



Информационные процессы в пространстве «больших данных»



Борис ЛЁВИН
Boris A. LYOVIN

Виктор ЦВЕТКОВ
Victor Ya. TSVETKOV



Лёвин Борис Алексеевич — доктор технических наук, профессор, ректор Российского университета транспорта (МИИТ), Москва, Россия.

Цветков Виктор Яковлевич — доктор технических наук, профессор, заместитель руководителя центра стратегического анализа и развития НИИАС, Москва, Россия.

Information Processes in Big Data Environment

(текст статьи на англ. яз. —
English text of the article — p. 27)

Статья описывает результаты исследования объектных и ситуационных моделей при управлении транспортом. Раскрываются смысл и содержание такого типа информационных моделей. Показано сходство и различие между ними. Рассматриваются особенности статических и динамических информационных ситуаций, их дифференциация по критериям взаимодействия и критериям состояния. Предложен метод использования модели информационной ситуации для оценки информационной неопределённости при принятии решений.

Ключевые слова: информация, транспорт, управление, моделирование, информационная модель объекта, информационная модель ситуации, информационная неопределённость, функция доверия, функция правдоподобия.

Современное управление транспортом характеризуется возрастанием объёмов информации, ростом сложности управленческих ситуаций и сокращением допустимого времени принятия решений, особенно для высокоскоростного транспорта. В определённой степени эти факторы являются следствием проблемы «больших данных» [1], характерной для разных направлений. Дополнительным фактором управления стала проблема уменьшения или устранения информационной неопределённости [2], которая имеет две причины. Первая причина неопределённости состоит в нехватке необходимой информации для принятия решения. Вторая — в чрезмерном избытке информации, которая требует значительного времени для анализа и нахождения полезного контента, что может быть сопоставимо с допустимым временем принятия решения.

Одним из радикальных для управления транспортом зарекомендовал себя информационный подход [3], включающий применение различных информационных моделей. Модели транспортных систем [4, 5] предназначены для выявле-

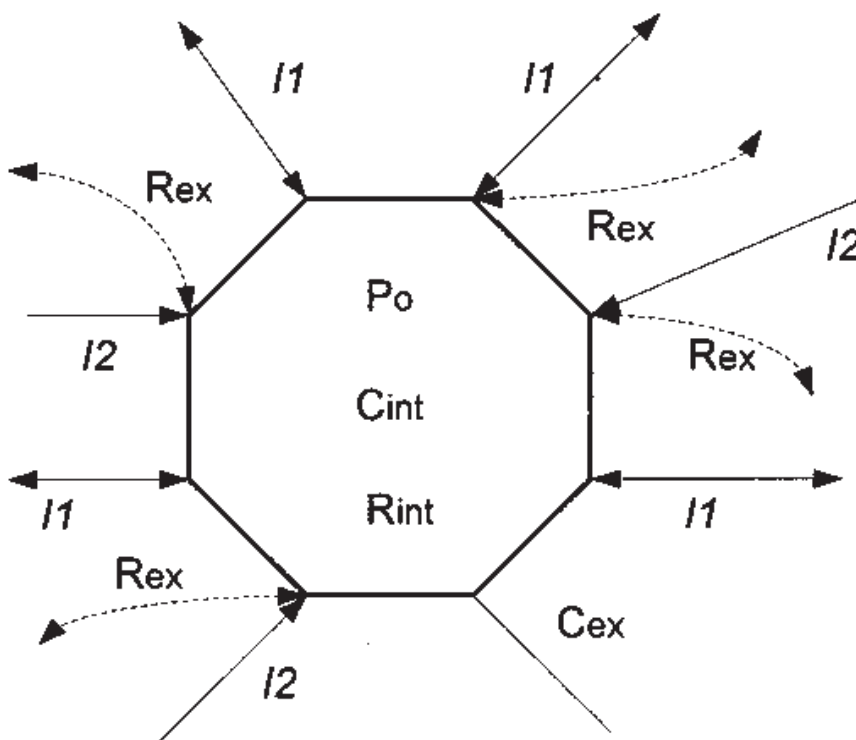


Рис. 1. Информационная модель объекта как совокупность связей и отношений с внешней средой.

ния резервов развития сферы транспортных услуг, повышения качества транспортной работы и прогнозирования будущего транспорта. Информационное моделирование не является простым переносом методов информатики и моделей в сферу транспорта. Оно требует новой организации информационных и электронных ресурсов и применения инновационных решений.

ВИДЫ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ

Объектные модели описывают объекты. Наиболее ярким представителем в информационной области является информационная модель объекта (ИМО). Ситуационные модели описывают ситуацию, в которой находятся объекты. Одной из таких стала информационная модель ситуации (ИМС), или информационной ситуации. Обе эти информационные модели близки между собой, но имеют и различия. Хотя не являются и единственными для описания объектов и ситуаций. Существуют обобщающие модели, например, модель информаци-

онной конструкции [6], которая при необходимости может описывать как объект, так и ситуацию.

Основное различие между ИМО и ИМС – масштаб действия. Общее у них то, что они производное от понятия «информационная модель».

Информационная модель объекта (ИМО) [7] определяется как взаимосвязанная совокупность параметров, наиболее важных связей и отношений. Термин «наиболее важных» означает, что в модель включаются существенные связи и отношения, а несущественные исключаются. Это общее свойство любых моделей, в том числе и неинформационных. Формальное описание ИМО приведено в выражении:

$$\text{ИМО} = \text{FI}(\text{Po}, \text{Cint}, \text{Cex}, \text{Rint}, \text{Rex}, \text{I1}, \text{I2}, \text{SO}), \quad (1)$$

где Po – параметры объекта, Cint – внутренние связи между частями объекта, Cex – внешние связи с другими объектами и со средой, Rint – внутренние информационные отношения между частями объекта, Rex – внешние информационные отношения, I1 – информационные



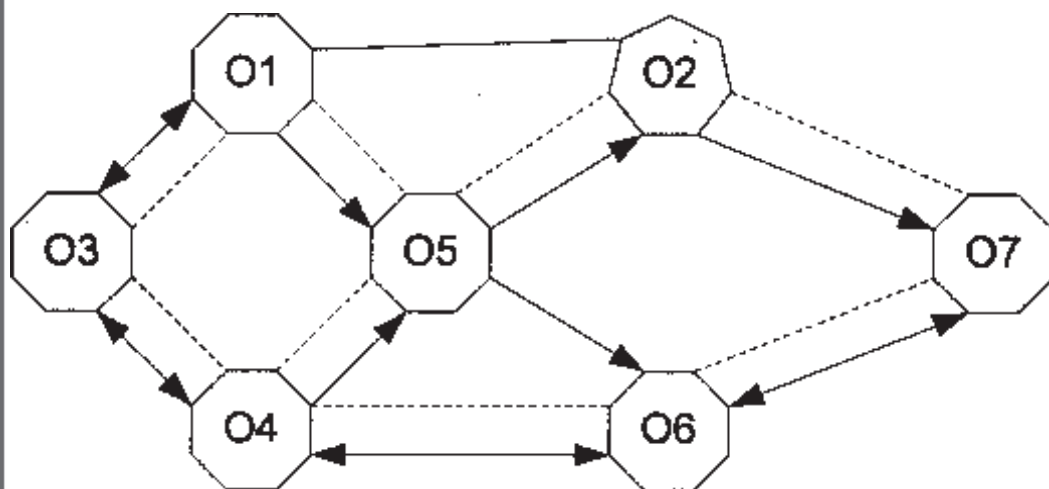


Рис. 2. Пример закрытой информационной ситуации по взаимодействиям.

взаимодействия объекта с другими объектами и со средой, I2 – информационные воздействия на объект, SO – системность объекта (необязательное свойство). Объект может быть частью другого объекта как более сложной системы или может быть самостоятельной системой, обладающей целостностью и системностью. В последнем случае появляется свойство системности объекта SO.

На рис. 1 приведена информационная модель объекта. Внутренняя структура не показана, а обозначена через Po, Cint, Rint. Связи и взаимодействия показаны сплошными линиями, отношения – пунктирными. Взаимодействия – двойными стрелками, воздействия – односторонними стрелками. Связи без взаимодействия – прямыми линиями без стрелок.

Область существенного влияния на объект выделена. Остальные объекты в этой модели отсутствуют и показаны через внешние связи и отношения. Целевое назначение ИМО – описание индивидуального объекта.

Информационная модель ситуации (ИМС) – взаимосвязанная совокупность параметров, наиболее важных связей и отношений для данной ситуации. Целевое назначение ИМС – описание качественно разных ситуаций: взаимодействия объектов, поведения одного или совокупности объектов в данной ситуации, динамики ситуации безотносительно к объектам. Ситуация имеет всегда больший масштаб, чем модель объекта. Информационная ситуация более разнообразна, чем ИМО. У неё предметная

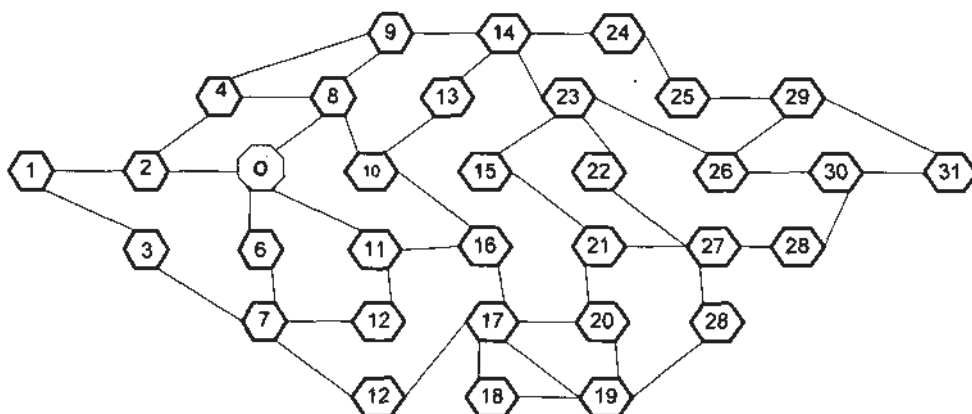


Рис. 3. Информационная ситуация по состояниям.

ориентация. Например, информационная ситуация взаимодействия объектов, информационная ситуация движения объекта, информационная ситуация состояния объекта. Формально модель ИМС приведена в выражении:

$ИМС = F_2(Ps, Co, Cp, Ros, Rps, IS1, IS2, S)$, (2)
где Ps — параметры ситуации, Co — связи между объектами, Cp — связи между параметрами объекта, Ros — отношения между объектами, Rps — отношения между параметрами, $IS1$ — информационные взаимодействия в ситуации между объектами, $IS2$ — информационные воздействия в ситуации, S — системность ситуации. На рис. 2 приведён пример информационной ситуации по взаимодействиям. На нём условно показаны семь объектов.

Объекты обозначены через O_i , отношения показаны пунктиром, связи и взаимодействия — сплошными линиями, отношения — пунктирными. Взаимодействия — двойными стрелками, воздействия — односторонними стрелками. Закрытая информационная ситуация чаще всего обладает свойством системности и может быть рассмотрена как сложная система со всеми системными свойствами. Отношения дополняют состояния. Это могут быть отношения иерархии, отношения эквивалентности и др.

На рис. 3 приведена информационная ситуация по состояниям. Связи между состояниями показаны непрерывными линиями. Состояния обозначены шестиугольниками и не заштрихованы. Объект (O) — восьмиугольником и заштрихован.

Рисунок может предполагать множество ситуаций. Например: поезд находится на станции отправления (условно состояние 1); поезд — на промежуточной станции (условно состояние 5 или O); поезд — на станции прибытия (условно состояние 31).

Информационная ситуация по состояниям описывает один объект, перемещающийся по возможным фиксированным состояниям. Она строится, когда есть начальное и целевое состояние. Если цель не определена чётко, а имеется некая целевая парадигма, то от управления по состояниям переходят к управлению по позициям [8]. Пример такой ситуации связан с рынком, когда одной из парадигм

управления является обеспечение конкурентоспособности. В этом случае проводят сравнительный анализ и определяют позицию объекта в информационной ситуации. Сравнивая позицию объекта с позицией других объектов, вырабатывают стратегию улучшения позиции объекта « O » с учётом изменения остальных позиций. Такое управление будет динамическим и требует применения динамической модели ситуации, термин «позиция» употребляется в двух значениях. Пространственная позиция, которая описывает перемещение объекта в пространстве, и параметрическая позиция, которая характеризует позицию объекта по выбранному критерию, например конкурентоспособность или надёжность.

Для полноты рассмотрения транспортных объектов необходимо применять динамические модели. Перемещение транспорта осуществляется по транспортной сети, аналог которой приведён на рис. 3.

Динамические модели включают фактор времени, по этой причине именно они служат основой управления как временного процесса [9, 10]. При перемещении ИМО, как правило, стационарна по внутренним характеристикам. Динамика проявляется в первую очередь на модели информационной ситуации:

$$ИМС(t) = F_3 [Ps(t), Co(t), Cp(t), Ros(t), Rps(t)]. \quad (3)$$

В процессе динамики информационная ситуация рис. 3 характеризует перемещение объекта, изменение его состояния и позиции. Часто при таком анализе применяют топологические модели [11], которые решают не только задачи выбора пути, но и оценивают риск перемещения, текущую и итоговую стоимость перевозки.

При моделировании следует принимать во внимание, что объекты транспорта и транспортная инфраструктура находятся в реальном пространстве. Поэтому для построения информационных моделей необходимо использовать пространственную информацию и пространственные модели [12]. В отдельных случаях для моделирования управления нужны пространственные знания [13]. Включение пространственных факторов в моделирование делает востребованным



ми методы геоинформатики и геоинформационных моделей [14]. Кроме того, современное пространственное моделирование широко использует космические технологии. В силу этого применяют интеграцию технологий дистанционного зондирования и геоинформатики.

ИНСТРУМЕНТ ОЦЕНКИ НЕОПРЕДЕЛЁННОСТИ

Причинами информационной неопределённости могут быть разные факторы, но из них можно выделить три, из которых два диаметрально противоположные: недостаток информации; избыточность информации, которая в когнитивных терминах проявляется как «необозримость» и «невосприимчивость». Третий фактор обусловлен развитием высокоскоростного транспорта и состоит в требовании сокращения времени на принятие решений.

Эти все причины можно назвать объективными. Дополнительно к ним появилась ещё одна — повышение информационной нагрузки на руководителей, рост риска, обусловленного «человеческим фактором». Информационная неопределённость мотивирована разными «не факторами» [16]: неосведомлённость, неинформированность, неистинность информации (заблуждение), неадекватность моделирования.

Неосведомлённость и неинформированность близкие понятия, но между ними есть различие. Неосведомлённость больше носит субъективный оттенок. Она может возникать при наличии осведомляющей информации, но не полученной принимающими решения по субъективным причинам. Неинформированность же обусловлена лишь отсутствием информации.

Неистинность информации (заблуждение) связана с наличием «правдоподобной», но не до конца точной информации. Целью информационного взаимодействия и диагностики, как и любого научного исследования, является познание истины. Однако вследствие неправильных исходных посылок, неправильной трактовки условий, логических ошибок и т.п. — результатом информационного процесса может быть заблуждение. Под заблуждением [17] обычно понимают определённый

вид ложных высказываний, отличающихся от прочих тем, что такое ложное высказывание принимается за истинное.

Проблемы моделирования, нацеленного на устранение неопределённости, проистекают в основном от того, что информационная неопределённость обусловлена неадекватностью моделей, отражающих реальную ситуацию. В настоящее время информационное моделирование стало более точным инструментом, позволяющим создавать качественно разные модели: ситуаций, процессов, объектов, явлений. Это помогает эффективно устранять неопределённость.

Рассмотрим ситуационные модели, содержащие неопределённость. Простейшая: «Поезд опаздывает на станцию прибытия». Это констатирующая модель, которая содержит неопределённость без какой-либо количественной меры.

На практике возможно применение моделей, содержащих количественные оценки неопределённости. Например, поезд прибывает вовремя на станцию с вероятностью 0,5; другая ситуация — поезд опаздывает на станцию прибытия с вероятностью 0,15. Эти варианты разнородны и вместе создают сложную информационную ситуацию.

Сложная информационная ситуация содержит истинную и не истинную информацию. Для обработки такой информации применяют теорию Демпстера—Шафера (ТДШ) [18]. Это математическая теория свидетельств, основанная на функции доверия (belief functions — Bel) и функции правдоподобия (plausible reasoning — Pl), которые используют, чтобы скомбинировать части разнородной информационной ситуации для вычисления вероятности событий.

Функция доверия представляет собой вероятность события: $Bel: P(X) \rightarrow [0,1]$.

Пусть X — универсум: набор всех состояний системы (рассматриваемых утверждений). Мощность множества 2^X содержит все подмножества множества X , включая пустое множество $\emptyset$. Например, если $X = \{a, b\}$, то $2^X = \{\emptyset, \{a\}, \{b\}, X\}$.

Элементы множества мощности могут быть взяты для представления предложений относительно фактического состояния системы, содержащего все и только те

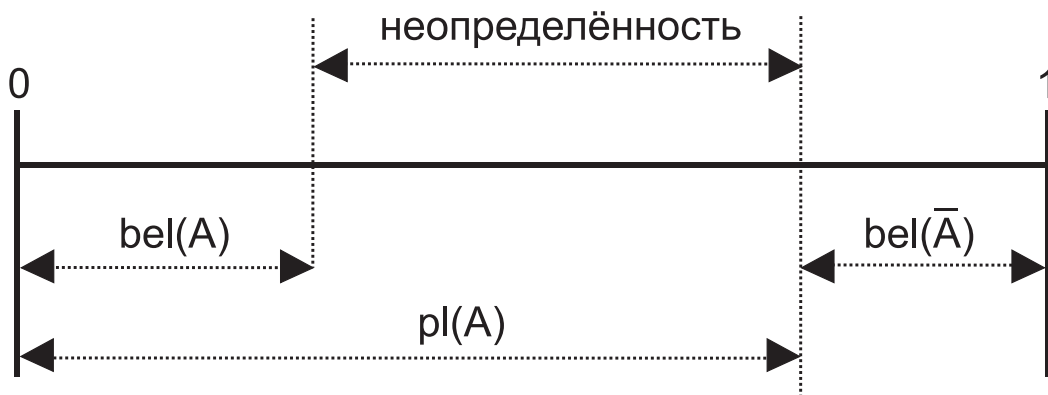


Рис. 4. Отношения между функцией доверия и функцией правдоподобия.

состояния, в которых утверждение верно. Теория доказательства присваивает массу убеждения каждому элементу набора мощности. Вводится понятие массы (m), заимствованное частично из физики, частично из математики, из теории вероятности. Формально функция $m: 2^x \rightarrow [0, 1]$ называется базовым убеждением (basic belief assignment – BBA), когда оно имеет два свойства:

1. Масса пустого множества равна нулю: $m(\emptyset) = 0$.

2. Массы оставшихся элементов показательного множества нормированы на единичную сумму:

$$\sum_{A \in P(x)} m(A) = 1.$$

Масса $m(A)$ элемента A , заданного члена набора, выражает долю всех релевантных и доступных доказательств, которые подтверждают утверждение о том, что фактическое состояние принадлежит A , но не подмножеству A . Значение $m(A)$ относится только к множеству A и не создаёт никаких дополнительных утверждений о других подмножествах A , каждое из которых, по определению, имеет свою собственную массу. Исходя из приписанных масс, можно определить верхнюю и нижнюю границы возможностей вероятностного интервала.

Этот интервал содержит точную величину вероятности рассматриваемого подмножества $P(A)$ (в классическом смысле) и ограничен двумя неаддитивными непрерывными мерами, называемыми доверием $\{belief\}$ (или поддержка $\{support\}$) и правдоподобие $\{plausibility\}$: $Bel(A) \leq P(A) \leq Pl(A)$.

Доверие $Bel(A)$ к множеству A определяется как сумма всех масс собственных подмножеств рассматриваемого множества:

$$Bel(A) = \sum_{B \downarrow B \subseteq A} m(B).$$

Правдоподобие $Pl(A)$ – сумма масс всех множеств B , пересекающихся с рассматриваемым множеством A :

$$Pl(A) = \sum_{B|A \cap B \neq \emptyset} m(B).$$

$$Pl(A) = 1 - Bel(\bar{A}) \text{ – для всех } A \in P(X).$$

Отношения между функцией доверия и функцией правдоподобия иллюстрирует рис. 4.

ТДШ позволяет интерпретировать доверие и правдоподобие как границы интервала возможного значения истинности гипотезы: доверие $\leq$ какая-то мера истинности $\leq$ правдоподобие.

При этом постулируется, что:

- доверие к гипотезе определяется как сумма масс свидетельств, однозначно поддерживающих гипотезу;
- правдоподобие определяется как единица – сумма масс всех свидетельств, противоречащих или опровергающих гипотезу.

По существу, ТДШ использует нечёткий оппозиционный подход. Различие между утверждением и оппозиционным утверждением определяет область неопределённости.

Например, пусть у нас есть гипотеза «поезд прибывает по расписанию». Если для неё доверие 0,5 и правдоподобие 0,85,





то это значит, что у нас есть свидетельства (общей массой 0,5), однозначно указывающие, что поезд прибывает по расписанию; но имеются и свидетельства (общей массой 0,15), однозначно указывающие, что поезд не прибывает по расписанию (доверие «поезд не прибывает» = $1 - 0,85 = 0,15$). Оставшаяся масса (дополняющая 0,5 и 0,15 до $1,0 = 0,35$), она же — зазор между правдоподобием 0,85 и доверием 0,5, соответствует «неопределённости» или наличию свидетельства, что поезд точно есть, но не говорящих ничего о том, прибывает он вовремя или не прибывает. Интервал $[0,5; 0,85]$ характеризует неопределённость истинности гипотезы, исходя из имеющихся свидетельств.

Таким образом, модель информационной ситуации с включением вероятностных характеристик позволяет оценивать неопределённость не только качественно, но и количественно. Это дает возможность использовать такие оценки в автоматизированных и интеллектуальных системах для повышения качества и надежности управления.

ЗАКЛЮЧЕНИЕ

Применение объектных и ситуационных моделей для управления транспортом является одним из направлений повышения эффективности деятельности транспорта и основой развития управления. Современное управление на транспорте всё больше переходит от эвристического к автоматизированному и интеллектуальному. Трудность такого перехода связана с трудностью формализации ситуаций, содержащих неопределённость.

Предложенный метод даёт формализовать неопределённость и проводить автоматизированную обработку ситуаций. Объектные и ситуационные модели позволяют на порядки уменьшить объём информации, анализируемый человеком, и использовать такие модели для принятия решений в ситуационных комнатах. В то же время теория Демпстера—Шафера остаётся только одним из подходов к оцен-

ке неопределённости и нуждается в дальнейшем развитии.

ЛИТЕРАТУРА

1. Чехарин Е. Е. Большие данные: большие проблемы // Перспективы науки и образования. — 2016. — № 3. — С. 7–11.
2. Цветков В. Я. Информационная неопределённость и определённость в науках об информации // Информационные технологии. — 2015. — № 1. — С. 3–7.
3. Коваленко Н. И. Информационный подход при построении картины мира // Перспективы науки и образования. — 2015. — № 6. — С. 7–11.
4. Лёвин Б. А., Мамаев Э. А., Багинова В. В. О концепции построения моделей производственно-транспортных систем // Наука и техника транспорта. — 2003. — № 4. — С. 8–17.
5. Мамаев Э. А. Особенности транспортных систем и специфика их моделирования // Системное моделирование социально-экономических процессов: Тезисы докл. и сообщ. XXV межд. науч. школы семинара им. акад. С. Шаталина. — Часть 2. — М.: ЦЭМИ РАН, 2002. — С. 92.
6. Дешко И. П. Информационное конструирование: Монография. — М.: МаксПресс, 2016. — 64 с.
7. Цветков В. Я. Информационные модели объектов, процессов и ситуаций // Дистанционное и виртуальное обучение. — 2014. — № 5. — С. 4–11.
8. Цветков В. Я. Когнитивное управление: Монография — М.: МаксПресс, 2017. — 72 с.
9. Неймарк Ю. И., Коган Н. Я., Савельев В. П. Динамические модели теории управления. — М.: Наука, 1985. — 400 с.
10. Барабанов И. Н. и др. Динамические модели информационного управления в социальных сетях // Автоматика и телемеханика. — 2010. — № 11. — С. 172–182.
11. Маркелов В. М. Применение топологических моделей геоанных для оптимизации транспортных маршрутов // Славянский форум. — 2012. — № 2. — С. 56–61.
12. Павлов А. И. Цифровое моделирование пространственных объектов // Славянский форум. — 2015. — № 4. — С. 275–282.
13. Кужелев П. Д. Пространственные знания для управления транспортом // Государственный советник. — 2016. — № 2. — С. 17–22.
14. Бутко Е. Я. Геоинформатика как метод построения картины мира // Славянский форум. — 2017. — № 1. — С. 34–41.
15. Цветков В. Я. Ресурсность и интегративность сложной организационно-технической системы // Международный журнал прикладных и фундаментальных исследований. — 2016. — № 5 (часть 4). — С. 676.
16. Нариньяни А. С. НЕ-факторы: краткое введение // Новости искусственного интеллекта. Вып. 2. — 2004. — М.: Комкнига, 2006. — С. 52–63.
17. Тихонов А. Н., Цветков В. Я. Методы и системы поддержки принятия решений. — М.: МаксПресс, 2001. — 312 с.
18. Shafer Glenn. Dempster—Shafer theory, 1992. [Электронный ресурс]: <http://www.glennshafer.com/assets/downloads/articles/article48.pdf>. Доступ 30.05.2017.

Координаты авторов: **Лёвин Б. А.** — tu@miit.ru,
Цветков В. Я. — cvj2@mail.ru.

Статья поступила в редакцию 30.05.2017, принята к публикации 23.08.2017.