j = 1$, если оптимальное значение j -го показателя минимально, m_j — наименьшее значение j -го показателя по всей выборке (глобальный минимум), M_j — наибольшее значение j -го показателя по всей выборке (глобальный максимум).

Если исходный показатель связан с анализируемым интегральным свойством качества жизни немонотонной зависимостью (т.е. внутри диапазона изменения данного показателя существует значение x_j^{opt} , при котором достигается наивысшее качество), то значение соответствующего унифицированного показателя рассчитывается по формуле:

$$a_{ij} = \left(1 - \frac{|x_{ij} - x_j^{opt}|}{\max((M_j - x_j^{opt}), (x_j^{opt} - m_j))} \right).$$

2 шаг. Определение априорных собственных векторов и собственных чисел для каждого наблюдения t

- центрирование (стандартизирование) унифицированных переменных: каждый элемент центрированной матрицы данных вычисляется по формуле:

$$z_{ij} = \frac{a_{ij} - \overline{a_j}}{s_j},$$

где $\overline{a_j} = \frac{1}{m} \sum_{i=1}^m a_{ij}$ — выборочная средняя, $s_j^2 = \frac{1}{m} \sum_{i=1}^m (x_{ij} - x_j)^2$ — выбо-

рочная дисперсия, т.е. s_j — выборочное среднеквадратичное отклонение;

- вычисление по центрированным значениям унифицированных показателей ковариационной матрицы, совпадающей в этом случае с матрицей корреляций;

– нахождение для каждого t собственных значений ковариационной (корреляционной) матрицы λ_i^t , $i = 1, \dots, n$, ранжирование их по убыванию;
 нахождение собственных векторов, соответствующих собственным числам λ_i^t для каждого наблюдения: $w^t_i = \langle w^t_{i1}, w^t_{i2}, \dots, w^t_{in} \rangle$, $t = 1, \dots, T$.

3 шаг. Определение эмпирических собственных чисел. Согласование направления собственных векторов и определение эмпирических главных компонент

- вычисление эмпирических значений собственных чисел как среднего значения собственных чисел для различных выборок $\lambda_i = \frac{1}{T} \sum_{t=1}^T \lambda_i^t$;
- согласование направлений собственных векторов (главных компонент) для разных наблюдений:

Рассматриваем по числу переменных n вариантов выбора направления главных компонент с номером i (ГК $_i$). Для k -го варианта максимизируем вычисляемое отношение сигнал/шум у k -ой переменной $SNR_{ij}(k)$ — отношение среднего значения факторных нагрузок к среднеквадратичному отклонению этих величин для всех наблюдений:

$$SNR_{ij}(k) = \frac{m_{ij}(k)}{s_{ij}(k)},$$

где $m_{ij}(k) = \frac{1}{T} \sum_{t=1}^T w^t_{ij}(k)$ — средние значения нагрузок в наборе из T i -ых главных компонент для k -го варианта, $s_{ij}^2 = \frac{1}{T} \sum_{t=1}^T (w^t_{ij}(k) - m_{ij}(k))^2$ — среднеквадратичное отклонение значения нагрузок для k -го варианта; $i, j, k = 1, \dots, n$.

Максимальное значение вычисляемого отношения сигнал/шум у k -ой переменной $SNR_{ij}(k)$ достигается, если все нагрузки k -ой переменной имеют одинаковые знаки.

Нагрузки, для которых вычисленное значение отношения сигнал/шум $SNR_{ij}(k)$ не менее порогового значения

$$SNR_{ij}(k) \geq SNR_{thr.},$$

считаем *значимыми*, а переменные с такой нагрузкой *действующими*. Сумму вычисленных значений отношения сигнал/шум у действующих переменных для варианта k запоминаем:

$$S_{1i}(k) = \sum_j SNR_{ij}(k),$$

где j — номер действующей переменной для k -го варианта выбора направления ГК $_i$.

Из n рассмотренных вариантов выбора направлений выбирается вариант k^* , для которого сумма отношений сигнал/шум у действующих переменных максимальна

$$k^* : S_{1i}(k^*) = \max_k S_{1i}(k), \quad S_{1i} = S_{1i}(k^*).$$

- определение эмпирических главных компонент $w_i = \langle w_{i1}, w_{i2}, \dots, w_{in} \rangle$, $i = 1, \dots, n$: факторные нагрузки у *недействующих* переменных обнуляются, для *действующих* переменных в качестве значения факторной нагрузки принимаем

среднее значение по всем наблюдениям (при согласовании направлений собственных векторов):

$$w_{ij} = \frac{1}{T} \sum_{t=1}^T w_{ij}^t(k^*);$$

– выбор числа эмпирических главных компонент (ЭГК) N ($N \leq n$).

Количество выбранных ЭГК N должно обеспечивать максимально возможную информативность вычисляемой интегральной характеристики γ :

$$\gamma = \gamma_{\sigma} \cdot \gamma_{SNR} \quad ,$$

где $\gamma_{SNR} = \frac{S_{11} + S_{12} + \dots + S_{1N}}{S_{21} + S_{22} + \dots + S_{2N}}$ — SNR-информативность,

$\gamma_{\sigma} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_N}{\lambda_1 + \lambda_2 + \dots + \lambda_n}$ — дисперсионная информативность.

4 шаг. Согласование направлений эмпирических главных компонент

– нормирование ЭГК умножением на квадратный корень из соответствующего эмпирического собственного числа:

$$w^*_i = \langle \sqrt{\lambda_i} \cdot w_{i1}, \sqrt{\lambda_i} \cdot w_{i2}, \dots, \sqrt{\lambda_i} \cdot w_{in} \rangle = \langle w^*_{i1}, w^*_{i2}, \dots, w^*_{in} \rangle;$$

– согласование направлений нормированных ЭГК.

Критерием выбора направления ЭГК является максимизация значений целевой функции, вычисляемое как свертка значений данных на компоненты вектора весов. Максимальное значение неотрицательных переменных есть единица, поэтому для максимизации значений целевой функции выбираем из двух возможных направлений эмпирической главной компоненты то, при котором сумма ее координат максимальна;

– нахождение единого для всех наблюдений вектора весовых коэффициентов $W^* = \langle w^*_{11}, w^*_{12}, \dots, w^*_{1n} \rangle$ суммированием нагрузок согласованных эмпирических главных компонент $w^*_j = \sum_{i=1}^N w_{ij}^*$, $j = 1, \dots, n$, $i = 1, \dots, N$.

– вычисление интегрального показателя для наблюдения t как суммы произведений значений вектора весов и матрицы унифицированных показателей A^t

$$q^t = A^t \cdot W^* . \quad (5)$$

Для фиксированного момента t интегральную оценку для объекта с номером i записывают в виде аддитивной свертки данных с весами, определяемыми вектором W^* :

$$q_i^t = \sum_{j=1}^n w_j^* \cdot a_{ij}^t, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \quad (6)$$

5 шаг. Определение весовых коэффициентов блоков

– вес блока l \tilde{w}_l , где $l = 1, \dots, L$, (L — количество блоков) пропорционален величине суммы действующих ОСШ этого блока S_1^l :

$$\tilde{w}_l = \frac{S_1^l}{S_1^1 + S_1^2 + \dots + S_1^L}, \quad \sum_{l=1}^L \tilde{w}_l = 1, \quad \tilde{w}_l \geq 0, \quad S_1^l = \sum_{i=1}^{N_l} S_{1i}^l,$$

где N_l — число выбранных эмпирических главных компонент в блоке l .

6 шаг. Вычисление интегральной характеристики системы

– вычисляем характеристику всей системы, учитывая веса блоков:

$$y_i^t = \sum_{l=1}^L \tilde{w}_l \cdot q_{il}^t,$$

где q_{il}^t — интегральная оценка блока l для объекта с номером i в момент t .

Шаги 2,3,4 и 5 выполняются независимо для каждого блока. На шаге 6 используются результаты предыдущих шагов.

4. Вычисление интегральных индикаторов качества жизни населения с помощью предложенного алгоритма

Рассмотрим работу алгоритма для вычисления интегральной характеристики качества жизни населения субъектов Российской Федерации. Выбор переменных и деление их на блоки является принципиальным вопросом, лежащим вне рассматриваемого обсуждения. Поэтому воспользуемся списком переменных из исследования, выполненного под руководством С.А. Айвазяна (табл.1), являющегося признанным авторитетом в исследовании качества жизни [7]. Деление переменных на блоки также произведено авторами исследования [7]. Отсутствующие в открытом доступе данные заменим на близкие к ним (выделено цветом в табл.1). Все значения переменных взяты из открытых справочников Росстата [12].

Вменение отсутствующих значений является уязвимым местом методик вычисления интегральной характеристики. Наличие ряда наблюдений во многом снимает эту проблему. Единичные отсутствующие значения восстановлены линейной интерполяцией по соседним имеющимся, в случае отсутствия значений показателей для ряда лет их дополняем их средними значениями по тому федеральному округу, к которому относится рассматриваемый субъект Федерации.

Среди перечисленных переменных переменные 1, 2, 5, 7, 9, 10, 12, 21, 22, 23 связаны с вычисляемой характеристикой монотонной возрастающей зависимостью, когда оптимальное значение j -го показателя максимально. Для остальных показателей, кроме переменной 27, оптимальное значение показателя минимально. Для переменной 27 «Коэффициент миграционного прироста» примем $x_j^{opt} = 0$, $j=27$. Значение соответствует ситуации, когда количество выбывших из региона совпадает с количеством приехавших.

Определение четвертой эмпирической главной компоненты представлено в табл.2. Выбранные направления собственных векторов максимизируют сумму значений ОСШ у действующих переменных, для которых отношение сигнал/шум — отношение среднего к среднеквадратичному отклонению — не менее 2,2. В этой ЭГК оказалась одна действующая переменная (выделена темным цветом в табл.2). Факторные нагрузки этой переменной в ЭГК определит средняя величина нагрузки по годам, остальные переменные в ЭГК обнуляются (так же, как в методах факторного анализа).

Таблица 1

Переменные для вычисления интегральных индикаторов качества жизни населения

Уровень благосостояния населения	
1	Отношение ВРП на душу населения к прожиточному мин, единиц
2	Покупательская способность среднедушевых денежных доходов по отношению к наборам прожиточного минимума (%)
3	Доля численности населения с денежными доходами ниже прожиточного минимума (%)
4	Отношение совокупных доходов 20% самых богатых и 20% самых бедных
5	Обеспеченность населения собственными легковыми автомобилями (автомобилей/1 000 чел.)
6	Доля семей, состоящих на учете на получение жилья (%)
7	Приходится общей площади жилищного фонда на одного жителя (кв. м/10 чел.)
8	Доля ветхого и аварийного жилья (%)
9	Плотность автомобильных дорог общего пользования (км/10 000 кв. км)
Качество населения	
10	Ожидаемая продолжительность жизни при рождении (лет)
11	Число умерших детей в возрасте до 1 года на 1000 населения (чел.)
12	Коэффициент естественного прироста (на 1000 населения)
13	Число умерших от инфекционных и паразитарных болезней и туберкулеза на 100 000 нас.
14	Число умерших от новообразований на 100 000 населения (чел.)
15	Число умерших от болезней системы кровообращения на 100 000 населения (чел.)
16	Число умерших от болезней органов дыхания на 100 000 населения (чел.)
17	Число умерших от болезней органов пищеварения на 100 000 населения (чел.)
18	Заболеваемость от травм, отравлений и некоторых других внешних причин на 100 000 нас.
19	Число инвалидов на 1 000 населения (чел.)
20	Зарегистрировано случаев заболевания врожденными аномалиями на 1 000 населения
21	Доля специалистов с высшим образованием среди занятых в экономике (%)
22	Приведенная производительность труда (ВРП/среднегодовая численность занятых в экономике, тыс. руб./чел.)
23	Количество выпускников высших и средних учебных заведений на 1000 населения
Качество социальной сферы	
24	Уровень безработицы (%)
25	Доля работников занятых во вредных и опасных условиях труда в среднегодовой численности занятых в экономике (%)
26	Численность пострадавших на производстве со смертельным исходом или с утратой трудоспособности на 1 рабочий день и более на 1 000 работающих
27	Коэффициент миграционного прироста на 10 000 населения
28	Число зарегистрированных умышленных убийств и покушений на убийство на 100 000 нас.
29	Число зарегистрир. фактов умышл. причинения тяжкого вреда здоровью на 100 000 нас.
30	Число зарегистрированных изнасилований и покушений на изнасилование на 100 000 нас.
31	Число разбоев, грабежей, краж из квартир граждан в расчете на 100 000 населения
32	Число зарегистрированных незаконных присвоений или растрат в расчете на 100 000 нас.
33	Число больных, состоящих на учете с диагнозом наркомания и токсикомания на 100 000 нас.
34	Число больных, состоящих на учете с диагнозом алкоголизм в расчете на 100 000 нас.
35	Число больных, инфицированных туберкулезом в расчете на 100 000 населения
36	Смертность от внешних причин на 100 000 населения
37	Число больных психическими расстройствами на 100 000 населения

Таблица 2

Определение четвертой эмпирической главной компоненты Блока 1

4 ГК	Переменные								
	1	2	3	4	5	6	7	8	9
2007	0,21	-0,23	-0,31	-0,09	0,42	0,46	-0,14	0,04	-0,62
2008	0,08	-0,07	-0,44	-0,35	0,37	0,32	-0,56	0,31	-0,16
2009	-0,06	0,04	-0,36	-0,43	0,18	0,39	-0,64	0,30	0,03
2010	-0,01	-0,04	-0,42	-0,40	0,31	0,18	-0,60	0,40	0,12
2011	-0,10	-0,02	-0,40	-0,41	0,27	-0,07	-0,47	0,60	0,07
2012	-0,03	-0,04	-0,37	-0,34	0,22	-0,30	-0,29	0,72	0,02
2013	0,02	0,00	-0,20	-0,05	-0,03	-0,69	0,24	0,63	-0,20
2014	-0,06	-0,04	-0,28	-0,26	-0,37	0,27	-0,29	0,72	-0,20
<i>Среднее, т</i>	0,01	-0,05	-0,35	-0,29	0,17	0,07	-0,34	0,46	-0,12
<i>Выб. ср.кв.др.откл., s</i>	0,10	0,08	0,08	0,15	0,26	0,40	0,29	0,24	0,24
ОСШ	0,07	0,62	4,26	1,97	0,67	0,18	1,17	1,92	0,49
	Сумма ОСШ по строке								11,34
	Действующие ОСШ								4,26

В табл.3 представлены все нормализованные и согласованные по направлению эмпирические главные компоненты (ЭГК) первого блока. Поскольку четвертое собственное число $\lambda_4 = 0,79 < 1$, значения факторной нагрузки в этой ЭГК уменьшились по сравнению с табл.2. Направления ЭГК выбираются, чтобы максимизировать значение вычисляемой линейной свертки определяемых весов (как правило, положительных) с переменными, чьи значения лежат на интервале $[0, 1]$.

Таблица 3

Эмпирические главные компоненты Блока 1

№ ГК	Эмпирические собственные числа	Переменные								
		1	2	3	4	5	6	7	8	9
1	3,34	0,80	0,90	0,82	-0,75	0,53		0,38	0,31	
2	2,04	-0,27	-0,31		0,43	0,43	0,64	0,64	0,64	
3	1,14		0,19			-0,37			0,18	0,79
4	0,79			0,31						
5	0,61								0,34	
6	0,50					0,43			-0,26	
7	0,31	0,35	-0,12	-0,24						0,19
8	0,24				0,26			-0,20	0,12	
9	0,03		0,14	-0,08	0,08					
Вектор весов		0,88	0,79	0,81	0,02	1,02	0,64	0,83	1,33	0,98

Максимальное значение, вносимое ЭГК в композитный индекс, достигается, если все значения переменных равны единице, т.е. определяется суммированием факторных нагрузок ЭГК. Если сумма факторных нагрузок у ЭГК отрицательна, значение направление меняем. Так четвертая ЭГК, определяемая табл.2, изменила направление.

Искомый вектор весов определяется суммированием одноименных переменных выбранного числа эмпирических главных компонент, представленных в табл.3. Число та-

ких компонент выбирается, чтобы максимизировать информативность полученного решения согласно (4).

Таблица 4

Определение информативности интегрального показателя Блока 1.

Уровень благосостояния населения

	Номер эмпирической главной компоненты								
	1	2	3	4	5	6	7	8	9
Сумма ОСШ k -ой ЭГК	102,2	57,6	19,7	11,3	7,9	10,7	20,9	16,5	100,5
Сумма действующих ОСШ k -ой ЭГК	101,4	55,0	15,4	4,2	3,4	4,9	16,5	10,8	95,7
Накопленный вклад, γ_{SNR}	0,99	0,98	0,96	0,92	0,90	0,88	0,87	0,86	0,88
Эмпирические собственные числа	3,34	2,04	1,14	0,79	0,61	0,50	0,31	0,24	0,03
Накопленный вклад, γ_{σ}	0,37	0,60	0,72	0,81	0,88	0,94	0,97	1,00	1,00
Информативность, $\gamma_{\sigma} \cdot \gamma_{SNR}$	0,37	0,58	0,69	0,75	0,79	0,82	0,85	0,85	0,88

В табл.4 приведен пример определения информативности интегрального показателя Блока 1 «Уровень благосостояния населения». При рассмотрении всех 9 ЭГК суммарная информативность максимальна и составит около 88%. Поэтому для вычисления интегрального показателя используем все эмпирические главные компоненты. Чем больше переменных описывают систему, тем меньшее их относительное количество участвует в построении композитного индекса. Для второго блока выбираем 11 из 14 ЭГК, для третьего — 10 из 14 ЭГК. (В другом случае в системе, которую описывают 51 переменных, было выбрано 21 ЭГК). Веса блоков определяем пропорционально действующим ОСШ переменных этого блока (табл.5), что аналогично пропорциональности силе принятого сигнала. Блок 3 «Качество социальной сферы», в котором 14 переменных, оказался менее значим, чем блок 1 «Уровень благосостояния населения», в котором 9 переменных.

Таблица 5

Определение весов блоков

Блок	1	2	3
Сумма действующих ОСШ	307,4	397,7	254,3
Вес блока	0,32	0,41	0,27

Основной недостаток методик определения весов с помощью аппарата факторного анализа состоит в необоснованных резких изменениях рейтингов объектов для последовательных наблюдений. При рассмотрении коротких временных промежутков при отсутствии катастроф и потрясений естественно ожидать плавное изменение как регистрируемых показателей, так и интегрального показателя качества системы и рейтингов, выставляемых на основании этого показателя. Отсутствие резких колебаний является естественным следствием предположения об адекватности рассматриваемой модели объективной реальности.

В проведенном исследовании число объектов, где значения рейтинга изменились по сравнению с предыдущим годом более чем на 15% от возможного максимального изменения, составило 1,1% случаев. Для сравнения, в работах [3, 4], где вычислялись интегральные характеристики муниципальных образований Тюменской и Самарской области для ряда лет, эта величина составляет 42,7% и 43,2%. Среднее изменение рейтинга за год в этих исследованиях составляет около 20%, в приведенном исследовании — 3,7%. Приведенную в [3, 4] интегральную характеристику нельзя назвать надежной, и на её основе оценивать качество жизни субъектов РФ. Источником неудовлетворительного качества вычисленных характеристик может являться, в частности, погрешность исходных данных, которую не учитывает классический метод главных компонент. Отметим, что ранги 25 развитых европейских стран за 2009–2011 годы, выставленные на основании значений ИРПЧ, дают среднее изменение рейтинга за год 7,7% [18].

На рис.2 показаны изменения интегральной характеристики качества жизни для Уральского ФО. Ожидаемо северные субъекты округа, ориентированные на добычу нефти и газа, имеют более высокие показатели качества жизни, чем промышленные области. За 100% принято максимальное значение интегральной характеристики в 2007 году (Москва), за 0% — минимальное значение в 2007 году (Тыва).

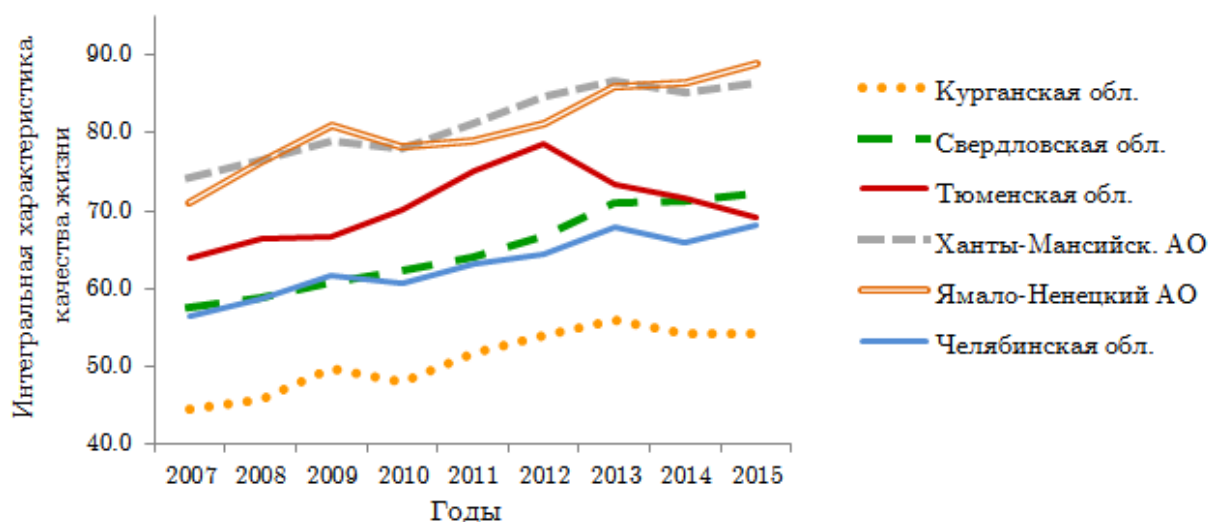


Рис. 2. Интегральная характеристика качества жизни некоторых субъектов Российской Федерации за 2007–2014 годы (2015 год — прогноз)

Заключение

В работе рассмотрено решение задачи построения латентной интегральной характеристики изменения качества системы на основании регистрируемых измерений для ряда наблюдений. Задача относится к классу задач выделения полезного сигнала из массива данных в условиях априорной неопределенности и решается на основании задаваемого отношения сигнал/шум. Предложен алгоритм построения интегральных характеристик с определением неслучайных составляющих главных компонент, характеризующих структуру рассматриваемой системы. Также предложен новый подход к выбору числа главных компонент, к определению весов рассматриваемых подсистем и к определению информативности полученной характеристики на основании дисперсионного критерия и выбранного параметра — отношения сигнал/шум. Алгоритм был использован для получения интегральных индикаторов качества жизни населения субъектов Российской Фе-

дерации. Предлагаемая методика может быть использована для вычисления интегральных оценок изменения качества плохо формализуемых систем. В дальнейшем следует проанализировать результаты применения методики построения интегральной характеристики качества системы не для заведомо хорошего (и не очень большого) набора входных переменных, а используя широкую базу измерений разных социальных систем, фиксируемую Росстатом.

Работа выполнена при финансовой поддержке проектной части государственного задания в сфере научной активности Министерства образования и науки Российской Федерации, проект № 1.949.2014/К.

Литература

1. Айвазян С.А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении межрегиональных сопоставлениях. М.: ЦЭМИ РАН, 2000. 56 с.
2. Айвазян С.А. К методологии измерения синтетических категорий качества жизни населения // Экономика и математические методы. 2003. Т. 39. № 2. С. 33–53.
3. Айвазян С.А., Степанов В.С., Козлова М.И. Измерение синтетических категорий качества жизни населения региона и выявление ключевых направлений совершенствования социально-экономической политики (на примере Самарской области и ее муниципальных образований) // Прикладная Эконометрика. 2009. №3(19). С. 18–84.
4. Гайдамак И.В., Хохлов А.Г. Моделирование интегральных показателей качества жизни населения юга Тюменской области // Вестник Тюменского государственного университета. 2009. № 6. С. 176–186.
5. Голяндина Н.Э., Усевич К.Д., Флоринский И.В. Анализ сингулярного спектра для фильтрации цифровых моделей // Геодезия и картография. 2008. №5. С. 21–28.
6. Голяндина Н.Э. Метод «Гусеница» — SSA: анализ временных рядов. СПб.: Санкт-Петербургский университет, 2004. 74 с.
7. Исакин М.А. Модификация метода k -средних с неизвестным числом классов // Прикладная эконометрика. 2006. Выпуск № 4. С. 62–70.
8. Жгун Т.В. Построение интегральной характеристики изменения качества системы на основании статистических данных как решение задачи выделения сигнала в условиях априорной неопределенности // Вестник Новгородского государственного университета. Серия: Технические науки. 2014. № 81. С. 10–16.
9. Жгун Т.В. Построения интегральной характеристики демографического развития территорий на примере муниципальных образований Новгородской области // Региональная экономика: теория и практика. 2013. № 36(315), сентябрь. С. 2–12.
10. Жгун Т.В. Вычисление интегрального показателя эффективности функционирования динамической системы на примере интегральной оценки демографического развития муниципальных образований Новгородской области // Вестник Новгородского государственного университета. Серия: Физико-математические науки. 2013. №75. Т.2. С.11–16.
11. Жгун Т.В. Исследование формальных методов построения латентной характеристики качества систем // Вестник Новгородского государственного университета. Серия: Физико-математические науки. 2014. № 80. С. 13–19.

12. Федеральная служба гос. статистики / Федеральная служба гос. статистики. — М., 2002–2014. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/stats/publications/catalog/doc_1137674209312(дата обращения: 13.01.2016).
13. Bandura R., del Campo C.M. Indices of National Performance: A Survey. Office of Development Studies, United Nations Development Programme, New York, November 17, 2006. URL: http://www.eldis.org/vfile/upload/1/document/1112/measuring_country_performance_2006update.pdf (дата обращения 01.07.2016).
14. Bandura R. A Survey of Composite Indices Measuring Country Performance: 2008 Update. Office of Development Studies United Nations Development Programme, New York, 2008/ 96 p. URL: http://web.undp.org/developmentstudies/docs/indices_2008_bandura.pdf (дата обращения 01.07.2016).
15. Bandura R. Composite Indicators and Rankings: Inventory 2011. Working Paper. Abbreviated version. 257 p. URL: <http://nebula.wsimg.com/beb0f5b0e9b6f0c298ae72e7519b22a4> (Дата обращения 01.07.2016).
16. Foa R.; Tanner J.C. Methodology of the Indices of Social Development / ISD Working Paper Series from International Institute of Social Studies of Erasmus University Rotterdam (ISS), The Hague. 2012. No 4. .66 p. URL: <http://repub.eur.nl/pub/50510/ISD-WP-2012-4.pdf> (дата обращения 01.07.2016).
17. Handbook on Constructing Composite Indicators: Methodology and User Guide. – OECD Publication. Paris CEDEX 16. 2008. 162 p.
18. Human Development Reports. 1990–2013. /United Nations Development Programme, 1990–2014. URL: <http://hdr.undp.org/en/reports/> (дата обращения: 13.06.2016).
19. Hightower W.L. Development of an index of health utilizing factor analysis // Medical Care, 1978. Vol. 16. P. 245–55.
20. Krishnan V. Constructing an Area-based Socioeconomic Index: A Principal Components Analysis Approach. Early Child Development Mapping Project (ЕСMap), Community–University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton, Alberta T5J 4P6, Canada. URL: http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf (дата обращения 01.07.2016).
21. Lindman C., Sellin J. Measuring Human Development. The Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011. 45 p. URL: <http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A464378&dswid=-5179> (дата обращения 01.07.2016).
22. McKenzie, D.J. Measuring Inequality with Asset Indicators. Journal of Population Economics. 2005. Vol. 18, Issue 2. P. 229–260.
23. Nardo M., Saisana M., Saltelli A., Tarantola S. Tools for composite indicators building. european commission, EUR 21682 EN. Joint Research Centre, Ispra, Italy.2005/URL: <http://farmweb.jrc.ec.eu.int/ci/bibliography.htm> (дата обращения 01.07.2016).
24. Nicoletti G., Scarpetta S., Boylaud O. Summary indicators of product market regulation with an extension to employment protection legislation, Economics department working papers NO. 226, ECO/WKP(99)18. 2000. URL: <http://www.oecd.org/eco/eco> (дата обращения 01.07.2016).
25. Saltelli A. Composite indicators between analysis and advocacy // Social Indicators Research, March 2007. Vol. 81, Issue 1. P. 65–77

26. Saltelli A., Munda G., Nardo M. From Complexity to Multidimensionality: the Role of Composite Indicators for Advocacy of EU Reform // Tijdschrift voor Economie en Management. 2006. Vol. LI, No 3 URL: http://feb.kuleuven.be/rebel/jaargangen/2001-2010/2006/TEM%202006-3/TEM_2006-3_03_Saltelli.pdf (дата обращения 01.07.2016).
27. Sharpe A. Literature Review of Frameworks for Macro-indicators, Ottawa, Canada: Centre for the Study of Living Standards. 2004. URL: <https://ideas.repec.org/p/sls/resrep/0403.html> (дата обращения 01.07.2016).
28. Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe // Social Indicators Research, Oct. 2009. Vol. 94, No 1. P. 115-133.
29. Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical details of the methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA) Italy. 2002. URL: http://ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf (дата обращения 01.07.2016).
30. Vyas S., Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. All rights reserved. DOI:10.1093/heapol/czl029 Advance Access publication 9 October 2006. URL: <http://heapol.oxfordjournals.org> (дата обращения 01.07.2016).

Жгун Татьяна Валентиновна, к.ф.-м.н., доцент, кафедра прикладной математики и информатики, Новгородский государственный университет имени Ярослава Мудрого (Великий Новгород, Российская Федерация)

DOI: 10.14529/cmse170101

AN ALGORITHM FOR CONSTRUCTING INTEGRAL QUALITY INDICATOR OF COMPLEX SYSTEMS FOR A SEQUENCE OF OBSERVATIONS

© 2017 T. V. Zhgun

Novgorod State University a. Yaroslav the Wise

(Bolshaya Sankt-Peterburgskaya Avenue, 41, Veliky Novgorod, 173003 Russia)

E-mail: Tatyana.Zhgun@novsu.ru

Received: 26.10.2016

In the paper we propose an algorithm of constructing the integral characteristics of changes in the quality system based on the recorded measurements. The algorithm provides allotment signal in a multidimensional array of data in terms of a priori uncertainty on the basis of the specified signal-to-noise ratio. Constructing of latent integral characteristics of changes of the quality system on the basis of statistical indicators for a number of consecutive observations is based on the principal component method, taking into account the presence of noise in the measured data (SNR-based algorithm). In classical PCA informativity of integral characteristics is given a priori and is provided by selecting the number of principal components. In the proposed algorithm the information content of the solution is evaluated a posteriori on the basis of variance criterion and the selected parameter signal-to-noise ratio. Based on the proposed algorithm we construct integral indicators of the quality of life of subjects of the Russian Federation for the years of 2007-2014.

Keywords: quality management system, an integral characteristic of the quality, principal component analysis change the quality characteristics, the integral indicators of quality of life, noise in the measured data, the signal-to-noise ratio, informativity of the principal component analysis.

FOR CITATION

Zhgung T.V. An algorithm for constructing integral quality indicator of complex systems for a sequence of observations. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017. vol. 6, no. 1. pp. 5–25. (in Russian) DOI: 10.14529/cmse170101.

References

1. Ayvazyan S.A. Integral'nye indikatory kachestva zhizni naselenija: ih postroenie i ispol'zovanie v social'no-jekonomicheskom upravlenii mezhregional'nyh sopostavlenijah. [Integrated Indicators of Quality of Life of the Population: Their Construction and use in Social and Economic Management Interregional Comparisons]. M.: TsEMI Russian Academy of Sciences, 2000. 56 p.
2. Ayvazyan S.A. *K metodologii izmerenija sinteticheskikh kategorij kachestva zhizni nasele-nija*. [To Methodology of Measurement of Synthetic Categories of Quality of Life of the Population]. *Jekonomika i matematicheskie metody*. [Economy and Mathematical Methods]. 2003. vol. 39, no. 2. pp. 33–53. (in Russian)
3. Ayvazyan S.A., Stepanov V.S., Kozlova M.I. *Izmerenie sinteticheskikh kategorij kachest-va zhizni naseleniya regiona i vyyavlenie klyuchevykh napravleniy sovershenstvovaniya sotsial'no-ekonomicheskoy politiki (na primere Samarskoy oblasti i ee munitsipal'nykh obrazovaniy)* [Measurement of Synthetic Categories of the Quality of Life of Population in the Region and Identify Key Areas of Improvement in Socio-Economic Policy (on the Example of Samara Region and its Municipalities)]. *Prikladnaja Jekonometrika* [Applied Econometrics]. 2009. no. 3(19). pp. 18–84. (in Russian)
4. Gaydamak I.V., Khokhlov A.G. *Modelirovanie integral'nykh pokazateley kachestva zhizni naseleniya yuga Tyumenskoy oblasti* [Modeling of Integral Indicators of the Quality of Life of the Population of the South of the Tyumen Region]. *Vestnik Tjumenskogo gosudarstvennogo universiteta*. [Bulletin of the Tyumen State University]. 2009. no. 6. pp. 176–186. (in Russian)
5. Goljandina N.J., Usevich K.D., Florinskij I.V. *Analiz singular'nogo spektra dlja fil'tracii cifrovyyh modelej* [The Analysis of a Singular Range for a Filtration of Digital Models]. *Geodeziya i kartografiya*. [Geodesy and Cartography] 2008, no. 5. pp. 21–28. URL: <http://iflorinsky.narod.ru/Florinsky-2008ts.pdf>. (accessed 13.01.2015). (in Russian)
6. Goljandina N.J. *Metod "Gusenica" — SSA: analiz vremennykh rjadov* [The Method of "Caterpillar" — SSA: Analysis of Time Series, St. Petersburg.: St. Petersburg University Press, 2004. 74 p.
7. Isakin M.A. *Modifikacija metoda k –srednih s neizvestnym chislom klassov* [Modification of a Method to k -Means with Unknown Number of Classes]. *Prikladnaya ekonometrika* [Applied Econometrics]. 2006. Release no. 4. pp. 62–70. (in Russian)
8. Zhgung T.V. *Postroenie integral'noj harakteristiki izmenenija kachestva sistemy na osno-vanii statisticheskikh dannykh kak reshenie zadachi vydelenija signala v uslovijah apriornoj neopredelennosti* [Creation of the Integrated Characteristic of Change of Quality of System on the Basis of Statistical Data as the Solution of a Problem of Allocation of a Signal in the Conditions of Aprioristic Uncertainty]. *Vestnik Novgorodskogo gosudarstvennogo universiteta. Seriya: Tekhnicheskie nauki* [Bulletin of Novgorod State University. Series: Engineering sciences.]. 2014. no. 81. pp. 10–16. (in Russian)

9. Zhgun T.V. *Postroeniya integral'noj harakteristiki demograficheskogo razvitija territorij na primere municipal'nyh obrazovaniy Novgorodskoj oblasti* [Creation of the Integrated Characteristic of Demographic Development of Territories on the Example of Municipalities of the Novgorod Region] *Regional'naya ekonomika: teoriya i praktika* [Regional Economy: Theory and Practice]. 2013. no. 36(315), September. pp. 2–12. (in Russian)
10. Zhgun T.V. *Vychislenie integral'nogo pokazatelya jeffektivnosti funkcionirovaniya dinamicheskoy sistemy na primere integral'noj ochenki demograficheskogo razvitija municipal'nyh obrazovaniy Novgorodskoj oblasti* [Calculation of an Integrated Indicator of Efficiency of Functioning of Dynamic System on the Example of an Integrated Assessment of Demographic Development of Municipalities of the Novgorod Region]. *Vestnik Novgorodskogo gosudarstvennogo universiteta. Seriya: Fiziko–matematicheskie nauki* [Bulletin of Novgorod State University. Series: Physico–Mathematical Sciences]. 2013. no. 75. vol. 2. pp. 11–16. (in Russian)
11. Zhgun T.V. *Issledovanie formal'nyh metodov postroeniya latentnoj harakteristiki kachestva system* [Research of Formal Methods of Creation of the Latent Characteristic of Quality of Systems. *Vestnik Novgorodskogo gosudarstvennogo universiteta Seriya: Fiziko–matematicheskie nauki* [Bulletin of Novgorod State University. Series: Physico–Mathematical Sciences]. 2014. no. 80. pp. 13–19. (in Russian)
12. *Federal'naja sluzhba gos. Statistiki* [Federal service state. statistics]. Federal Service State. Statistics.–M, 2002–2014. URL: http://shchshchshch.gks.ru/shchps/shchtsm/tsonetst/rosstat_main/rosstat/ru/statistitss/publitsations/tsatalog/dots_1137674209312 (accessed 13.01.2016).
13. Bandura R., del Campo C.M. *Indices of National Performance: A Survey*. Office of Development Studies, United Nations Development Programme, New York, November 17, 2006. 91 p. URL: http://www.eldis.org/vfile/upload/1/document/1112/measuring_country_performance_2006update.pdf (accessed: 01.07.2016).
14. Bandura R. *A Survey of Composite Indices Measuring Country Performance: 2008 Update*. Office of Development Studies United Nations Development Programme, New York, 2008/ 96 p. URL: http://web.undp.org/developmentstudies/docs/indices_2008_bandura.pdf (accessed: 01.07.2016).
15. Bandura R. *Composite Indicators and Rankings: Inventory*. 2011. Working Paper. Abbreviated version. 257 p. URL: <http://nebula.wsimg.com/beb0f5b0e9b6f0c298ae72e7519> (accessed: 01.07.2016).
16. Foa R., Tanner J. C. *Methodology of the Indices of Social Development / ISD Working Paper Series from International Institute of Social Studies of Erasmus University Rotterdam (ISS), The Hague*. 2012. No 4. P.66.
17. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publication. Paris CEDEX 16. 2008. 162 p.
18. *Human Development Reports. 1990–2013/ United Nations Development to the Program, 1990–2014*. URL: <http://hdr.undp.org/en/reports> (accessed 13.06.2016).
19. Hightower W. L. (1978). *Development of an Index of Health Utilizing Factor Analysis*. *Medical Care*. 1978. no. 16. pp. 245–55.
20. Krishnan V. *Constructing an Area–based Socioeconomic Index: A Principal Components Analysis Approach*. Early Child Development Mapping Project (ECMap), Community–University Partnership (CUP), Faculty of Extension, University of Alberta, Edmonton,

- Alberta T5J 4P6, Canada. URL: http://www.cup.ualberta.ca/wp-content/uploads/2013/04/SEICUPWebsite_10April13.pdf (accessed: 01.07.2016).
21. Lindman C., Sellin J. Measuring Human Development. The Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011, 45 p. URL: <http://uu.diva-portal.org/smash/record.jsf?pid=diva2%3A464378&dsid=-5179> (accessed: 01.07.2016).
 22. McKenzie D. J. Measuring Inequality with Asset Indicators. *Journal of Population Economics*. 2005. vol. 18, issue 2. pp. 229–260.
 23. Nardo M., Saisana M., Saltelli A., Tarantola S. Tools for Composite Indicators Building. European Commission, EUR 21682 EN. Joint Research Centre, Ispra, Italy. 2005. URL: <http://farmweb.jrc.cec.eu.int/ci/bibliography.htm> (accessed: 01.07.2016).
 24. Nicoletti G., Scarpetta S., Boylaud O. Summary Indicators of Product Market regulation with an Extension to Employment Protection Legislation, Economics Department Working Papers NO. 226, ECO/WKP(99)18. 2000. URL: <http://www.oecd.org/eco/eco> (accessed: 01.07.2016).
 25. Saltelli A. Composite Indicators Between Analysis and Advocacy. *Social Indicators Research*. March 2007. vol. 81, issue 1. pp 65–77
 26. Saltelli A., Munda G., Nardo M. From Complexity to Multidimensionality: the Role of Composite Indicators for Advocacy of EU Reform. *Tijdschrift voor Economie en Management*. 2006. Vol. LI, No 3 URL: http://feb.kuleuven.be/rebel/jaargangen/2001-2010/2006/TEM%202006-3/TEM_2006-3_03_Saltelli.pdf (accessed: 01.07.2016).
 27. Sharpe A. Literature Review of Frameworks for Macro-indicators, Ottawa, Canada: Centre for the Study of Living Standards. 2004. URL: <https://ideas.repec.org/p/sls/resrep/0403.html> (accessed: 01.07.2016).
 28. Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe. *Social Indicators Research*, 2009. vol. 94, issue 1. pp. 115–133.
 29. Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical Details of the Methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA) Italy. 2002. URL: http://ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf (accessed: 01.07.2016).
 30. Vyas S., Kumaranayake L. *Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis*. Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine. DOI:10.1093/heapol/czl029 Advance Access publication 9 October 2006. URL: <http://heapol.oxfordjournals.org> (accessed: 01.07.2016).