

МЕТОДИКА ПРОГНОЗИРОВАНИЯ НАУЧНОЙ ПРЕДМЕТНОЙ ОБЛАСТИ НА ОСНОВЕ ГРАФОСЕМАНТИЧЕСКОГО МОДЕЛИРОВАНИЯ

В.Н. Афанасьев,
Д.А. Баранов,
И.В. Влацкая

В данной работе рассматриваются варианты решения (посредством использования математико-статистического инструментария) задачи моделирования научных предметных областей и прогнозирования их состояния. Представлены методики графосемантического моделирования научных предметных областей и прогнозирования их состояния на основе имитационного моделирования. В качестве примера приводится модель предметной области научных проектов, поддержанных Российским гуманитарным научным фондом (РГНФ) в 2010–2013 гг. Приведены результаты прогнозирования состояния предметной области на 2014 г.

Ключевые слова: научная предметная область, графосемантическая модель, прогнозирование, имитационное моделирование, ретроспективный анализ.

JEL: C53, C38.

Введение

В настоящее время сформировалась практика оценки результатов научной деятельности на основе наукометрических показателей и применения данной оценки в качестве повода для поощрения сотрудников научных организаций. К таким показателям относятся различные индексы цитируемости, импакт-фактор и т. д. Такая практика все чаще подвергается критике: утверждается, что показатели цитируемости исследователей, импакт-факторов журналов часто не соответствуют истинному положению дел ни в отношении самой статистики, ни в отношении качества исследований, порождают негативные процессы в науке [1, 2].

В то же время важным фактором успешной научной деятельности является финансовая поддержка со стороны различных государственных и негосударственных фондов. Кроме того, поддержка в виде грантов отдельных научных работ (проектов) характеризует актуальность их предметных областей, соответствие современным требованиям науки. Следовательно, актуальна задача моделирования и прогнозирования состояния предметных областей, поддерживаемых научными фондами. В публикуемой статье приводится решение описываемой задачи на примере

моделирования предметной области научных работ, поддержанных Российским гуманитарным научным фондом (РГНФ).

Графосемантическая модель предметной области

В данном исследовании для моделирования научной предметной области использовался метод графосемантического моделирования [3, 4, 5], ранее апробированный на различных задачах моделирования предметных областей, например предметной области журнала «Вопросы экономики» за 2010–2012 гг. [6]. Моделирование выполнялось в системе ИС «Семограф» [7, 8].

Основными понятиями метода графосемантического моделирования являются:

контекст – основная структурная сущность модели предметной области, определяет связи между смысловыми компонентами предметной области;

метатип – тип данных, определяющий конкретную описательную характеристику контекстов (источник, территориальная принадлежность, возраст и т. д.);

метаполе – значение определенного метатипа для данного контекста;

Афанасьев Владимир Николаевич (vAfanassyev@gmail.com) – д-р экон. наук, профессор, зав. кафедрой статистики и эконометрики Оренбургского государственного университета.

Баранов Дмитрий Александрович (baranov@semograph.com) – аспирант кафедры математического обеспечения информационных систем Оренбургского государственного университета.

Влацкая Ирина Валерьевна (mois@mail.osu.ru) – канд. техн. наук, доцент, зав. кафедрой математического обеспечения информационных систем Оренбургского государственного университета,

семантический компонент - сущность, характеризующая существование определенной семантики (значения, объекта и т. п.) в контексте;

семантическое поле - сущность, включающая множество компонентов, через которые определяется ее присутствие в контекстах (в обобщенном методе может включать другие поля, определяя иерархию);

семантическая карта - двумерная матрица, содержащая численные характеристики сил взаимных связей семантических полей (в обобщенной модели может быть многомерной и определяться не только для полей);

семантический граф - способ представления семантической карты, позволяющий наглядно оценить связность семантических полей.

Алгоритм метода графосемантического моделирования состоит из следующих шагов:

1. Сбор исходных данных, описывающих предметную область (контекстов);
2. Ввод описательной информации (заполнение метаполей) в собранных контекстах;
3. Выделение семантических компонентов в контекстах;
4. Объединение семантических компонентов в семантические поля;
5. Формирование выборок контекстов на основе метаполей;
6. Вычисление семантических карт;
7. Визуализация (построение семантического графа);
8. Анализ полученных данных.

Ниже приведено формальное математическое описание графосемантической модели:

S - множество символьных строк неограниченной длины;

$\Sigma \subset S$ - множество контекстов;

$C \subset S$ - множество компонентов;

$F \subset S$ - множество семантических полей;

T - множество типов метаполей;

$\Psi(\Sigma, T) = \{\psi: \psi(\sigma, t), \forall t \in T, \forall \sigma \in \Sigma\}$ - функция, определяющая множество метаполей для данного множества контекстов;

$\Phi(C, \Sigma) = \{\phi: \phi(c, \sigma) \in \{0, 1\}, \forall c \in C, \forall \sigma \in \Sigma\}$ - функция, определяющая множество связей компонентов с контекстами;

$\Lambda(C, F) = \{\lambda: \lambda(c, f) \in \{0, 1\}, \forall c \in C, \forall f \in F\}$ - функция, определяющая множество связей компонентов с полями;

$\Gamma(F, \Sigma) = \{\gamma: \gamma(f_1, f_2, \Sigma) \in \mathbb{N}, \forall f_1, f_2 \in F, f_1 \neq f_2\}$ - функция, определяющая множество числовых значений (сил, весов) связей между полями;

$\Omega[\Sigma, C, F, T, \Psi(\Sigma), \Phi(C, \Sigma), \Lambda(C, F), \Gamma(F, \Sigma)]$ - графосемантическая модель.

Также можно обозначить множество Z , содержащее факты существования связей полей с контекстами. Такое множество вычисляется на основе множеств F и Σ и является вспомогательным для нахождения $\Gamma(F)$:

$$Z(F, \Sigma) = \{\xi: \xi(f, \sigma) \in \mathbb{N}, \forall f \in F, \forall \sigma \in \Sigma\}.$$

Определим основные функции, используемые в обозначениях:

$\psi(\sigma, t), t \in T, \sigma \in \Sigma$ - функция, определяющая значение метаполя типа t для контекста σ ;

$$\phi(c, \sigma) = \begin{cases} 0, & c \notin \sigma \\ 1, & c \in \sigma \end{cases}, c \in C, \sigma \in \Sigma$$
 - функция, опре-

деляющая факт существования связи между компонентом c и контекстом σ , где $c \in \sigma, \forall c \in C, \forall \sigma \in \Sigma$ означает присутствие компонента c в контексте σ ;

$$\lambda(c, f) = \begin{cases} 0, & c \notin f \\ 1, & c \in f \end{cases}, c \in C, f \in F$$
 - функция, опре-

деляющая факт существования связи между компонентом c и полем f , где $c \in f, \forall c \in C, \forall f \in F$ означает присутствие (связь) компонента c в поле f ;

$\xi(f, \sigma) = \sum_{c \in C} \phi(c, \sigma) \cdot \lambda(c, f), f \in F, \sigma \in \Sigma$ - функция, определяющая силу связи между полем f и контекстом σ через компоненты $c \in C$.

Функция γ может быть определена двумя способами:

$$\gamma(f_1, f_2, \Sigma) = |\{\sigma: \sigma \in \Sigma, \xi(f_1, \sigma) > 0, \xi(f_2, \sigma) > 0\}|;$$

$$\gamma(f_1, f_2, \Sigma) = \sum_{\sigma \in \Sigma} \min\{\xi(f_1, \sigma), 1\} \cdot \min\{\xi(f_2, \sigma), 1\}. \quad (1)$$

В ходе исследования нами были собраны исходные данные о 889 поддержанных проектах РГНФ в 2009-2013 гг. в области знаний «02 - Экономические науки». РГНФ размещает основную информацию о поддержанных проектах в открытом доступе на своем сайте [9]. Собранные данные включают следующие типы описательной информации (метаданных):

1. Название проекта;
2. Область знаний;
3. Тип проекта;
4. Год начала проекта;

5. Ссылка на описание проекта на сайте РГНФ;

6. Ключевые слова, описывающие публикацию с точки зрения ее авторов.

Среди описанных типов метаданных особый интерес представляет «Тип проекта». Поля данного типа содержат одно из возможных 13 значений:

1) а - проект проведения научных исследований, выполняемый научным коллективом или отдельным ученым;

2) а1 - проект проведения научных исследований, выполняемый коллективом (до 9 человек) молодых ученых под руководством ведущего ученого без ограничения возраста;

3) а2 - проект проведения научных исследований, выполняемый коллективом (до 10 человек), состоящим полностью из молодых ученых, включая руководителя;

4) а(м) - совместный проект проведения научных исследований, выполняемый научным коллективом (международный конкурс);

5) а(р) - проект проведения научных исследований, выполняемый научным коллективом или отдельным ученым (региональный конкурс);

6) а(ф) - проект проведения научных исследований, выполняемый отдельным ученым;

7) а(ц) - проект проведения междисциплинарных исследований с изданием научных трудов по результатам исследований;

8) в - проект создания и приобретения программного обеспечения для информационных систем в области гуманитарных наук, способствующих распространению гуманитарных знаний в обществе;

9) г - проект организации в рамках реализации научных проектов мероприятий, в том числе конференций и семинаров;

10) г(р) - проект организации в рамках реализации научных проектов мероприятий, в том числе конференций и семинаров (региональный конкурс);

11) д - проект издания научных трудов по результатам научных исследований, проводимых в рамках научных проектов, профинансированных РГНФ;

12) е - проект экспедиций, полевых и социологических исследований, научно-реставрационных работ, необходимых для получения новых данных в области гуманитарных наук;

13) к - проект подготовки научно-популярного издания.

«Тип проекта» имеет важное значение, поскольку естественным образом определяет классы моделей предметных областей научных проектов, разделяя их по целям, составу исполнителей, территориальным и другим признакам. На рис. 1 приведено распределение количества проектов по приведенным выше значениям.

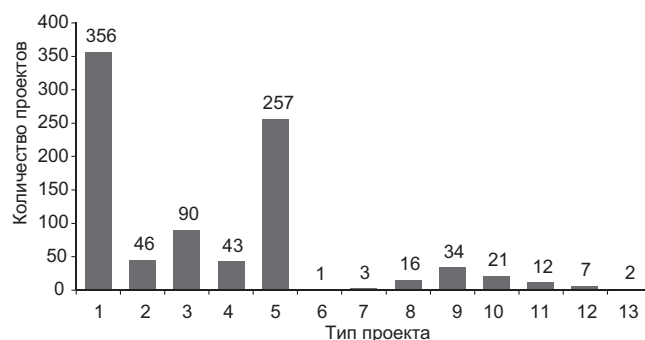


Рис. 1. Распределение количества проектов по типам проектов

Как видно из рис. 1, наибольшее число поддержанных проектов приходится на 1-й и 5-й типы. В данном исследовании рассматривается только множество проектов первого типа (а - проект проведения научных исследований, выполняемый научным коллективом или отдельным ученым) как наиболее унифицированное и широко представленное. При построении графосемантической модели предметной области на основе данного множества проектов было сформировано множество контекстов. Полученное множество включает 356 контекстов.

В процессе подготовки исходных данных было выделено 4468 уникальных ключевых слов. Из наиболее частотных ключевых слов было сформировано множество семантических компонентов S (3830 элементов). Полученные компоненты были объединены экспертом в семантические поля. Для этого были определены множества семантических полей F (17 полей) и множество связей поле-компонент $\Lambda(S, F)$ (2833 связи). Ниже приведено полученное множество семантических полей:

1. Методология исследований.
2. Научные направления.
3. Виды деятельности.
4. Территориальные образования и географическое расположение.
5. Институциональные структуры.
6. Денежно-кредитная политика.
7. Безопасность.

8. Фундаментальная наука.
9. Система образования.
10. Политическая деятельность.
11. Макроэкономика.
12. Микроэкономика.
13. Демография и миграционные процессы.
14. Качество жизни населения.
15. Ресурсный потенциал.
16. Особенности рыночного хозяйствования.
17. Стратегические направления развития экономики.

В качестве эксперта выступил один из авторов данной статьи, заведующий кафедрой статистики и эконометрики Оренбургского государственного университета, д-р экон. наук, проф. В.Н. Афанасьев.

На рис. 2 приведено распределение количества контекстов по количеству связанных полей (мощности предметной области).

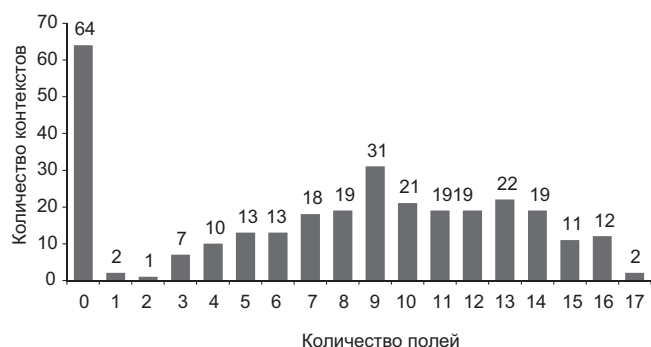


Рис. 2. Распределение количества контекстов по количеству связанных полей

Из рис. 2 видно, что значительное количество контекстов не связано с семантическими полями. Появление таких контекстов возможно в результате отсутствия ключевых слов в описании исходного проекта, либо использования низкочастотных ключевых слов, не связанных с семантическими полями в процессе построения модели ($|C| < |\Lambda(C, F)|$). Подобные контексты, не содержащие предметной области в виде связей с семантическими полями, не представляют интерес в данном исследовании. Фактически, при анализе графосемантической модели предметной области рассматриваются только связи между семантическими полями, возникающие при связи контекста с более чем одним семантическим полем, поэтому контексты, связанные с одним полем, также не представляют интереса.

Как следствие, из рассматриваемой выборки были исключены контексты, связанные с менее чем двумя семантическими полями, благодаря чему размер выборки сократился до 237 контекстов.

В результате нами была получена графосемантическая модель Ω_{rfh} и построен семантический граф, представленный на рис. 3.



Рис. 3. Семантический граф исходной модели предметной области

В семантическом графе вершины соответствуют семантическим полям, а ребра - связям между ними. Размер вершины пропорционален частоте поля, а толщина ребра - силе связи между полями.

Частнонаучные предметные области

Для подробного изучения структуры модели предметной области в ней выделяют частнонаучные предметные области. *Частнонаучная предметная область* - это представленная в виде математической модели композиция терминоплей, репрезентирующая отдельный научный сегмент в общей предметной области [10]. В данном исследовании для выделения частнонаучных предметных областей использовался кластерный анализ. Кластерный анализ производится над множеством предметных областей, представленных в отдельных контекстах (единичных предметных областей), описываемых бинарными векторами. В качестве алгоритма кластерного анализа в данном исследовании используется алгоритм нечеткой кластеризации C-means [11]. Ниже приведено подробное описание данного алгоритма.

Алгоритм нечеткой кластеризации C-means основывается на минимизации целевой функции $J(x, c, u, m)$:

$$J(x, c, u, m) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, m \in R, m > 1,$$

где m - нечеткий параметр; u - степень принадлежности кластеру; $\|\cdot\|$ - норма, характеризующая близость элементов анализируемого пространства.

Процесс оптимизации целевой функции заключается в итеративном пересчете степеней $u_{ij}, i = \overline{1, N}$ и центров кластеров $c_j, j = \overline{1, C}$:

$$u_{ij} = \frac{1}{\sum_{k=1}^K \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}};$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}.$$

Условие окончания итерации: $\max_{ij} (u_{ij}^{(k+1)} - u_{ij}^{(k)}) < \varepsilon$, где ε - заданный критерий, $\varepsilon > 0$.

Число кластеров было выбрано равным 5, параметр $m = 2$. В качестве исходных центров кластеров были выбраны наиболее высокочастотные наборы полей. Набор полей F_σ - множество семантических полей, связанных через единственный контекст σ , $F_\sigma \subseteq F$. Набор полей также называют единичной предметной областью.

Полученные нами кластеры определяют частнонаучные предметные области; их описание представлено в таблице. Для каждого кластера приведен наиболее *репрезентативный проект* - проект, единичная предметная область которого наиболее близка к центру кластера (для определения расстояния применяется евклидова норма). Следует отметить, что в описании каждого кластера представлен один репрезентативный проект, однако их может быть несколько (находящихся на одном расстоянии от центра кластера), но их единичные предметные области совпадают.

Таблица

Описание частнонаучных предметных областей

№	Наборы полей	Наиболее репрезентативные проекты
1	7, 9, 13, 15	Разработка принципов функционирования региональной инновационной системы
2	3, 5, 6, 7, 9, 10, 12, 15, 16	Разработка методологии моделирования процессов преодоления социodemографического кризиса в России

Окончание таблицы

№	Наборы полей	Наиболее репрезентативные проекты
3	7, 8, 9, 10, 13, 14, 15, 17	Методология формирования экономики знаний
4	1, 4, 5, 7, 8, 9, 10, 11, 12, 13, 15, 17	Системный анализ стратегий устойчивого развития на примере Бурятской части Байкальского региона
5	1-17	Технологии мониторинга и прогнозирования приоритетных направлений развития региональной инновационной и кластерной политики в лесном хозяйстве

Частнонаучные предметные области фактически описывают наиболее актуальные направления современной науки в рассматриваемой области знаний. При этом выбор количества кластеров позволяет регулировать детализацию модели. В рассмотренном случае представлен достаточно низкий уровень детализации, о чем свидетельствует высокая мощность наборов полей, составляющих центры полученных кластеров. Следует отметить кластер № 5, в центре которого приведены все доступные семантические поля. Данный кластер является устойчивым (сохраняется при изменении параметра m и количества кластеров) и включает контексты, не содержащие выраженных признаков прочих кластеров.

Значительный интерес представляет изменение предметных областей во времени. Для изучения этого процесса частнонаучные предметные области были визуализированы в виде графика, приведенного на рис. 4. На этом графике отображены кумулятивные суммы количества статей в каждом из рассматриваемых направлений (частнонаучных предметных областей) за рассматриваемый период (2010-2013 гг.).

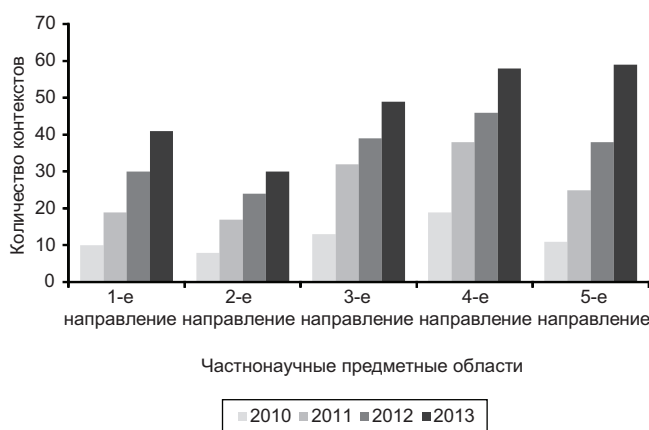


Рис. 4. Изменение мощности частнонаучных предметных областей за 2010-2013 гг.

Ретроспективный анализ

Следующим этапом исследования было прогнозирование состояния изучаемой предметной области на 2014 г. Деятельность агентов научного производства является длительным, трудоемким и невоспроизводимым процессом, что делает практически невозможным проведение натурных экспериментов с реальными агентами научного производства. Поэтому для исследования процесса изменения состояния научной предметной области нами была разработана методика на основе имитационного моделирования [12]. В качестве имитационной модели используется дискретный марковский процесс первого порядка. Состояниями этого процесса являются графосемантические модели предметной области на заданном шаге. Переход между состояниями осуществляется посредством поддержки новых проектов (добавления новых контекстов в модель). Ниже приведено краткое описание процесса имитации. Подробное описание данной модели приведено в [13].

Пусть задана графосемантическая модель научной предметной области агента научного производства Ω и модель изменения ее состояния в виде марковского процесса первого порядка с матрицей перехода $p_{N \times N}$, где N - количество всех допустимых состояний предметной области. Тогда имитационное моделирование первых T переходов может быть осуществлено следующим образом:

1. В качестве начального состояния X_0 выбирается заданная модель Ω .

2. Для каждого шага $t = 1, \overline{T}$:

(а) выбирается отрезок $x: [0; \sum_{i=1}^N p_{ii}]$;

(б) генерируется («разыгрывается») случайное число r , равномерно распределенное на отрезке x ;

(с) в качестве состояния X_t выбирается такое состояние $\Omega^{(j)}$, для которого выполняется условие:

$$\sum_{i=1}^{j-1} p_{ii} < r \leq \sum_{i=1}^j p_{ii};$$

(д) переход к шагу $t+1$.

3. Полученная последовательность состояний $X_t, t = 0, T$ принимается в качестве результата имитационного моделирования.

Вероятности p_{ii} вычисляются на основе модели перехода:

$$P(F_\sigma, \overline{F \setminus F_\sigma}) = P(F_\sigma) \prod_{i=1}^{n-k} \frac{P(\overline{f_{k+i}}) \prod_{j=1}^k P(f_j | \overline{f_{k+i}}) \prod_{j=1}^{i-1} P(\overline{f_{k+j}} | \overline{f_{k+i}})}{P(f_1, f_2, \dots, f_k, f_{k+1}, \dots, f_{k+i-1})}, \quad (2)$$

где $P(F_\sigma, \overline{F \setminus F_\sigma})$ - вероятность присутствия в следующем контексте набора полей F_σ при отсутствии других семантических полей; $P(\overline{f_j})$ - вероятность присутствия поля $f_j \in F$ в новом контексте.

Вывод формулы (2) приведен в [13].

Принцип, используемый на шагах (б), (с), также называется «принципом рулетки» [14]. Полученная последовательность $X_t, t = 0, T$ является прогнозом состояния исследуемой предметной области.

Для оценки ошибки прогноза был применен ретроспективный анализ. На основе контекстов за 2009-2012 гг. была построена исходная модель Ω_{rth}^1 , используемая в качестве исходной для прогноза состояния научной предметной области на 2013 г. Для оценки количества прогнозируемых состояний было рассмотрено распределение количества контекстов по годам, приведенное на рис. 5.

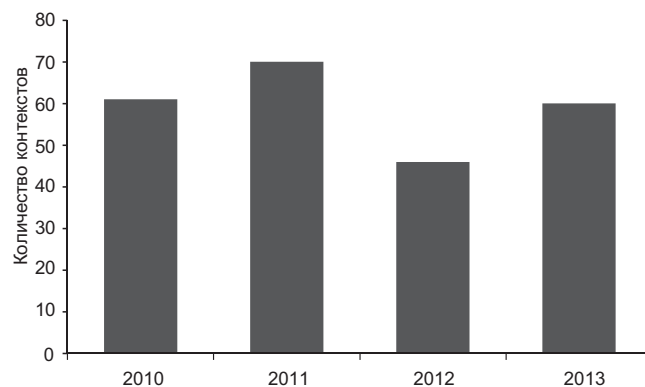


Рис. 5. Распределение количества контекстов по годам

Реальное количество контекстов в 2013 г. составляет 60, а среднее арифметическое за 2009-2013 гг. равно 59,25:

$$N = \frac{1}{4} \sum_{i=2009}^{2013} n_i = 59,25,$$

где n_i - количество контекстов в i -м году; N - количество прогнозируемых контекстов.

Количество прогнозируемых контекстов N было решено округлить до 60. Полученные в ре-

зультате прогнозирования единичные предметные области были распределены по ранее выделенным частнонаучным предметным областям; результат приведен на рис. 6.

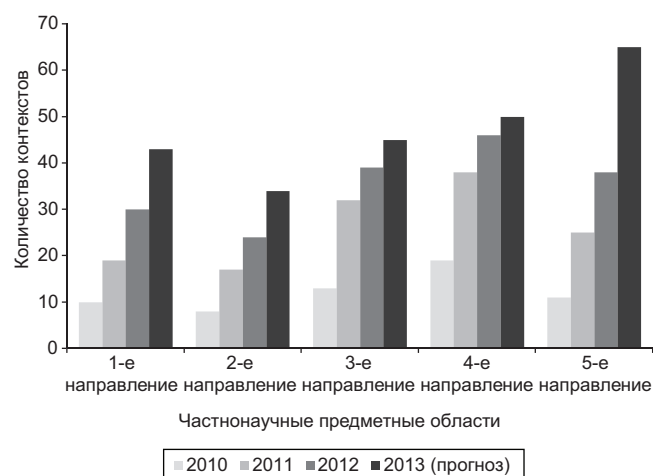


Рис. 6. Изменение мощности частнонаучных предметных областей за 2010-2012 гг. и прогноз на 2013 г.

В качестве оценки ошибки прогноза использовалась распространенная мера MAPE (Mean Absolute Percentage Error, средняя абсолютная ошибка в процентах) [15]:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|Z(t) - \hat{Z}(t)|}{Z(t)} 100\%,$$

где $Z(t)$ - реальное значение в момент времени t ; $\hat{Z}(t)$ - прогнозируемое значение в момент времени t .

Поскольку в рассматриваемой задаче состояние описывается семантической картой, то есть матричной величиной, в качестве $Z(t)$ используется матричная норма. Обычно матричная норма выбирается как подчиненная ранее используемой векторной норме [16, 17]. Поскольку ранее в данном исследовании применялась евклидова норма вектора $\|x\|_2$, при решении данной задачи была выбрана подчиненная ее спектральная матричная норма $\|A\|_2$:

$$\|x\|_2 = \sqrt{\sum_i x_i^2};$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^{\dagger}A)},$$

где $A^{\dagger}$ - матрица, сопряженная с A ; $\lambda_{\max}(A^{\dagger}A)$ - наибольшее собственное число матрицы $A^{\dagger}A$.

При прогнозе 60 контекстов за 2013 г. был получен график MAPE, приведенный на рис. 7.

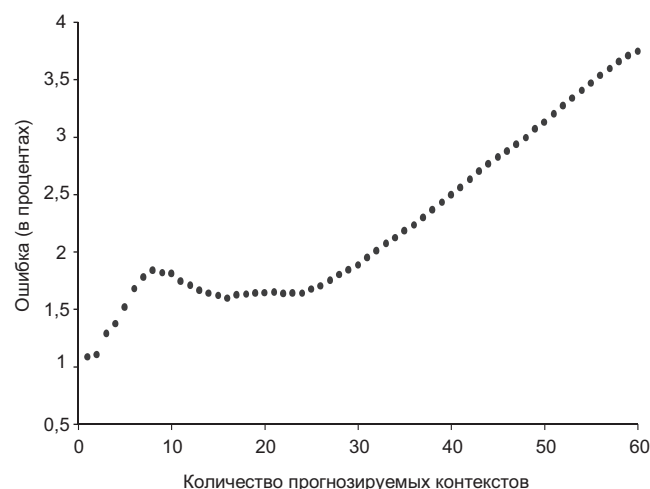


Рис. 7. Зависимость MAPE от числа прогнозируемых контекстов

Из рис. 7 следует, что средняя абсолютная ошибка в процентах вычисления семантической карты при прогнозировании состояния предметной области на год для данной задачи не превышает 4%; следовательно, полученную модель можно считать адекватной.

Прогноз

На основе описанной методики нами был построен прогноз состояния изучаемой предметной области на 2014 г., на основе которого была построена графосемантическая модель, семантический граф которой приведен на рис. 8.



Рис. 8. Семантический граф исходной модели предметной области

Полученные в результате прогнозирования единичные предметные области были распре-

делены по ранее выделенным частнонаучным предметным областям (результат приведен на рис. 9).

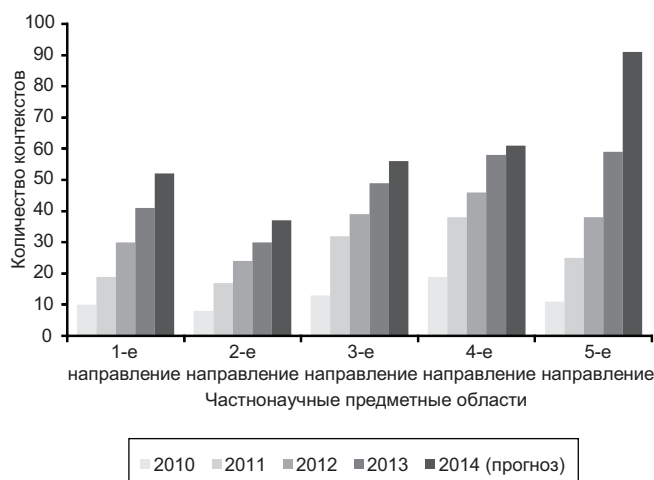


Рис. 9. Изменение мощности частнонаучных предметных областей за 2010-2013 гг. и прогноз на 2014 г.

Результаты

1. Построена графосемантическая модель предметной области научных проектов, поддержанных РГНФ в 2010-2013 гг.

2. На основе полученной модели выделены пять основных частнонаучных предметных областей, описывающих наиболее актуальные направления современной науки в области знаний «02 - Экономические науки».

3. Апробирована методика прогнозирования состояния предметной области научных проектов, поддержанных РГНФ в 2013 г. С помощью ретроспективного анализа оценена погрешность прогноза. Показано, что при прогнозировании последовательности из 60 состояний модели оценка ошибки $MARE$ не превышает 4%; следовательно, модель можно считать адекватной.

4. Произведен прогноз состояния предметной области научных проектов, поддержанных РГНФ в 2014 г. Приведена оценка ее характеристик и построен семантический граф.

Литература

1. Campbell P. Escape from the impact factor // Ethics in Science and Environmental Politics. 2008. Vol. 8. P. 5-7.
2. Lawrence P.A. Lost in publication: how measurement harms science // Ethics in Science and Environmental Politics. 2008. Vol. 8. P. 9-11.

3. Белоусов К.И. Теория и методология полиструктурного синтеза текста. - М.: Флинта, 2009. - 216 с. [Belousov K.I. Theory and methodology of polystructural text synthesis. - M.: Flint, 2009. - 216 p.].

4. Белоусов К.И., Зелянская Н.Л., Баранов Д.А. Концептуально гипертекстовая модель управления контентом в ИС «Семограф» // Вестник ОГУ. 2012. Т. 11, № 147. С. 56-61. [Belousov K.I., Zelyanskaya N.L., Baranov D.A. Conceptual model of hypertext content management in «Semograf» IS // Vestnik OSU. 2012. Vol. 11, No 147. P. 56-61].

5. Математическая формализация метода графосемантического моделирования: Техника и технология: новые перспективы развития / Материалы VIII Международной научно-практической конференции; исполн.: Д.А. Баранов. - М.: 2013. - С. 70-78. [Mathematical formalization of graphosemantic modeling method: Techniques and technology: new development prospects / Proceedings of the VIII International Scientific Conference; executed.: D.A. Baranov. - M.: 2013. P. 70-78].

6. Применение графосемантического моделирования для анализа предметных областей агентов научного производства (на примере журнала «Вопросы экономики») / В.Н. Афанасьев, Д.А. Баранов, И.В. Влацкая, Д.А. Ичкинеева // Вестник ОГУ. 2013. Т. 9. № 158. С. 161-169 [Application of graphosemantic modeling for the analysis of subject domains of scientific production agents (on the example of the journal «Voprosy ekonomiki») / V.N. Afanasyev, D.A. Baranov, I.V. Vlatkaya, D.A. Ichkineeva // Vestnik OSU. 2013. Vol. 9. No 158. P. 161-169].

7. Семограф. URL: <http://new.semograf.com> (дата обращения: 4.08.2013). [Semograf. URL: <http://new.semograf.com> (Accessed: 04.08.2013)].

8. Система графосемантического моделирования / Д.А. Баранов, К.И. Белоусов, И.В. Влацкая, Н.Л. Зелянская. - М.: Свидетельство о государственной регистрации в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам. Зарегистрировано в Реестре программ для ЭВМ № 20111617192 от 15.09.2011. [Graphosemantic modeling system / D.A. Baranov, K.I. Belousov, I.V. Vlatkaya, N.L. Zelyanskaya. - M.: State Registration Certificate of the Federal Service for Intellectual Property, Patents and Trademarks. Registered in the Register of Computer Programs No 20111617192 from 15.09.2011].

9. Проекты, поддержанные РГНФ. URL: <http://www.rfh.ru/index.php/en/grant/isintegr> (дата обращения 08.04.2014) [Projects supported by the Russian Foundation for Humanities. URL: <http://www.rfh.ru/index.php/en/grant/isintegr> (Accessed: 08.04.2014)].

10. Белоусов К.И., Баранов Д.А., Зелянская Н.Л. Научный коллектив и его предметные области (К вопросу о методах эффективного планирования научной деятельности) // Научно-техническая информация.

2014. № 4. С. 13-26. [Belousov K.I., Baranov D.A., Zelyanskaya N.L. Research team and its subject areas (On the question of methods for effective planning science activities) // Scientific and Technical Information. 2014. No 4. P. 13-26].

11. **Каумак У., Сетнес М.** Extended fuzzy clustering algorithms // Erasmus Research Institute of Management. - 2000. - 24 p.

12. **Лукьянов В.С., Слесарев Г.В.** Проектирование компьютерных сетей методами имитационного моделирования. - Волгоград, 2001. - 72 с. [Lukyanov V.S., Slesarev G.V. Design of computer networks by simulation methods. - Volgograd, 2001. - 72 p.].

13. Вероятностный подход к графосемантическому моделированию: «Формирование основных направлений развития современной статистики и эконометрики» / Материалы I Международной научной конференции; исполн.: Дмитрий Александрович Баранов. - Оренбург, 2013. С. 166-175 [A probabilistic approach to

graphosemantic modeling: «Creating main development directions for modern statistics and econometrics»/ Materials from the Ist International Scientific Conference; executed.: D.A. Baranov. - Orenburg, 2013. P. 166-175].

14. **Осовский С.** Нейронные сети для обработки информации. - М.: Финансы и статистика, 2002. - 344 с. [Osovski C. Neural networks for information processing. - M.: Finance and Statistics, 2002. - 344 p.].

15. **Swanson D.A., Tayman J., Bryan T.M.** MAPE-R: a rescaled measure of accuracy for cross-sectional forecasts // Journal of Population Research. 2011. Vol. 28. P. 225-243.

16. **Самарский А.А., Гулин А.В.** Численные методы. - М.: Наука, 1989. - 432 с. [Samarskii A.A., Gulina A.V. Numerical methods. - M.: Nauka, 1989. - 432 p.].

17. **Волков Е.А.** Численные методы. - М.: Наука, 1987. - 248 с. [Volkov E.A. Numerical methods. - M.: Nauka, 1987. - 248 p.].

GRAPHOSEMANTIC MODELING TECHNIQUE FOR THE SCIENTIFIC DOMAINS

Vladimir Afanas'ev

Author affiliation: Orenburg State University (Orenburg, Russia). E-mail: vAfanassyev@gmail.com.

Dmitry Baranov

Author affiliation: Orenburg State University (Orenburg, Russia). E-mail: baranov@semograph.com.

Irina Vlatskaya

Author affiliation: Orenburg State University (Orenburg, Russia). E-mail: mois@mail.osu.ru.

In this paper tasks of simulating scientific domains and forecasting their state are considered. Graphosemantic modeling techniques for scientific domains and simulation-based forecasting of their state are introduced. The domain model of research projects supported by the Russian Humanitarian Foundation (RFH) in 2010-2013 serves as an example. Forecasting results for the state of subject domain in 2014 are presented.

Keywords: scientific domain, graphosemantic model, forecasting, simulation, retrospective analysis.

JEL: C53, C38.