

Анализ последовательности состояний риска при выполнении информационных проектов

Наумов П. В.

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации (Северо-Западный институт управления РАНХиГС), Санкт-Петербург, Российская Федерация; naumov122@gmail.com

РЕФЕРАТ

В статье рассмотрены вопросы менеджмента риска выполнения информационных проектов. Предложен не традиционно используемый при анализе рисков дедуктивный подход, основанный на знании основных положений и их конкретизации в реальной ситуации, а индуктивный подход, требующий накопления данных и построения обучающих наборов данных в результате ретроспективного анализа хода и выполнения отдельных проектов. Предложено использовать методы траекторного анализа последовательностей событий, каждое из которых характеризует состояние риска выполнения проекта. Для проверки данного подхода сгенерированы случайные последовательности событий риска и применены к данным последовательностям методы анализа данных.

Ключевые слова: информационный проект, риски выполнения проекта, методы менеджмента рисков, случайный процесс, дискретная марковская цепь, случайная последовательность событий, показатели случайных последовательностей, методы машинного обучения, кластерный анализ

Для цитирования: Наумов П. В., Куклина Е. А. Анализ последовательности состояний риска при выполнении информационных проектов // Управленческое консультирование. 2021. № 4. С. 159–173.

Analysis of the Sequence of Risk States in the Implementation of Information Projects

Pavel V. Naumov

Russian Presidential Academy of National Economy and Public Administration (North-West Institute of Management of RANEPА), Saint-Petersburg, Russian Federation; naumov122@gmail.com

ABSTRACT

The article deals with the issues of risk management of information projects. We propose not a deductive approach traditionally used in risk analysis, based on knowledge of the main provisions and their concretization in a real situation, but an inductive approach that requires data accumulation and the construction of training data sets as a result of a retrospective analysis of the progress and implementation of individual projects. It is proposed to use methods of trajectory analysis of sequences of events, each of which characterizes the state of risk of project implementation. To test this approach, random sequences of risk events are generated and data analysis methods are applied to these sequences.

Keywords: Information project, project execution risks, risk management methods, random process, discrete Markov chain, random sequence of events, random sequence indicators, machine learning methods, cluster analysis

For citing: Naumov P.V., Kuklina E. A. Analysis of the Sequence of Risk States in the Implementation of Information Projects // Administrative consulting. 2021. N4. P. 159–173.

Введение

Цифровая трансформация экономики предусматривает увеличение доли информационных проектов, их внедрение в различные сферы деятельности, в различные отрасли. Принятая национальная программа и выполняемые национальные проекты направлены на формирование экосистемы цифровой экономики, создание и совершенствование информационных систем в государственном управлении и в бизнесе. Бюджет национального проекта «Цифровая экономика» составляет более 1,5 трлн руб., в том числе на развитие информационной инфраструктуры предусмотрено около 800 млрд руб. Только в Санкт-Петербурге на централизованное сопровождение и развитие ГИС в 2020 г. из регионального бюджета выделяется 1 275 075,4 тыс. руб. и 1 295 381 тыс. руб. соответственно [4; 7]¹. Такой масштаб решаемых задач, как и масштаб финансирования, предполагает необходимость совершенствования системы управления ее развитием.

Несмотря на усилия, предпринимаемые в теории управления проектами, существует большое число проектов, которые завершаются неудачей. Так, консалтинговая компания Standish Group ежегодно публикует отчет о результатах выполнения программных проектов. База данных компании содержит данные о более 50 тыс. проектах в государственном или частном секторах. Анализ доступных сведений об их результативности показывает, что риски выполнения информационных проектов остаются высокими. Так, например, в отчете за 2015 г. указывается, что при использовании водопадной модели жизненного цикла проекта более 20% проектов завершаются неудачей. Около 60% проектов выполняются частично с превышением стоимости или сроков и только 11% выполняются полностью в рамках выделенного бюджета. Использование гибких технологий проектирования лишь частично улучшает ситуацию. Доля полностью успешных проектов составляет 39%. Таким образом, существует большая проблема в области управления информационными проектами в условиях рисков. Напомним девиз разработчиков проектов из Microsoft: «Мы не боремся с рисками — мы ими управляем». Управление рисками проекта направлено на снижение вероятности возникновения и/или значимости воздействия неблагоприятных для проекта событий.

Методы управления рисками определены в стандартах и сводах знаний, например, в стандарте ГОСТ Р ИСО/МЭК 31010-2011². В нем приведено описание различных методов оценки рисков, сформулированы рекомендации по их применению в зависимости от этапа жизненного цикла проекта, а также требуемых степени глубины и детализации оценки. Данные методы в основном ориентированы на прогнозирование опасных событий в будущем, а не ретроспективное исследование ситуаций риска, организации управления ими в ходе выполнения проекта.

В своде РМВОК [9; 12] выделена отдельная область знаний управления рисками. Использованный в своде знаний процессный подход позволил определить основные процессы: планирования, идентификации, качественного и количественного анализа, реагирования, мониторинга и управления рисками. При их описании произведена характеристика рекомендуемых к использованию методов и инструментальных средств.

Методы программной инженерии, основанные на теории Б. Бозма, позволяют выполнить количественное обоснование информационных проектов, оценку затрат, сроков проектирования, а также оценить риски управления проектами.

¹ См. также: Закон Санкт-Петербурга от 29 ноября 2019 года № 614-132 «О бюджете Санкт-Петербурга на 2020 год и на плановый период 2021 и 2022 годов» [Электронный ресурс]. URL: <https://rg.ru/2019/12/02/spb-zakon614-132-reg-dok.html> (дата обращения: 14.12.2020).

² Менеджмент риска. Методы оценки риска. ГОСТ Р ИСО/МЭК 31010-2011 [Электронный ресурс]. URL: <http://docs.cntd.ru/document/gost-r-iso-mek-31010-2011> (дата обращения: 03.01.2021).

В публикациях Р. Валерди предложен специальный метод COSYSMO, в котором в зависимости от сложности выполняемого проекта оцениваются его экономические показатели [22; 23]. Используемый метод Монте-Карло позволяет построить эмпирические законы распределения трудозатрат, сроков и стоимости проектирования. Метод СОСUALMO, входящий в разрабатываемую школой Б. Боема теорию экономики информационных проектов [13; 14], позволяет оценивать число программных ошибок при выполнении проектов, используя модель системной динамики Дж. Форрестера [3; 10; 11]. Разработанный онлайн-калькулятор позволяет в зависимости от драйверов программного проекта, программного продукта, команды проекта оценить риски проекта, платформы, персонала и процесса. В зависимости от уровня риска формируются производственные правила-рекомендации по действиям по снижению классифицированных рисков. Данные правила базируются на дедуктивном подходе, который предусматривает, что для каждой выявленной ситуации известен рекомендуемый порядок действий. Альтернативой дедуктивному подходу является индуктивный подход, который основан на накоплении опыта необходимых действий.

Одним из наиболее популярных методов, рассмотренных в данной области знаний, является метод вероятности и воздействия. Его основу составляет построение и исследование специальной матрицы, ячейками которой являются категории рисков. В том случае, если такие категории оценивать периодически на различных стадиях жизненного цикла проекта, возможно построить траекторию развития ситуации риска, которую можно представить ориентированным графом G с вершинами — состояниями риска, а дугами — отношения следования. Для классификации состояния риска могут быть использованы различные методы машинного обучения, учитывающие ход выполнения информационного проекта. Один из подходов к решению задачи классификации рассмотрен в [8]. Здесь предполагается, что существует набор классификационных признаков, определяющих успешность или неуспешность проектов. Это позволяет использовать индуктивный подход к решению задач оценки рисков, использовать методы машинного обучения и интеллектуального анализа данных.

В теории менеджмента риска такой граф рассматривается как граф непрерывной или дискретной марковской цепи, что позволяет для количественного анализа рисков использовать методы теории марковских цепей и определять вероятностные характеристики случайных процессов. В отличие от общепринятого подхода, позволяющего решать задачи прогнозирования, предлагается использовать историю процесса проектирования для того, чтобы выполнить ретроспективный анализ накопленного опыта проектирования.

Предположим, что имеется опыт управления проектами, который описан как траектории — реализации случайного процесса $X_i(t)$. Каждая реализация представляет собой последовательность состояний $X_i(t) = (s_i(t_1), s_i(t_2), \dots, s_i(t_n))$, где t_1, t_2, \dots, t_n — моменты времени (вехи выполняемого проекта), $s_i(t_1), s_i(t_2), \dots, s_i(t_n)$ — состояния ситуации риска.

Первые задачи анализа последовательностей связаны с биоинформатикой. Фред Сэнгер исследовал последовательности белка инсулина для понятия функции молекул ДНК¹. Накопленный объем баз данных первичных последовательностей растет экспоненциально. Так, база данных GeneBank составляет более 160 млн записей (150 млрд пар оснований)². Задачи анализа случайной последовательности решались,

¹ Анализ последовательности [Электронный ресурс]. URL: https://ru.qaz.wiki/wiki/Sequence_analysis (дата обращения: 02.01.2021).

² Зленко Д. В. Анализ последовательностей [Электронный ресурс]. URL: http://media.biophys.msu.ru/MMB_2013/lecture17.pdf (дата обращения: 08.01.2021).

например, при анализе последовательности покупки товаров в интернет-магазине [2]. По мнению авторов данной публикации, анализ случайных последовательностей событий позволит вывести маркетинговые мероприятия на новый качественный уровень, будет способствовать повышению качества обслуживания покупателей.

Такие последовательности можно рассматривать как набор специальных знаков и дальше работать с ними как с обычными числами. Метод анализа последовательностей событий или знаков использовался, например, при анализе семейного статуса швейцарских граждан на протяжении 16 лет. Данный набор приведен в библиотеке языка R и подробно проанализирован в [6].

Материалы и методы

Для решения задачи ретроспективного анализа ситуации риска во время проектирования предположим, что существуют ситуации риска, которые будут кодировать специальными символами — цифрами от 0 до 5, как показано в табл. 1.

Таблица 1

Ситуации риска выполнения проекта

Table 1. Project Risk Situations

Код ситуации риска	Ситуация риска
0	Риска нет
1	Малый риск
2	Средний риск
3	Риск выше среднего
4	Высокий риск
5	Очень высокий риск

Если каждой ситуации риска поставить в соответствие вершину графа марковской цепи, то данный граф можно представить полносвязным псевдографом с петлями. Предположим, что каждый проект выполняется в течение одного года. Тогда траектория случайной последовательности на данном графе будет иметь длину 12 и содержать отдельные состояния — вершины графа. Данная траектория определяется спецификой выполнения проекта. Множество таких траекторий позволит сформировать обучающую выборку, анализ которой является содержанием траекторного анализа TramineR [1; 15; 16–19].

Основное содержание такого анализа базируется на исследовании имеющейся выборки. Так, при анализе швейцарских граждан используется выборка из 2000 наблюдений. При анализе данных об ирландских школьниках с данными об образовании и об их трудоустройстве рассматривается выборка из 712 наблюдений и 86 переменных. Отметим, что получение таких наборов данных является проблемой. В дальнейшем для исследования будем использовать синтетические выборки, которые формируются с помощью метода Монте-Карло.

Предположим, что выполняется проект, риски которого могут быть оценены. Для такого проекта может быть построен граф дискретной марковской цепи с шагом, равным одному месяцу. Ниже приведены две матрицы вероятностей переходов за один шаг. Размерность матрицы определена числом ситуаций риска. Элементы матрицы определяют вероятность перехода между состояниями случайного процесса за один шаг. Матрицы отличаются значениями данных вероятностей. Первая

матрица соответствует практически безрисковой ситуации выполнения информационного проекта. Вторая матрица ситуации сравнительно невысокого риска.

$$W = \begin{pmatrix} 0,7 & 0,2 & 0,1 & 0 & 0 & 0 \\ 0,3 & 0,4 & 0,2 & 0,1 & 0 & 0 \\ 0,2 & 0,4 & 0,2 & 0,1 & 0,1 & 0 \\ 0 & 0,2 & 0,3 & 0,3 & 0,1 & 0,1 \\ 0 & 0 & 0,3 & 0,3 & 0,2 & 0,2 \\ 0 & 0 & 0,1 & 0,3 & 0,3 & 0,3 \end{pmatrix} \cdot W = \begin{pmatrix} 0,5 & 0,3 & 0,2 & 0 & 0 & 0 \\ 0,2 & 0,5 & 0,2 & 0,1 & 0 & 0 \\ 0,2 & 0,3 & 0,3 & 0,1 & 0,1 & 0 \\ 0 & 0,1 & 0,4 & 0,3 & 0,1 & 0,1 \\ 0 & 0 & 0,3 & 0,3 & 0,2 & 0,2 \\ 0 & 0 & 0,1 & 0,3 & 0,3 & 0,3 \end{pmatrix}$$

При допущении, что в начальный момент времени риски отсутствуют и процесс находится в нулевом состоянии, пользуясь уравнением Колмогорова, можно вычислить распределение вероятностей его нахождения в различных состояниях в каждый момент времени дискретного случайного процесса:

$$P(n) = (1 \ 0 \ 0 \ 0 \ 0 \ 0)W^n,$$

где n — шаг прогнозирования (номер месяца выполнения проекта).

На рис. 1 приведена диаграмма распределения вероятностей нахождения процесса в различных состояниях, показывающая, что в процессе проектирования вероятность возникновения ситуации риска повышается. К сожалению, приведенный граф предполагает, что известны все законы распределения вероятностей, что скорее будет исключением, чем правилом. С другой стороны, полученные

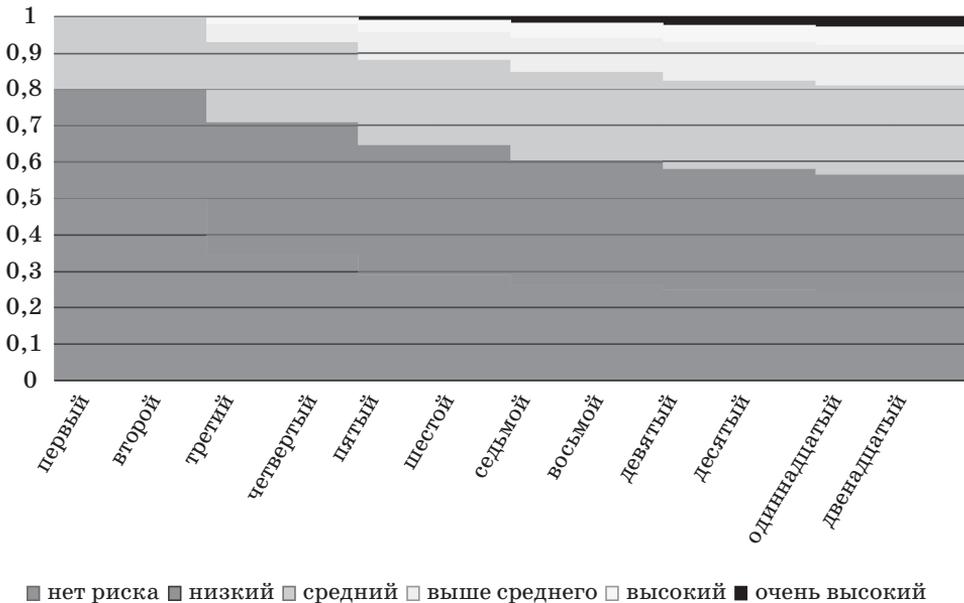


Рис. 1. Распределение вероятностей нахождения процесса в различных состояниях для второго графа

Fig. 1. Distribution of probabilities of finding a process in different states for the second graph

результаты позволяют проверить адекватность результатов моделирования при создании синтетических выборок.

Полученные синтетические выборки для двух рассмотренных примеров позволяют использовать методы анализа последовательностей для оценки рисков проектов. Каждое наблюдение полученных выборок может быть рассмотрено как траектория выполнения проекта. Случайная последовательность будет представлена последовательностью символов из алфавита 0–5, что соответствует ситуации риска. Пусть две синтетические выборки будут содержать 1000 наблюдений. Результаты генерации позволят в дальнейшем использовать методы машинного обучения для решения задач исследования последовательностей. Такие последовательности можно представить множеством пар чисел. Первым элементом пары является номер состояния, а вторым — число последовательных шагов нахождения в данном состоянии. Такое представление (SPS-представление) позволяет проанализировать динамику изменения состояний — неопределенность, а также сложность процесса.

Результаты

На рис. 2 приведены 10 наиболее популярных траекторий для первого из двух рассмотренных примеров. Данные траектории показывают, что, несмотря на случайный характер анализируемых процессов, можно определить их стабильность (нахождение несколько моментов времени в одном и том же состоянии) или изменчивость. Приведенный рисунок показывает, что наиболее популярной является траектория (нижняя строка рисунка), для которой в первые шесть моментов времени процесс находится в нулевом состоянии — состоянии отсутствия риска. В нотации STS данный процесс описывается последовательностью (0,6)-(2,1)-(1,2)-(3,2)-(1,1). Используемое представление позволяет сравнить различные траектории, определить расстояние между ними, определить стационарность анализируемых случайных процессов, например, с помощью критериев серий, исследующих подпоследовательности в анализируемом временном ряду.

Полученные последовательности можно количественно анализировать с помощью различных статистических показателей: средней длины каждой из подпоследовательностей; доли нахождения процесса проектирования в состояниях невысокого, высокого и других категорий риска; сложности; энтропии, турбулентности. Используя графические средства анализа последовательностей, построим частотные распределения состояния риска по мере выполнения проектов. На рис. 3, рис. 4 приведены такие распределения для двух заданных вариантов дискретной марковской цепи.

Приведенные диаграммы, с одной стороны, соответствуют теоретическому распределению, приведенному на рис. 1, что подтверждает адекватность решения задачи создания выборок с помощью метода Монте-Карло, а с другой стороны, позволяют выполнить исследования ситуации выполненных проектов. Приведенные рисунки показывают, что в основном проекты находятся в состоянии невысокого риска, что позволяет сделать вывод о хорошей организации управления рисками в исследуемых информационных проектах. Доля состояний отсутствия риска для первого и второго примеров свидетельствует, что в целом проекты выполняются в условиях невысокого риска и, возможно, они будут успешными. Первая диаграмма показывает, что наиболее вероятным является состояние отсутствия риска. Для второго процесса наиболее вероятным является состояние невысокого риска (рис. 5).

Наличие дополнительных признаков, характеризующих проекты, позволяет решать традиционные задачи машинного обучения, например, задачи их кластеризации. С этой целью должно быть использовано понятие близости, схожести наблюдений или признаков. Для этого могут применяться: расстояние Хэмминга, позволяющее определить число несовпадающих позиций в траектории, рассматриваемой как

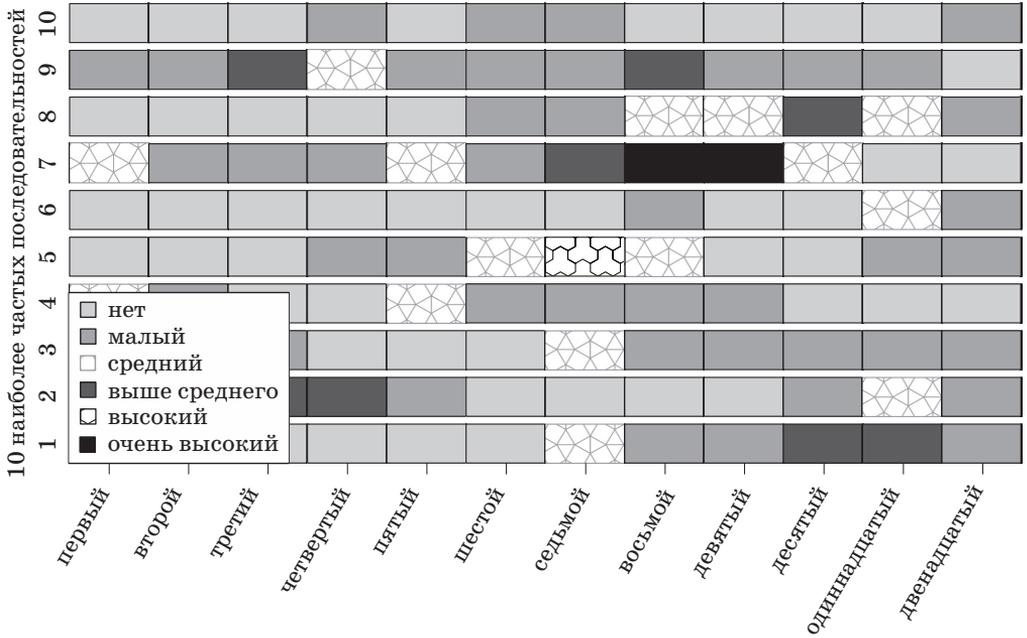


Рис. 2. Приоритетные траектории реализации процесса
 Fig. 2. Priority process paths

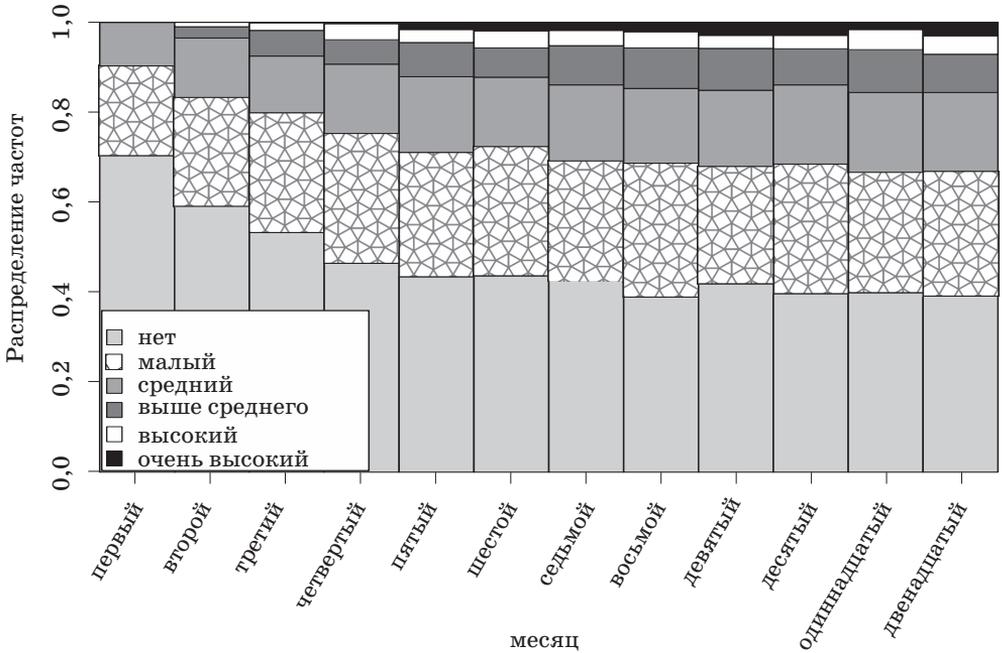


Рис. 3. Распределение частот нахождения проекта в различных состояниях риска для первой дискретной марковской цепи
 Fig. 3. Distribution of project frequencies in different risk states for the first discrete Markovian chain

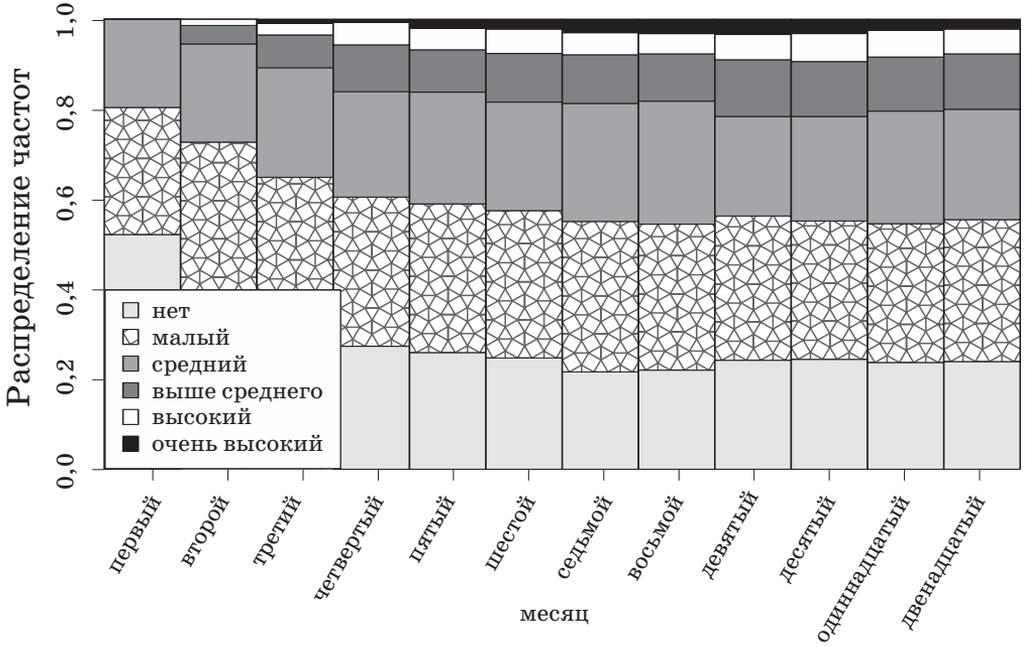


Рис. 4. Распределение частот нахождения проекта в различных состояниях риска для второй дискретной марковской цепи
 Fig. 4. Distribution of project frequencies in different risk states for the second discrete Markovian chain

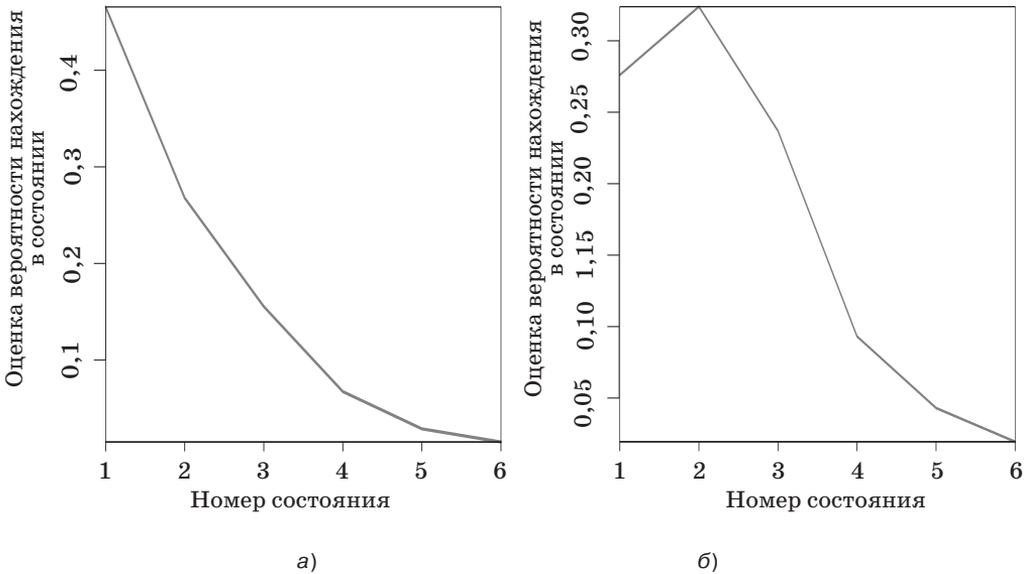


Рис. 5. Диаграммы распределения вероятностей нахождения процесса в различных состояниях риска: а) для первого примера; б) для второго примера
 Fig. 5. Diagrams of the probability distribution of finding a process in different risk states: а) for the first example; б) for the second example

бинарный вектор; длина совпадающих частей от начала последовательности; максимальная длина общей подпоследовательности; суммарная «стоимость» преобразования одной последовательности в другую, так называемая метрика Левенштейна [5; 20; 21; 24].

Вспользуемся последней метрикой, которая позволяет построить матрицу расстояний между анализируемыми последовательностями выполняемых проектов, так называемую матрицу оптимального редактирования ОМ (optimal matching distances). На рис. 6, рис. 7 приведены результаты решения задачи кластерного анализа для двух рассматриваемых примеров. Число кластеров было определено с помощью агломеративного метода кластеризации с использованием ОМ-матрицы и метода Уорда.

Столбчатые диаграммы, приведенные на рисунках, показывают распределение наблюдений по среднему числу нахождения процесса в различных состояниях. В табл. 2 приведены частоты попадания анализируемых процессов в различные кластеры.

Первый кластер первого примера характеризуется сравнительно большим числом этапов выполнения проекта, для которых процесс находится в нулевом или первом состоянии. При этом возможны и состояния среднего и высокого рисков. Во второй кластер попадают наблюдения, для которых мода распределения приходится на состояние малого риска. Однако сравнительно большое число этапов характеризуется средним, выше среднего, высоким и очень высоким риском. В третий кластер

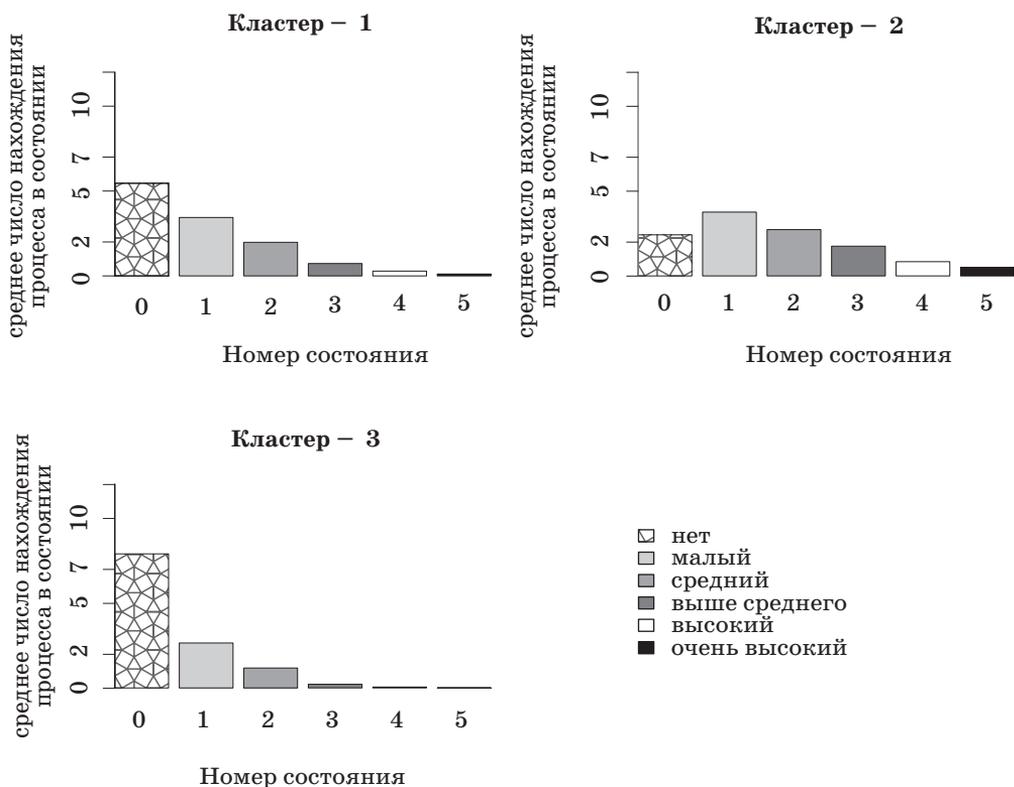


Рис. 6. Кластеризация процессов для первого примера
 Fig. 6. Process clustering for the first example

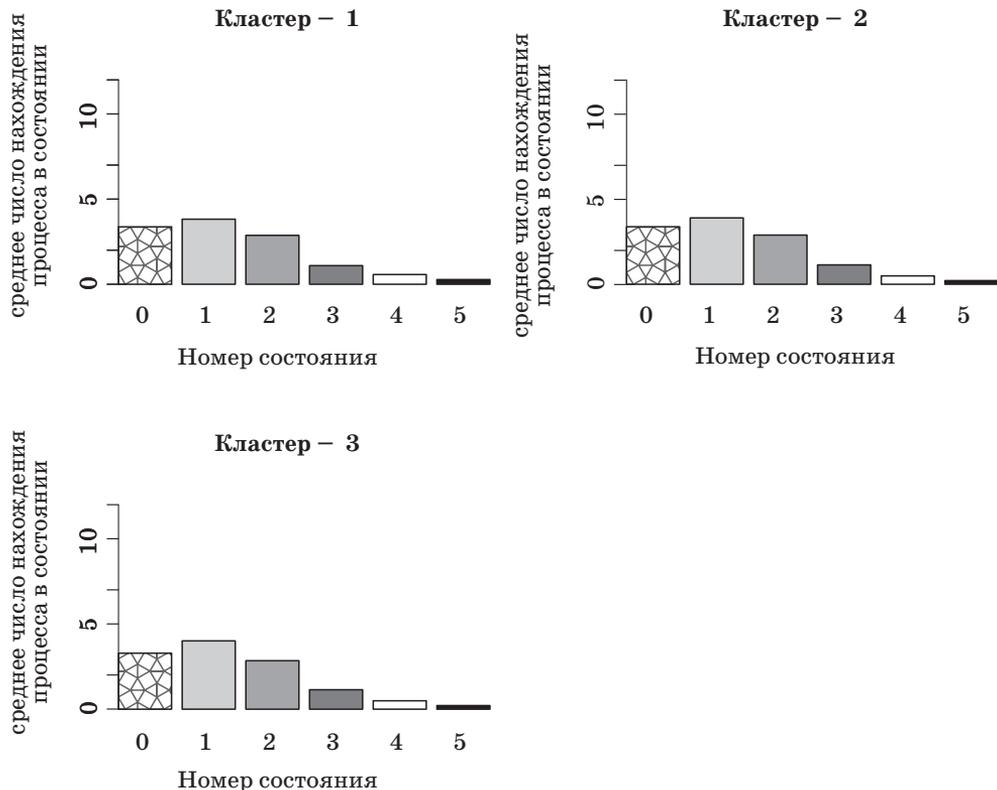


Рис. 7. Кластеризация процессов для второго примера
Fig. 7. Process clustering for the second example

Таблица 2

Распределение наблюдений по кластерам
Table 2. Distribution of observations by clusters

Номер примера	Номер кластера		
	Первый	Второй	Третий
Первый	324	277	398
Второй	258	507	234

падают наблюдения, для которых состояния высокого риска практически не возникают. Условно первый кластер можно назвать кластером проектов невысокого риска. Второй — кластером рискованных проектов. И, наконец, третий кластер — это безрисковые проекты. Для второго примера ситуация отличается. Все кластеры можно классифицировать как кластеры риска. При этом для третьего кластера модальным является состояние среднего риска. В данный кластер приходится почти треть всех проектов, что в целом соответствует исходным данным для этого примера и заданной матрице переходов случайного процесса за один шаг.

Отметим, что объединение сгенерированных выборок в одну обобщенную позволяет ввести дополнительные признаки, которые можно использовать для реше-

ния других задач аналитики. Введем дополнительный бинарный признак «риск проекта», который будет принимать нулевое значение, что соответствует значению этого признака — «низкий риск» или единичное значение, что будет соответствовать значению признака «средний риск». Аналогично можно сформировать и другие значения признака, перейдя на большое число категорий.

Для сравнительного анализа проектов двух категорий будем использовать энтропию, под которой понимается мера неопределенности, непредсказуемости процесса проектирования

$$H = -\sum_i p_i \log_2(p_i),$$

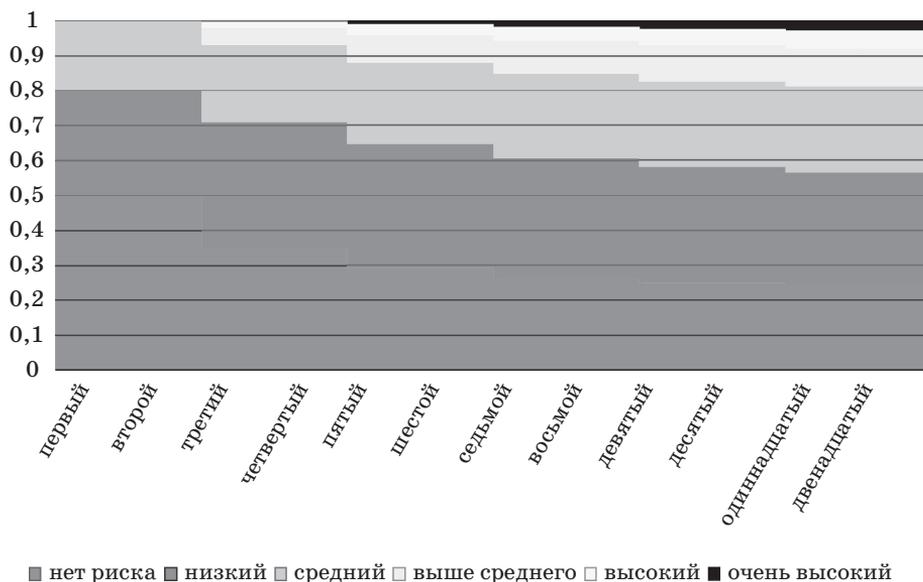
где p_i — вероятность нахождения в i -м состоянии.

Оценим так называемую горизонтальную энтропию (longitudinal, within-sequence entropy); в этом случае анализируется мера неопределенности каждого процесса.

Формула для оценки энтропии примет вид:

$$H = -\sum_{i=1}^{n=6} p_i \log_2(p_i).$$

На рис. 8 приведены две зависимости энтропии процесса от этапа проектирования.



а)

б)

Рис. 8. Зависимость энтропии от этапа проекта: а) первый пример; б) второй пример
Fig. 8. Dependence of entropy on the project stage: a) the first example; b) second example

Приведенные диаграммы показывают, что энтропия, определяющая неопределенность, риск выполнения проекта, растет. При этом степень неопределенности для второго примера выше, чем для первого примера. То есть с увеличением рисков выполнения проекта растет и энтропия. Таким образом, данный показатель может быть использован для оценки рисков проекта.

Решение задачи парного регрессионного анализа позволяет получить статистическую модель:

$$H = 0,57 + 0,08x,$$

где x — показатель сложности проекта.

Однофакторный дисперсионный анализ показывает, что существует значимый эффект показателя сложности на энтропию проекта:

```
> anova(lm(Entropy ~ Сложность, data = dataproj))
Analysis of Variance Table
Response: Entropy
  Df Sum Sq Mean Sq F value Pr(>F)
Сложность 1 3.511 3.5115 112.93 < 2.2e-16 ***
Residuals 1996 62.063 0.0311
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Наблюдаемое значение критерия Фишера, равное 112,93, определяет очень малое значение уровня значимости p-value данного критерия.

Обсуждение

Результаты выполненного анализа показывают, что ретроспективный анализ результатов выполнения проектов позволяет сделать выводы о проблемах, возникающих при проектировании. Для их решения следует использовать методы машинного обучения, например, методы кластерного анализа. При этом, при решении задачи группирования существует возможность выполнения сравнительного анализа траектории риска для различных значений факторов, определяющий организацию выполнения проекта (например, для разных этапов времени выполнения проекта, для разных проектировщиков, для разных типов выполняемых проектов, для разной технологии проектирования). Получаемая информация позволяет скорректировать опыт проектирования, выбрать более добросовестных проектировщиков, вовремя реагировать на возможные риски с целью управления ими. Данное обстоятельство определяет необходимость не только оперативного управления рисками при выполнении проектами, но и накопления, агрегирования опыта выполнения проекта. Постоянный мониторинг хода и результатов выполнения проектов позволит создать наборы данных менеджмента рисков, и, следовательно, использовать методы машинного управления в риск-менеджменте. Внедрение таких методов позволяет повысить качество управления проектами.

Наличие данных о выполненных проектах позволяет реализовать индуктивный метод оценки рисков проектов, обобщать результаты выполненных проектов, получать новые знания о проектировании информационных проектов. В теоретическом смысле такую задачу можно рассматривать как задачу построения скрытых марковских цепей. В этом случае наблюдаемая случайная последовательность состояния рисков выполнения проекта ведет себя как марковская цепь с неизвестными параметрами. Задача прогнозирования на таких графах может быть рассмотрена как задача идентификации данных параметров.

Выводы

Отметим, что не только проектное управление является средой исследования и использования случайных последовательностей событий. Такие последовательности целесообразно использовать и при исследовании других случайных процессов, число состояний которых конечно, а время может быть представлено дискретной последовательностью. В том случае, если ее фрагменты можно рассматривать как логические ассоциации, возможно формирование логических последовательностей, позволяющих прогнозировать будущие состояния анализируемого процесса.

Литература

1. *Бирюкова С.* Анализ последовательностей в R: TraMineR [Электронный ресурс]. М. : НИУ ВШЭ, 2014. URL: <https://www.hse.ru> (дата обращения: 02.01.2021).
2. *Горбач А. Н., Цейтлин Н. А.* Покупательское поведение: анализ спонтанных последовательностей и регрессионных моделей в маркетинговых исследованиях. Киев : Освіта України, 2011. 220 с.
3. *Каталевский Д. Ю.* Основы имитационного моделирования и системного анализа в управлении : учеб. пособие. М. : Изд-во Московского университета, 2011.
4. *Кучеренко Д. В., Матвеев А. В.* Анализ структуры проектных организаций государственных информационных систем органов исполнительной власти Санкт-Петербурга // Материалы II международной научно-практической конференции «Актуальные теоретические и прикладные вопросы управления социально-экономическими системами». Москва, 10–11 декабря 2020. М. : Институт развития дополнительного профессионального образования, 2020. С. 270–274.
5. *Левенштейн В. И.* Двоичные коды с исправлением выпадений, вставок и замещений символов // Доклады Академии наук СССР. 1965. № 4. С. 845–848.
6. *Мастицкий С. Э., Шитиков В. К.* Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс]. URL <https://github.com/ranalytics/data-mining> (дата обращения: 02.01.2021).
7. *Наумов В. Н., Кучеренко Д. В.* Планирование трудозатрат на создание и развитие государственных информационных систем и их функциональных программных комплексов // Управленческое консультирование. 2017. № 6. С. 140–154.
8. *Наумов В. Н., Наумов П. В.* Прогнозирования рисков проектирования информационных систем // Автоматизация процессов управления. 2019. № 4. С. 34–44.
9. *Павлов А. Н.* Эффективное управление проектами на основе стандарта PMI PMBOK 6 Edition. М. : Лаборатория знаний, 2019.
10. *Форрестер Дж.* Мировая динамика. СПб. : Terra Fantastica, 2003.
11. *Форрестер Дж.* Основы кибернетики предприятия (индустриальная динамика). М. : Прогресс, 1971.
12. *PMBOK.* Руководство к Своду знаний по управлению проектами. 6-е изд. М. : Олимп-бизнес, 2016.
13. *Buettner D.* Dynamic COQUALMO: Defect profiling over development cycles [Электронный ресурс] // Conference: Proceedings of the International Conference on Software Process: Trustworthy Software Development Processes. URL: https://www.researchgate.net/publication/225240099_Dynamic_COQUALMO_Defect_profiling_over_development_cycles/ (дата обращения: 07.01.2021).
14. *Chulani S., Boehm D.* Modeling Software Defect Introduction and Removal: COQUALMO (CONstructive QUALity MOdel) [Электронный ресурс]. URL: https://www.researchgate.net/publication/228778578_Modeling_software_defect_introduction_and_roval_COQUALMO_CONstructive_QUALity_MOdel/ (дата обращения: 06.01.2021).
15. *Gabadinho A., Ritschard G., Studer M., Müller N. S.* Mining Sequence Data in R with the TraMineR package: A user's guide. Department of Econometrics and Laboratory of Demography, University of Geneva [Электронный ресурс]. URL: https://www.researchgate.net/publication/342889620_Mining_sequence_data_in_R_with_the_TraMineR_package_A_user's_guide/ (дата обращения: 5.01.2021).
16. *Gabadinho A., Ritschard G., Müller N. S., Studer M.* Analyzing and Visualizing State Sequences in R with TraMineR [Электронный ресурс] // Journal of Statistical Software 2011. № 40 (4). P. 1–37.

17. Mayer D. The Complexity of Some Problems on Subsequences and Super sequences // Journal of the ACM. Vol. 25, Iss. 2. April 1978. P. 322–336.
18. *Sequence Analysis in Demographic Research* Special Issue on Longitudinal Methodology [Электронный ресурс] // Canadian Studies in Population. 2001. Vol. 28(2). P. 439–458.
19. *Software Project Risk Assessment and Efforts Contingency Model based on COCOMO Cost Factor COCOMO*. [Электронный ресурс]. URL: <https://www.researchgate.net/publication/236679817> (дата обращения: 28.09.2019).
20. Studer M., Ritschard G. A Comparative Review of Sequence Dissimilarity Measures [Электронный ресурс] // LIVES Working Papers, 33. NCCR LIVES, Switzerland, 2014. URL: https://www.lives-nccr.ch/sites/default/files/pdf/publication/33_lives_wp_studer_sequencedissmeasures.pdf (дата обращения: 03.01.2020).
21. Studer M., Ritschard G. What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures // Journal of the Royal Statistical Society, Series A. 2016. Vol. 179, Iss. 2. P. 481–511.
22. Valerdi R., Boehm Barry W. COSYSMO: A system engineering cost model [Электронный ресурс]. URL: <http://csse.usc.edu/TECHRPTS/2010/usc-csse-2010-508/usc-csse-2010-508.pdf> (дата обращения: 20.09.2019).
23. Valerdi R., John E. Gaffney Jr. Reducing Risk and Uncertainty in COSYSMO Size and Cost Drivers: Some Techniques for Enhancing Accuracy. 2007. [Электронный ресурс]. URL: <https://www.researchgate.net/publication/229034929> Reducing_Risk_and_Uncertainty_in_COSYSMO_Size_and_Cost_Drivers_Some_Techniques_for_Enhancing_Accuracy (дата обращения: 28.09.2019).
24. Wagner R. A., Fischer M. J. The string-to-string correction problem // Journal of the ACM. 1974. Vol. 21, Iss. 1. P. 168–173.

Об авторах:

Наумов Павел Владимирович, аспирант Северо-Западного института управления РАНХиГС (Санкт-Петербург, Российская Федерация); naumov122@gmail.com

References

1. Biryukova S. Sequence analysis in R: TraMineR [Electronic resource]. M.: HSE, 2014. URL: <https://www.hse.ru> (access date: 02.01.2021). (In Rus).
2. Gorbach A. N., Zeitlin N. A. Purchasing behavior: analysis of spontaneous sequences and regression models in marketing research. Kiev: Education of Ukraine, 2011. 220 p. (In Rus).
3. Katalovsky D. Yu. Fundamentals of simulation modeling and system analysis in management: tutorial. M.: Publishing House of Moscow University, 2011. (In Rus).
4. Kucherenko D. V., Matveev A. V. Analysis of the structure of design organizations of state information systems of executive authorities of St. Petersburg // Materials of the II International Scientific and Practical Conference «Topical Theoretical and Applied Issues of Management of Socio-Economic Systems». Moscow, December 10-11, 2020. M.: Institute for the Development of Vocational Education, 2020. P. 270–274. (In Rus).
5. Levenstein V. I. Binary codes with correction of fallout, insertions and substitutions of symbols // Reports of Academy of Sciences of the USSR [Doklady Akademii Nauk SSSR]. 1965. N 4. P. 845–848. (In Rus).
6. Mastitsky S. E., Shitikov V. K. Classification, regression and other Data Mining algorithms using R [Electronic resource]. URL <https://github.com/ranalytics/data-mining> (access date: 02.01.2021). (In Rus).
7. Naumov V. N., Kucherenko D. V. Planning labor costs for the creation and development of state information systems and their functional software complexes // Administrative consulting [Upravlencheskoe konsul'tirovanie]. 2017. N 6. P. 140–154. (In Rus).
8. Naumov V. N., Naumov P. V. Forecasting risks of information systems design // Automation of management processes [Avtomatizatsiya protsessov upravleniya]. 2019. N 4. P. 34–44. (In Rus).
9. Pavlov A. N. Effective project management based on the PMI PMBOK 6 Edition standard. M.: Knowledge Laboratory, 2019. (In Rus).
10. Forrester J. World dynamics. St. Petersburg: Terra Fantastica, 2003. (In Rus).
11. Forrester J. Fundamentals of cybernetics enterprise (industrial dynamics). M.: Progress, 1971. (In Rus).
12. PMBOK. Guide to Project Management Knowledge. 6th ed. M.: Olympus business, 2016. (In Rus).

13. Buettner D. Dynamic COQUALMO: Defect profiling over development cycles [Electronic resource] // Conference: Proceedings of the International Conference on Software Process: Trustworthy Software Development Processes. URL: https://www.researchgate.net/publication/225240099_Dynamic_COQUALMO_Defect_profiling_over_development_cycles/ (дата обращения: 07.01.2021).
14. Chulani S., Boehm D. Modeling Software Defect Introduction and Removal: COQUALMO (CONstructive QUALity MOdel) [Electronic resource]. URL: https://www.researchgate.net/publication/228778578_Modeling_software_defect_introduction_and_removal_COQUALMO_CONstructive_QUALity_MOdel/ (дата обращения: 06.01.2021).
15. Gabadinho A., Ritschard G., Studer M., Müller N. S. Mining Sequence Data in R with the TraMineR package: A user's guide. Department of Econometrics and Laboratory of Demography, University of Geneva [Electronic resource]. URL: https://www.researchgate.net/publication/342889620_Mining_sequence_data_in_R_with_the_TraMineR_package_A_user's_guide/ (дата обращения: 5.01.2021).
16. Gabadinho A., Ritschard G., Müller N. S., Studer M. Analyzing and Visualizing State Sequences in R with TraMineR [Electronic resource] // Journal of Statistical Software 2011. № 40 (4). P. 1–37.
17. Mayer D. The Complexity of Some Problems on Subsequences and Super sequences // Journal of the ACM. Vol. 25, Iss. 2. April 1978. P. 322–336.
18. Sequence Analysis in Demographic Research Special Issue on Longitudinal Methodology [Electronic resource] // Canadian Studies in Population. 2001. Vol. 28(2). P. 439–458.
19. Software Project Risk Assessment and Efforts Contingency Model based on COCOMO Cost Factor COCOMO. [Electronic resource]. URL: <https://www.researchgate.net/publication/236679817> (дата обращения: 28.09.2019).
20. Studer M., Ritschard G. A Comparative Review of Sequence Dissimilarity Measures [Electronic resource] // LIVES Working Papers, 33. NCCR LIVES, Switzerland, 2014. URL: https://www.lives-nccr.ch/sites/default/files/pdf/publication/33_lives_wp_studer_sequencedissmeasures.pdf (дата обращения: 03.01.2020).
21. Studer M., Ritschard G. What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures // Journal of the Royal Statistical Society, Series A. 2016. Vol. 179, Iss. 2. P. 481–511.
22. Valerdi R., Boehm Barry W. COSYSMO: A system engineering cost model [Electronic resource]. URL: <http://csse.usc.edu/TECHRPTS/2010/usc-csse-2010-508/usc-csse-2010-508.pdf> (дата обращения: 20.09.2019).
23. Valerdi R., John E. Gaffney Jr. Reducing Risk and Uncertainty in COSYSMO Size and Cost Drivers: Some Techniques for Enhancing Accuracy. 2007. [Electronic resource]. URL: https://www.researchgate.net/publication/229034929_Reducing_Risk_and_Uncertainty_in_COSYSMO_Size_and_Cost_Drivers_Some_Techniques_for_Enhancing_Accuracy (дата обращения: 28.09.2019).
24. Wagner R. A., Fischer M. J. The string-to-string correction problem // Journal of the ACM. 1974. Vol. 21. Iss. 1. P. 168–173.

About the author:

Pavel V. Naumov, Postgraduate Student of North-West Institute of Management, Branch of RANEPA (St. Petersburg, Russian Federation); naumov122@gmail.com