

С. А. ДЕРЖАВИН

Северо-западный институт управления - Российская академия народного хозяйства и государственной службы при Президенте РФ, Санкт-Петербург

А. С. ГЕЙДА

Северо-западный институт управления - Российская академия народного хозяйства и государственной службы при Президенте РФ; Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург

И. П. КОЛОСОВ, В. С. РЕЗАНОВА

Северо-западный институт управления - Российская академия народного хозяйства и государственной службы при Президенте РФ, Санкт-Петербург

МАЙНИНГ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИИ ДЛЯ ФУНКЦИОНИРОВАНИЯ СИСТЕМ

Цифровая трансформация экономики и общества вызывает много задач, решаемых практиками и теоретиками при проектировании и использовании информационных технологий. Для успешного решения многих из этих задач требуется исследование результатов использования информационных технологий еще на стадии проектирования систем, в которых эти технологии используются, на математических моделях, описывающих будущее использование информационных технологий при функционировании систем. Однако, прогнозные математические модели, связывающие показатели успешности использования информационных технологий с характеристиками этих технологий и характеристиками функционирования систем пока еще не разработаны в должной мере. Недостаточно подробно описаны концептуальные и методологические основания такого моделирования. Моделирование использования информации связано с построением значительного числа возможных последовательностей причинно-следственных связей информационных и других действий, вызываемых ими цепочек событий и состояний при функционировании. Частично такие цепочки могут быть получены за счет майнинга процессов. Однако, множество всех возможных цепочек может быть получено на этапе проектирования лишь частично, в связи с тем, что такие цепочки могут быть новыми. Для получения множеств возможных цепочек действий, событий и состояний предложено разработать технологии синтеза множеств возможных цепочек действий, событий и состояний. Представленный в докладе материал должен позволить преодолеть имеющееся несоответствие и перейти к построению требуемых моделей цепочек действий, событий и состояний на этапе проектирования.

Введение. При реализации цифровой трансформации многих современных систем возникают разнообразные задачи проектирования и совершенствования использования цифровых информационных технологий (ИТ). Так, например, оказывается необходимым соотносить затраты на внедрение цифровых технологий сначала с характеристиками разрабатываемых и внедряемых ИТ, а затем с теми результатами, которые могут быть получены при использовании внедренных ИТ и наконец – с финансовыми результатами, достигаемыми благодаря использованию цифровых ИТ [1, 2, 3].

В ранее опубликованных авторами трудах [4, 5, 6] показано, что для получения математических моделей, позволяющих описать связи характеристик разрабатываемых и используемых ИТ с получаемыми результатами использования в будущей практике необходимо породить возможные последовательности связанных причинно-следственными связями действий, событий и состояний в зависимости от получаемой, создаваемой и используемой информации о функционировании системы и ее среды.

Модели использования информации для функционирования системы. Возможные последовательности связанных причинно-следственными связями действий, событий и состояний [7] могут быть структурированы в виде теоретико-графовых моделей разного вида, а затем – функциональных и программных моделей использования информации при функционировании системы. Они дают далее возможность оценить результаты отдельных реализаций функциони-

рования в изменяющихся условиях при соответствующих реализациях применения ИТ и меры возможности таких реализаций. Это, затем, дает возможность оценить показатели потенциала системы, функционирующей с использованием ИТ, в изменяющихся условиях, и показатели, описывающие изменение энтропии при использовании ИТ – синтропию систем [8]. Эти показатели, оцениваемые в зависимости от условий функционирования и от характеристик используемых ИТ, позволяют далее перейти к функциональным моделям, описывающим связи характеристик разрабатываемых и используемых ИТ с получаемыми результатами использования в будущей практике.

Такие зависимости, в свою очередь, позволяют решать значительное число практических задач, в том числе – задач цифровизации, как математические задачи исследования операций, математического программирования [9].

Основная сложность при таком подходе к решению задач цифровой трансформации состоит в построении необходимых моделей, опирающихся на представление возможных последовательностей причинно-следственных связей между действиями разных видов, информацией об условиях, результатах и реализации действий, событиями в и результате и состояниями. Такие модели могут обладать разными особенностями и их можно порождать разными методами. Рассмотрим особенности предложенных авторами моделей использования информации для функционирования систем, опирающихся на представление в виде цепочек причинно-следственных связей.

Особенности моделей использования информации для функционирования систем, опирающихся на представление в виде цепочек причинно-следственных связей. Нами предложены модели, опирающиеся на представления последовательностей причинно-следственных связей между действиями разных видов, информацией об условиях, результатах и реализации действий, событиями в их результате и состояниями («цепочки причинно-следственных связей»). Такие модели предназначены далее для связывания характеристик информационных технологий и систем, в которых они используются, с показателями, характеризующими разнообразные стороны успешности использования ИТ и для расчета этих показателей в зависимости от характеристик ИТ. При этом важной особенностью указанных моделей является то, что вычисления по этим моделям реализуются в соответствии с закономерностями природных явлений (по причинно-следственным связям), результаты вычислений имитируют результаты реализации того или иного функционирования системы в тех или иных условиях (по тем же причинам, модель обеспечивает подобие функционированию). Кроме того, модель информационной операции подобна информационной операции (и может непосредственно использоваться в системе при функционировании). Наконец, вычисление результатов функционирования задается последовательностью элементов в цепочках модели, а цепочки могут храниться в последовательности вычислений.

Несмотря на то, что модели указанного вида обладают указанными достоинствами, их построение связано с существенными затруднениями, вызванными масштабностью множества возможных цепочек, неопределенностями и случайностями при их формировании, сложностями корректного описания причинно-следственных связей в цепочках так, чтобы можно было достаточно просто перейти к вычислительным моделям.

В связи с этим остро встает вопрос о методах порождения моделей «цепочек» и методах автоматизации их порождения.

Методы порождения моделей использования информации для функционирования систем. Рассмотрим ряд основных методов порождения моделей, опирающихся на представления последовательностей причинно-следственных связей между действиями разных видов, информацией об условиях, результатах и реализации действий, событиями в их результате и состояниями (моделями «цепочек»).

1. Метод непосредственного построения модели экспертом. В соответствии с этим методом, цепочки порождаются специалистом по реализации процессов, хорошо знающим особенности моделируемых процессов. Метод трудоемок, однако позволяет описывать все элементы модели (в том числе информационные) подробно. При этом, однако, могут быть пропущены некоторые реализации цепочек в результате ошибки эксперта.

2. Метод порождения модели по заданным экспертом мета-описаниям (мета-модели) и алгоритмам порождения. В соответствии с этим методом, цепочки порождаются не непосред-

ственно, а путем описания алгоритма их порождения, на основе некоторой мета-модели, описывающей сведения о совокупностях подобных элементов строящихся моделей в более сжатом виде (по отношению к строящейся модели) и алгоритма (правил) порождения модели. Так, например, путем обходов диаграммы, задающей мета-модель в виде графа, описывающего возможные реализации последовательностей причинно-следственных связей, которые могут перечисляться (для порождения модели) при обходе такой мета-модели с соблюдением заданных правил порождения. Метод характеризуется тем, что часть действий может быть автоматизирована. Однако он более сложен в конструировании корректной мета-модели и алгоритма. При его реализации алгоритм, если он корректен, гарантирует перечисление всех возможных цепочек. Описание информационных действий должно быть формализовано, что может вызвать сложности.

Указанные два метода порождения моделей – методы порождения человеком моделей применения информации в деятельности. Их отличие в том, что человеком в разных формах описываются возможные последовательности изменения (альтернирования) деятельности в различных условиях, в зависимости от выполняемых информационных действий. При этом при описании в форме мета-модели применение информации и возможности альтернирования необходимо формализовать, что и дает возможность поставить между человеком и моделью вычислительное устройство, на котором может быть выполнена часть действий по моделированию.

3. Майнинг процессов. В соответствии с этим методом log-файлы исследуются на предмет выявления цепочек причинно-следственных связей. По ним порождаются реализации цепочек (traces), которые затем следует систематизировать, породив модели процессов. Недостаток метода состоит в том, что в log файлах могут быть только те реализации процессов, которые были наблюдаемы. Однако это не все возможные реализации. Кроме того, на этапе проектирования никаких log-файлов функционирования проектируемой системы еще нет. Наконец, существующие log-файлы редко содержат сведения о выполненных информационных действиях, необходимые для майнинга использования информации для функционирования систем («майнинга использования информации»).

4. Метод обучения по log-файлам алгоритма непосредственного построения моделей на основе порождаемых мета-моделей. Используя как имеющиеся построенные экспертами мета-модели, так и полученные из log-файлов реализации процессов следует уточнить мета-модель использования информации и, на основе log-файлов обучить алгоритм, порождающий возможные цепочки в зависимости от условий. Алгоритм должен порождать корректные (с точки зрения эксперта) модели процессов в виде цепочек реализаций. В этом случае также говорят о майнинге использования информации. И в этом случае необходимо моделировать использование информации в log-файлах.

Указанные два метода порождения моделей – методы порождения моделей применения информации в деятельности из log-файлов. Их отличие в том, что последовательности изменения (альтернирования) деятельности в различных условиях заданы в log-файлах и в моделях использования информации при функционировании систем, построенных людьми. При этом, в настоящее время выстраивание цепочек деятельности в зависимости от выполняемых информационных действий не распространено достаточно широко и, как правило, игнорируется. Описывается только результат. Не практикуется формализация применения информации с использованием мета-моделей.

5. Метод обучения мета-модели и алгоритмов порождения. В рамках этого метода мета-модель и алгоритм ее использования для порождения модели создается машиной на основе данных, используемых и человеком, и машиной (в том числе и log-файлов) и аналогично тому, как это делает человек. Предполагается, что структура и характеристики информационных действий не меняются при машинном обучении, информационные действия заданы. Тем не менее, такие алгоритмы нам неизвестны.

6. Метод майнинга информационных действий. Метод состоит в обучении моделей информационных действий так, чтобы они реализовывали лучшие из возможных (по соответствию эффектов требованиям в возможных условиях) реакции на изменения системы и ее среды. При этом в качестве данных используются реализованные информационные действия,

модели и мета-модели процессов, описанные экспертами и log-файлы с записями о характеристиках информационных действий.

Указанные два метода порождения моделей – перспективные методы машинного порождения моделей применения информации в деятельности. Модели создаются машиной на основе данных, используемых и человеком, и машиной и аналогично тому, как это делает человек. Их отличие в том, что на основе как моделей, построенных человеком, так и log-файлов применяются алгоритмы создания возможных последовательностей изменения (альтернирования) деятельности, в зависимости от возможных выполняемых информационных действий в различных условиях и от характеристик этих действий. При этом применение информации и возможности альтернирования формализуются для того, чтобы обеспечить их формирование разрабатываемым алгоритмом на основе уже имеющихся моделей (машинное обучение созданию моделей использования информации для деятельности).

7. Комплексные методы. Такие методы представляют собой систему методов 1–6, сформированную таким образом, чтобы с ее использованием могли создаваться требуемые модели использования информации для функционирования систем.

Заключение. Предложены направления построения и совершенствования прогнозных математических моделей, связывающие показатели успешности использования информационных технологий с характеристиками этих технологий и характеристиками функционирования систем. Описаны концептуальные и методологические трудности, связанные с таким моделированием. Показано, что моделирование использования информации связано с построением значительного числа возможных последовательностей причинно-следственных связей информационных и других действий, вызываемых ими цепочек событий и состояний при функционировании. Частично такие цепочки могут быть получены за счет майнинга процессов. Однако, множество всех возможных цепочек может быть получено на этапе проектирования лишь частично, в связи с тем, что такие цепочки могут быть новыми. Для получения множеств возможных цепочек действий, событий и состояний предложено разработать ряд концепций, методов и технологий моделирования множеств возможных цепочек действий, событий и состояний с использованием информационных действий. В докладе показаны основные направления создания таких моделей, в том числе и с использованием машинного обучения. Описано новое направление моделирования использования информации, представляющее собой майнинг информационных действий при функционировании систем и машинное обучение характеристик таких действий по имеющимся сведениям. Представленный в докладе материал должен позволить преодолеть имеющееся несоответствие и перейти к построению требуемых моделей использования информации для функционирования систем.

Работа проводилась при поддержке бюджетной НИР FFZF-2022–0003

ЛИТЕРАТУРА

1. Юсупов Р.М., Мусаев А.А. Проблема оценивания эффективности информационных технологий. Материалы конференции «Информационные технологии в управлении (ИТУ-2018)». Электроприбор. 2018.
2. Юсупов Р.М., Мусаев А.А. Особенности оценивания эффективности информационных систем и технологий. Тр. СПИИРАН. 51 (2017), с. 5–34.
3. Юсупов Р.М., Мусаев А.А. К оцениванию эффективности информационных систем. Методологические аспекты. Информационные технологии. 2017. Том 23. №5. С. 323–332.
4. Ашимов А.А., Гейда А.С., Лысенко И.В., & Юсупов Р.М. Эффективность функционирования и другие операционные свойства систем: задачи и метод оценивания. Труды СПИИРАН, 5(60), 2018. С. 241–270.
5. Гейда А.С., Гурьева Т.Н., Наумов В.Н. Концептуальные и математические модели, методы и технологии исследования цифровой трансформации экономических и социальных систем: обзор предметного поля (часть I). Управленческое консультирование. 2021. № 11. С. 95–108.
6. Гейда А.С., Гурьева Т.Н., Наумов В.Н. Концептуальные и математические модели, методы и технологии исследования цифровой трансформации экономических и социальных систем: обзор предметного поля (часть II). Управленческое консультирование. 2021. № 12. С. 111–125.
7. Reichert M., Weber B. Enabling Flexibility in Process-Aware Information Systems, Springer-Verlag Berlin Heidelberg 2012. 511 p.
8. Vyatkin V. Syntropic criterion for removing restrictions during the COVID-19 pandemic, Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета, № 174, 2021, pp. 426-435.
9. Гейда А.С. Основы теории потенциала сложных технических систем: монография. М.: РАН, 2021. 408 с.

S.A.Derzhavin, A.S.Geyda, I.P.Kolesov, V.S.Rezanova (The North-West Institute of management – branch of the Russian Presidential Academy of National Economy and Public Administration, St. Petersburg)

Information use mining for system functioning

The digital transformation of the economy and society caused many problems in the design and use of information technologies to be solved by practitioners and theorists. To successfully solve many of these problems, it is necessary to study the results of the use of information technologies at the design stage of the systems. Such results of the use of information technologies shall be studies on mathematical models predicting the future use of information technologies in the functioning of systems. However, predictive mathematical models linking indicators of the success of the use of information technologies with the characteristics of these technologies and the characteristics of the functioning of systems have not yet been adequately developed. The conceptual and methodological foundations of such modeling are not described in sufficient detail. Modeling the use of information is associated with the construction of a significant number of possible sequences of cause-and-effect relationships of information and other actions, chains of events and states caused by them during functioning. In part, such chains can be obtained by process mining. However, the set of all possible chains can be obtained by process mining only partially at the design stage, because such chains may be new, not ever recorded ones. To obtain sets of possible chains of actions, events, and states, it is proposed to develop new technologies for synthesizing sets of possible chains of actions, events and states. The research results, presented in the article should make it possible to overcome the existing gap and proceed to the construction of the required models of chains of actions, events and states at the design stage.

Авторы готовы представить текст на английском языке для сборника материалов мультиконференции, который будет подан для индексирования в Scopus.