

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ НАУЧНОГО ПРОЦЕССА АКАДЕМИЧЕСКОГО ИНСТИТУТА

В. В. Моисеенко

Институт вычислительной математики
и математической геофизики СО РАН,
630090, Новосибирск, Россия

УДК 004.94

С использованием баз данных, включающих кадровые и производственные показатели, построена имитационная модель подсистемы „Публикация научных работ“ научного процесса академического института. Задачи, решаемые с помощью модели, позволяют анализировать структурные характеристики и трудовую деятельность научных кадров института.

Ключевые слова: имитационная модель, научная работа, научная активность, база данных, научный процесс, пространство событий.

With the use of the data including the indices of personnel and productive activities we have developed an imitating model of the sub-system „Publishing Scientific Papers“ of the Scientific Process of an academic institution. The problems to be solved with the help of the model proposed make possible to analyze structural characteristics and labor activity of the scientific staff of an institution.

Key words: imitating models, scientific research, scientific activity, database, scientific process, a set of events.

1. Определение научного процесса академического института.

1.1. Научный процесс академического института является сложной вероятностной системой и включает в себя ряд подсистем, название которых соответствует выполняемым ими функциям. Это:

- научные кадры;
- публикация научных работ;
- выполнение научных проектов;
- участие в семинарах и конференциях;
- создание программных продуктов;
- разработка изобретений;
- защита диссертаций;
- преподавательская деятельность;
- руководство дипломными проектами.

Эти подсистемы взаимосвязаны. Например, публикация научной работы может являться результатом участия в конференции или результатом выполнения научного проекта. Создание программного продукта, как правило, необходимо для расчетов, включаемых в научную работу. Руководство дипломными проектами является частью преподавательской деятельности. Участие в семинарах и конференциях обязательно в процессе

написания диссертаций. И, наконец, научный процесс института неразрывно связан с деятельностью научных кадров, которые обеспечивают функционирование всех остальных подсистем.

На первом этапе работы предлагается построить модель функционирования подсистемы „Публикация научных работ“. Это, в свою очередь, позволит исследовать работу подсистем, связанных с ней. В дальнейшем, говоря о научном процессе работника или всего института, будем считать, что на данном этапе исследование касается той части процесса, которая связана только с публикацией научных работ.

1.2. Во время трудовой деятельности работник ежегодно публикует определенный набор научных работ (статьи в центральных или зарубежных журналах, монографии и т. д.) индивидуально или с соавторами. В год, когда работник не имеет публикаций, набор является нулевым.

Обозначим набор научных работ, опубликованных в l -й год трудовой деятельности работника, через n_l , а в следующий год — через $n_{(l+1)}$. Первый набор назовем *исходным набором*, а второй — *конечным набором*. Тогда научный процесс работника, сформировавшийся от начала его трудовой деятельности до некоторого выбранного года l_w , включительно, можно представить в виде последовательности событий

$$\underbrace{n_{l_b} \rightarrow n_{l_b+1}}_{s_{l_b+1}}, \underbrace{n_{l_b+1} \rightarrow n_{l_b+2}, \dots, n_l \rightarrow n_{(l+1)}}_{s_{l_b+2}}, \dots, \underbrace{n_l \rightarrow n_{(l+1)}}_{s_{l+1}}, \dots, \underbrace{n_{l_w-1} \rightarrow n_{l_w}}_{s_{l_w}}, \quad (1)$$

где

l_b — год начала трудовой деятельности работника;

l_w — год, до которого следует сформировать научный процесс работника;

s_l — события, формирующие научный процесс работника. Здесь индекс l указывает год завершения события.

Научный процесс всего института, объединяющий все события персональных научных процессов работников, является, таким образом, пространством событий. Период, за который берутся публикации научных работ для формирования пространства событий, включает в себя годы с 1994 по 2013.

1.3. Как показал анализ динамики производственных показателей научных работников института [1], интенсивность их научных процессов неодинакова. Чем чаще публикуется научный работник, тем выше интенсивность его научного процесса, т. е. его научная активность. Необходимо также отметить, что научная активность работника, как правило, меняется в течение его трудовой деятельности.

При исследовании научных процессов работников удобнее оперировать количественными оценками научной активности. Определим текущую научную активность работника в заданный год как среднегодовое количество публикаций научных работ за период его трудовой деятельности до этого года включительно. Назовем это значение *текущим показателем научной активности* работника, в дальнейшем — *показателем I_A работника*. Под научной активностью здесь понимается не число баллов, являющееся суммарной оценкой научных работ по некоторой системе балльных оценок, а интенсивность их публикаций. Дело в том, что системы балльных оценок имеют субъективный характер. Действительно, существует своя система балльных оценок, принятая в СО РАН для отчетов институтов. В самих институтах функционируют свои системы балльных оценок, по которым оценивается деятельность работников. Все эти системы неравнозначны и непостоянны. Вывод

Таблица 1

Граничные значения показателя научной активности для групп

| Группы | 1-я | 2-я | 3-я | 4-я | 5-я |
|-----------------|-------------------|-------------------------|-------------------------|-------------------------|----------------|
| Значения границ | $G_{gr} \leq 0,5$ | $0,5 < G_{gr} \leq 1,0$ | $1,0 < G_{gr} \leq 1,7$ | $1,7 < G_{gr} \leq 2,5$ | $G_{gr} > 2,5$ |

же о научной ценности работы следует делать на основании мнений компетентных и объективных рецензентов.

У работников с разными уровнями научной активности персональные научные процессы имеют свои закономерности. В связи с этим предлагается распределить работников института по группам. Для этого необходимо задать граничные значения показателя I_A для каждой группы G_{gr} . При определении этих границ автор ориентировался на должностные инструкции работников института, в которых определено необходимое количество публикаций научных работ для каждой научной должности. Принятые граничные значения показателя I_A для каждой группы приведены в табл. 1.

В соответствии с группами научной активности пространство событий всего института разбивается на групповые подпространства.

2. Имитационная модель подсистемы „Публикация научных работ“.

2.1. Для исследования поведения подсистем научного процесса, оценки его качественных характеристик, прогнозирования его результатов правомерно использовать методологию имитационного моделирования. Создание имитационной модели любой сложной системы включает в себя ряд взаимосвязанных этапов. Одним из таких этапов является создание информационного и программного обеспечений модели.

Информационным обеспечением разрабатываемой имитационной модели являются базы данных автоматизированных информационных систем (АИС) „Кадров института“ и „Научный рейтинг“. Эти интегрированные базы включают в себя кадровые показатели работников и перечни публикаций научных работ за весь период их трудовой деятельности.

Разработанное автором программное обеспечение отражает структуру имитационной модели и включает в себя следующие задачи:

- расчет текущей научной активности работников на каждый год;
- распределение работников по группам научной активности в каждый год;
- формирование групповых подпространств событий;
- расчет фактических количеств публикаций научных работ в группах научной активности в каждый год;
- расчет оценки количеств публикаций научных работ в заданный год;
- расчет средних значений отклонений оценок количеств публикаций научных работ от реальных за выбранные годы;
- решение аналитических задач с помощью имитационной модели.

2.2. Проблема построения имитационной модели требует решения задачи идентификации. Понятие идентификации охватывает круг вопросов, связанных с отысканием таких параметров модели и их значений, которые бы уменьшали насколько возможно величину рассогласования между реальными показателями научного процесса и показателями, рассчитанными с помощью созданной модели. Поскольку моделируемая подсистема связана с публикацией научных работ, то логично в качестве таких показателей взять

Таблица 2

Динамика количества публикаций научных работ

| Годы | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Количество публикаций | 263 | 244 | 285 | 276 | 274 | 216 | 217 | 295 | 238 | 256 | 317 | 262 | 335 |

- количество публикуемых научных работ в группах научной активности за год;
- количество публикуемых научных работ по видам за год;
- суммарное количество публикуемых научных работ в институте за год.

Определение оценок этих показателей может базироваться на решении задач интерполяции и экстраполяции. Однако применение ряда известных методов (метод наименьших квадратов, разностные схемы, метод скользящих средних) в нашем случае не дало удовлетворительных результатов. Причина, вероятно, заключается в том, что при использовании формальных методов не учитывается случайный характер создания научного продукта.

В связи с этим предлагается следующая эвристическая процедура:

1) Задается год l_k , в который следует рассчитать количество публикуемых научных работ. Назовем его *контрольным годом*.

2) У очередного работника текущей группы научной активности определяется набор публикаций в год, предшествующий контрольному. Это — исходный набор событий.

3) В соответствующем групповом подпространстве событий находим подмножество событий, имеющих такой исходный набор. Конечные наборы событий являются теми наборами научных работ, среди которых один предположительно может быть изданным конкретным работником в контрольном году. Термин *предположительно может быть изданным* подчеркивает вероятностный характер имитационной модели научного процесса.

4) Прежде чем перейти к описанию алгоритма выбора конечных наборов, рассмотрим изменение количеств публикаций научных работ института по годам (табл. 2).

Как видно из таблицы, в разные годы количества публикаций научных работ значительно отличаются. Оказывается, что если мы обратимся к графику аттестаций научных работников за эти годы, то эти отличия вполне объяснимы. Действительно, аттестация большого числа работников в 2003, 2008, 2011 годах привела к увеличению количества публикаций научных работ в эти годы. И наоборот, охват аттестацией очень малого числа работников привел к спаду количества публикаций научных работ в 2002, 2006, 2007, 2009 годы. В другие же годы (2002, 2004, 2005, 2010, 2012) аттестацией было охвачено относительно не очень большое число работников. Т. е. налицо фактор, влияющий на количество публикаций научных работ. Назовем его *административным фактором*. Этот фактор может быть связан и с другими причинами. В частности, значительный рост количества публикаций научных работ в 2013 году с большой уверенностью можно связать с кардинальной реформой Российской академии наук. При моделировании влияния этого фактора вводится соответствующий параметр.

5) Из всего подмножества событий, имеющих одинаковый исходный набор, выбираем те события, конечные наборы которых удовлетворяют условию

$$n_{l_k} \leq \bar{n}_s \cdot k_{gl_k}, \quad (2)$$

n_{l_k} — количество научных работ в выбираемом наборе в l_k -й контрольный год;

\bar{n}_s — среднее количество научных работ в наборах, опубликованных s -м работником за годы трудовой деятельности, предшествующие контрольному году;

k_{gl_k} — поправочный коэффициент, учитывающий административный фактор для g -й группы научной активности в l_k -й год.

6) Анализ показал, что величина отношений количества публикаций одного вида научных работ к количеству публикаций другого вида работ для разных видов имеет свою динамику. Поэтому для каждого события j из подмножества, полученного после реализации п. 5, определяются величины

$$\delta_{ijz} = |\bar{\Theta}_{iz} - \Theta_{ijz} \cdot \rho_\theta|, \quad i = 2, 3, 4, 5, \quad (3)$$

где Θ_{ijz} — расчетное значение отношения количества публикаций статей в центральных журналах к количеству публикаций научных работ i -го вида, если будет выбрано j -е событие ($i = 2$ — статьи в зарубежных журналах; $i = 3$ — патенты; $i = 4$ — материалы международных конференций; $i = 5$ — монографии). Число публикаций научных работ в этом показателе равно сумме уже выбранных публикаций научных работ для всех предыдущих работников плюс число публикаций научных работ в j -м событии;

$\bar{\Theta}_{iz}$ — реальное среднее значение отношения количества публикаций статей в центральных журналах к количеству публикаций научных работ i -го вида за три года перед контрольным;

ρ_θ — коэффициент, сглаживающий погрешности расчетных значений отношений.

Для каждого события определяется сумма

$$\Delta_{jz} = \sum_{i=2}^5 \delta_{iz}. \quad (4)$$

Выбирается то событие, для которого эта величина минимальна.

В итоге определяется событие, у которого конечный набор n_{sl_k} является набором научных работ, возможно опубликованных s -м работником в l_k -м контрольном году. Опять же, формулировка *набор научных работ, возможно опубликованных* указывает на вероятностный характер выбора наборов.

7) Рассчитываются следующие величины

$$P_{gl_k} = \sum_{i=1}^5 \sum_{s \in S_{gl_k}} p_{isl_k}, \quad g = 1, 2, 3, 4, 5. \quad (5)$$

$$P_{il_k} = \sum_{g=1}^5 \sum_{s \in S_{gl_k}} p_{isl_k}, \quad i = 1, 2, 3, 4, 5. \quad (6)$$

$$P_{l_k} = \sum_{g=1}^5 \sum_{i=1}^5 \sum_{s \in S_{gl_k}} p_{isl_k}. \quad (7)$$

где P_{gl_k} — количество научных работ, возможно опубликованных работниками g -й группы научной активности в l_k -м контрольном году; P_{il_k} — количество научных работ i -го вида, возможно опубликованных всеми работниками института в l_k -м контрольном году; P_{l_k} — количество научных работ, возможно опубликованных всеми работниками института в

Таблица 3

Динамика численности работников в группах научной активности

| Группы научной активности | Годы | | | | | | | | | | | | |
|---------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| 1-я | 70 | 67 | 67 | 66 | 70 | 74 | 57 | 53 | 43 | 51 | 51 | 58 | 64 |
| 2-я | 58 | 57 | 60 | 64 | 61 | 57 | 52 | 48 | 49 | 46 | 49 | 45 | 41 |
| 3-я | 32 | 34 | 32 | 27 | 32 | 31 | 38 | 29 | 38 | 38 | 39 | 41 | 40 |
| 4-я | 17 | 20 | 18 | 22 | 25 | 24 | 23 | 23 | 22 | 24 | 26 | 21 | 19 |
| 5-я | 22 | 19 | 24 | 24 | 24 | 22 | 21 | 24 | 23 | 23 | 27 | 22 | 18 |

Таблица 4

Динамика среднего возраста работников в группах научной активности

| Группы научной активности | Годы | | | | | | | | | | | | |
|---------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| 1-я | 48,1 | 47,6 | 46,7 | 47,9 | 44,2 | 45,8 | 49,4 | 51,1 | 52,8 | 51,7 | 53,3 | 50,1 | 48,5 |
| 2-я | 44,6 | 44,1 | 45,3 | 43,1 | 48,1 | 47,4 | 48,8 | 51,2 | 49,8 | 49,2 | 47,1 | 51,3 | 53,8 |
| 3-я | 45,8 | 48,4 | 46,6 | 51,5 | 46,5 | 50,4 | 47,4 | 48,2 | 45,6 | 46,6 | 46,9 | 49,9 | 53,9 |
| 4-я | 52,2 | 53,5 | 53,7 | 54,3 | 53,1 | 54,6 | 56,1 | 57,9 | 55,7 | 57,0 | 60,9 | 61,1 | 61,5 |
| 5-я | 53,5 | 56,5 | 56,8 | 58,7 | 58,7 | 58,6 | 61,2 | 58,8 | 62,6 | 64,0 | 57,8 | 64,3 | 65,9 |

l_k -м контрольном году; p_{isl_k} — количество научных работ i -го вида, возможно опубликованных s -м работником в l_k -м контрольном году с учетом соавторов; S_{gl_k} — число работников g -й группы в l_k -м контрольном году.

В процессе моделирования в диалоговом режиме находятся такие параметры модели k_{gl_k} и ρ_θ , при которых отклонение показателей, рассчитанных по формулам (5)–(7), от фактических их значений было бы минимально.

3. Результаты решения задач. 3.1. После того, как у работников были рассчитаны текущие научные активности, решалась задача „Распределение работников по группам научной активности в каждый год“. Эта задача определяет динамику структурных показателей групп научной активности. К этим показателям можно отнести:

- численность работников в группах научной активности;
- средний возраст работников в группах научной активности;
- средняя научная активность работников в группах научной активности.

Результаты решения задачи приведены в табл. 3–5. Эти данные являются исходными для других задач.

Следует обратить внимание на тот факт, что численность работников в 3-й, 4-й и 5-й группах практически постоянна, а в 1-й и 2-й группах — имеет тенденцию к уменьшению. Последнее объясняется следующим. Основная часть работников, входящих в первые две группы имеют возраст от 40 до 50 лет. А как показали исследования, проведенные в работе [1], на протяжении последних 15 лет численность ученых именно этого возраста постоянно уменьшалась.

Как видно из таблицы, средний возраст у работников, входящих в 5-ю группу научной активности, больше 55 лет, а в 4-й группе он больше 55 последние 7 лет. Это подтверждает

Таблица 5

Динамика среднегодовой научной активности работников в группах научной активности

| Группы научной активности | Годы | | | | | | | | | | | | |
|---------------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
| 1-я | 0,22 | 0,20 | 0,20 | 0,19 | 0,15 | 0,15 | 0,17 | 0,18 | 0,18 | 0,17 | 0,17 | 0,15 | 0,14 |
| 2-я | 0,59 | 0,53 | 0,57 | 0,53 | 0,59 | 0,58 | 0,58 | 0,61 | 0,55 | 0,54 | 0,51 | 0,52 | 0,56 |
| 3-я | 0,99 | 1,05 | 0,98 | 1,09 | 0,86 | 1,01 | 0,90 | 1,01 | 0,89 | 0,87 | 0,91 | 0,96 | 1,05 |
| 4-я | 1,89 | 1,98 | 1,97 | 1,92 | 1,88 | 1,75 | 1,87 | 1,87 | 1,77 | 1,84 | 1,77 | 1,87 | 1,98 |
| 5-я | 3,68 | 3,99 | 3,61 | 3,58 | 3,48 | 3,58 | 3,62 | 3,27 | 3,40 | 3,39 | 2,95 | 3,26 | 3,29 |

тот очевидный вывод, что ученые в этом возрасте, как правило, более продуктивны. Расчеты показали, что доля этих ученых в 5-й группе научной активности от общего числа работников в этой группе меняется от 0,48 в 2001 до 0,89 в 2013 году. Для 4-й группы — 0,25 в 2001 году и 0,79 в 2013 году. Для определения указанных пропорций ученых разных возрастов в каждой группе научной активности была решена соответствующая задача.

Достаточно большой средний возраст в 1-й группе сформировался, потому что все же небольшая часть ученых в возрасте старше 55 лет малоактивна. Но доля их в этой группе относительно велика (от 0,23 до 0,47). Отсюда такой итог.

Практическое постоянство среднегодовой научной активности (табл. 5) во всех группах может говорить о правильном выборе граничных значений научной активности для групп.

3.2. Задача „Расчет оценки количеств публикаций научных работ в заданный год“ выполняет в разрабатываемой модели две функции. С одной стороны, она участвует в решении проблемы идентификации модели и в то же время является основой задачи прогнозирования количества публикаций научных работ в выбранный год (год прогноза). Задача реализует алгоритм, приведенный в п. 2.3. Для каждого работника определяется набор научных работ, которые он предположительно опубликовал в заданном контрольном году. По формулам (5)–(7) определяются показатели, введенные в п. 2.3. Определяются отклонения расчетных значений этих показателей от фактических значений (по группам научной активности, по видам научных работ и суммарно по всему институту), полученных в выбранном контрольном году. Отклонения определяются в процентном выражении

$$\Delta_{gl_k} = \frac{|P_{gl_k}^{\Phi} - P_{gl_k}|}{P_{gl_k}^{\Phi}} \cdot 100; \quad (8)$$

$$\Delta_{il_k} = \frac{|P_{il_k}^{\Phi} - P_{il_k}|}{P_{il_k}^{\Phi}} \cdot 100; \quad (9)$$

$$\Delta_{l_k} = \frac{|P_{l_k}^{\Phi} - P_{l_k}|}{P_{l_k}^{\Phi}} \cdot 100, \quad (10)$$

где $P_{gl_k}^{\Phi}$ — фактическое количество научных работ всех видов, опубликованных работниками g -й группы в l_k -м контрольном году;

$P_{il_k}^{\Phi}$ — фактическое количество научных работ i -го вида, опубликованных всеми работниками института в l_k -м контрольном году;

$P_{l_k}^{\Phi}$ — фактическое количество научных работ всех видов, опубликованных всеми работниками института в l_k -м контрольном году.

Результаты решения задачи идентификации приведены в табл. 6. Здесь годы сгруппированы по уровню влияния административного фактора на количество публикаций научных работ.

Таблица 6

Показатели и выбранные параметры имитационной модели

| Показатели и параметры модели | Годы | | |
|---------------------------------------------------------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| | 2002, 2006, 2007, 2009 | 2001, 2004, 2005 2010, 2012 | 2003, 2008, 2011, 2013 |
| Количество публикаций за год | $P_l < 260$ | $260 \leq P_l < 285$ | $P_l \geq 285$ |
| Коэффициент k_{gl_k} для 1-й, 2-й, 3-й, 4-й, 5-й групп соответственно | 3,14; 1,61; 1,89; 1,11; 0,99 | 4,18; 2,16; 1,84; 1,41; 1,40 | 4,47; 3,61; 2,36; 1,75; 1,39 |
| Коэффициент ρ_θ | 1,14 | 0,96 | 1,08 |
| Отклонения Δ_{gl_k} для 1-й, 2-й, 3-й, 4-й, 5-й групп, соответственно, % | 7,93; 4,95; 4,91; 5,03; 1,88 | 9,90, 5,50 8,03, 5,07 3,46 | 8,43; 8,22; 4,77; 4,51; 4,54 |
| Отклонения Δ_{il_k} для 1, 2, 4 видов научных работ соответственно, % | 2,72; 4,07; 2,75 | 8,98; 5,90; 5,25 | 9,17; 6,93; 4,08 |
| Отклонение Δ_{l_k} , % | 1,51 | 2,15 | 2,60 |

Отклонения для патентов (3-й вид научных работ) и монографий (5-й вид научных работ) не определялись в силу непрезентативности выборок.

3.3. Если решается задача определения количества научных работ, которые могут быть опубликованы в год прогноза, то вначале необходимо установить, какой фактор будет действовать в этот год, и уровень его влияния на научную активность работников. Эта процедура неформальна, и для ее решения возможен экспертный подход. Экспертами могут выступать администрация и сами ученые. Качество прогноза будет зависеть от адекватности итогов этой процедуры. Задача решается с параметрами для группы лет, в которую будет включен год прогноза (табл. 6).

Дальнейшее продолжение исследований в этом направлении состоит в построении имитационных моделей других подсистем научного процесса академического института и решении с их помощью актуальных задач.

Список литературы

1. КАРПАЧЕВ Г. И., МОИСЕЕНКО В. В. Некоторые вопросы ретроспективного анализа структуры и деятельности научных кадров академического института // Новосибирск: Проблемы информатики. № 1. 2009.

*Моисеенко Владислав Владимирович —
науч. сотр. Института вычислительной математики
и математической геофизики СО РАН,
р. т.: (+7383) 330-65-79*

Дата поступления — 06.11.2014