

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 9

Москва 2014

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 004.891 : 658.5

В.В. Мокеев

Об оценке деятельности предприятий методом собственных состояний*

Рассматривается анализ эффективности деятельности предприятий. Решение задачи базируется на построении эталонной модели предприятия и сравнении эталонной и фактической деятельности предприятия. При построении эталонной модели используется метод собственных состояний, позволяющий представить деятельность предприятия в виде суммы независимых составляющих (собственных состояний). Рассматриваются основные соотношения метода собственных состояний, формулируются свойства собственных состояний, предлагаются формулы для вычисления коэффициентов эффективности работы предприятий. Эффективность предлагаемого подхода демонстрируется на примере анализа эффективности работы филиалов энергосбытового предприятия. Полученные коэффициенты эффективности сравниваются с результатами, полученными методом анализа среды функционирования.

Ключевые слова: метод собственных состояний компонент, эффективность, экономический объект, анализ среды функционирования, анализ главных компонент

ВВЕДЕНИЕ

Создание моделей и методов для оценки эффективности деятельности предприятий и организаций

является актуальной задачей. В настоящее время существует большое количество простых методов для анализа деятельности предприятий. В таких методах для конкретного объекта вычисляется набор коэффициентов, которые затем усредняются, и полученное значение является оценкой развития предприятий.

* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 14-01-00054).

Распространенной формой оценки деятельности предприятий является вычисление коэффициентов эффективности в виде отношения суммы выходных показателей к сумме входных показателей. Так как разные показатели имеют различную степень важности, им присваиваются веса, значения которых определяют эксперты. Основным способом оценки эффективности экономических объектов является анализ среды функционирования (Data Envelopment Analysis) – непараметрический метод оценки эффективности объектов в различных отраслях экономики. Анализ среды функционирования (АСФ) рассматривает любую сложную систему как объект с множеством входов и выходов и определяет его эффективность как отношение суммы взвешенных выходов к сумме взвешенных входов, но веса входов и выходов вычисляются с помощью строгой математической процедуры [1–3]. Методология АСФ направлена на оценку эффективности ряда одноуровневых объектов, которые называются Единицами принятия решения (Decision Making Unit – DMU). В дальнейшем будем называть их объектами.

Считается, что эффективность объекта в АСФ хорошо согласуется со следующими определениями. Во-первых, полная эффективность достигается любым объектом, если и только если ни один из его входов или выходов не может быть улучшен, не ухудшая его другие входы или выходы. Во-вторых, объект может быть оценен как полностью эффективный относительно других объектов, если анализ эффективности других объектов показывает, что улучшение некоторых их входов или выходов не приводит к ухудшению других их входов или выходов. Эффективность объектов измеряется относительно других объектов с помощью простых ограничений, при которых все объекты лежат либо ниже, либо на границе относительной эффективности. АСФ анализирует каждый объект по отдельности и вычисляет меру максимальной эффективности для каждого объекта.

Классический вариант АСФ позволяет получать только коэффициенты относительной эффективности объектов. При этом объекты, расположенные на границе эффективности, могут улучшать свои показатели. В работе [4] предлагается в АСФ использовать искусственную границу эффективности для повышения качества функционирования объектов, эффективность которых подтверждена опытным путем. Для повышения эффективности АСФ в работах [5, 6] используется анализ главных компонент. Как известно, качество применения АСФ снижается, когда число входов и выходов становится большим. Для снижения числа входов и выходов используется метод главных компонент. Альтернативная методология – анализ оболочек со свободным распределением – описывается в работе [7]. Методология не учитывает выпуклость границы, как это делает АСФ, поэтому число эффективных объектов растет до такой степени, что результаты становятся слишком малоинформативными, чтобы делать какие-либо выводы [8].

Метод собственных состояний использует простую процедуру построения модели эффективной работы объекта, которая базируется на методе главных компонент [9, 10]. При этом деятельность предпри-

ятия описывается набором автономных состояний, которые можно анализировать независимо друг друга. Каждая из таких составляющих называется собственным состоянием. В рамках каждого такого состояния показатели могут меняться строго определенным образом, что позволяет формировать модель только из тех состояний, которые удовлетворяют требованиям эффективности.

МЕТОД СОБСТВЕННЫХ СОСТОЯНИЙ

Пусть состояние экономических объектов описывается набором показателей x_{ki}^0 , где i – номер показателя ($i = 1, 2, 3, \dots, n$), k – номер экономического объекта ($k = 1, 2, 3, \dots, m$), n – количество показателей, m – количество экономических объектов. Значения показателей для k -го экономического объекта образуют вектор $\mathbf{x}_k^0 = \{x_{k1}^0, x_{k2}^0, \dots, x_{kn}^0\}$. Пространство показателей экономических объектов можно представить в виде матрицы \mathbf{X}^0 , где каждый столбец содержит значения одного показателя для различных экономических объектов, а каждая строка включает значения всех показателей и описывает состояние одного экономического объекта. Таким образом, пространство состояний экономических объектов будет описываться в виде

$$\mathbf{X}^0 = [x_1^0 \quad x_2^0 \quad \dots \quad x_m^0]^T. \quad (1)$$

Главные компоненты представляют такую группировку исходных показателей, в которой члены группы (исходные показатели) связаны между собой, но группа (главная компонента) в целом является независимой от других групп (главных компонент). Для расчета весовых коэффициентов главных компонент решается задача собственных значений

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = 0, \quad (2)$$

где \mathbf{A} – матрица размерностью $n \times n$, \mathbf{I} – единичная матрица, \mathbf{v} – собственный вектор уравнения (2), λ – собственное значение. В качестве матрицы \mathbf{A} , как правило, используется ковариационная матрица (данные центрируются), однако можно использовать матрицу начальных вторых моментов (данные не центрируются), а также корреляционную матрицу (данные центрируются и нормируются).

Матрица собственных состояний \mathbf{V} формируется из собственных векторов уравнения (2). Собственные векторы уравнения (2) обладают свойством ортогональности и масштабируются так, что

$$\mathbf{V}^T \mathbf{A} \mathbf{V} = \mathbf{\Lambda} \quad \text{и} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}, \quad (3)$$

где $\mathbf{\Lambda}$ – диагональная матрица, диагональные коэффициенты которой равны собственным значениям уравнения (2).

Если при вычислении собственных векторов используется ковариационная или корреляционная матрица, показатели определяются по формуле

$$\mathbf{x}_k^0 = \bar{\mathbf{x}} + \sum_{i=1}^p \mathbf{v}_i z_{ik}, \quad (4)$$

где z_{ik} – значение главной компоненты i -го собственного состояния k -го объекта, \bar{x} – вектор средних значений показателей. Если при вычислении собственных векторов используется матрица начальных вторых моментов, то $\bar{x} = 0$.

Каждый собственный вектор имеет ту же размерность, что и вектор состояния экономического объекта, что позволяет интерпретировать его как некоторое собственное состояние объекта. Так как собственный вектор определяется с точностью до множителя, коэффициенты собственного вектора показывают не столько величину показателей, сколько их взаимосвязь друг с другом. В дальнейшем коэффициенты собственного вектора будем называть коэффициентами собственного состояния.

В общем случае главные компоненты вычисляются по формуле

$$z_{ki} = \sum_{h=1}^p v_{hi} (x_{ki}^0 - \bar{x}_i), \quad (5)$$

где z_{ki} – значение главной компоненты i -го собственного состояния k -го экономического объекта,

v_{hi} – коэффициент i -го собственного состояния, соответствующий h -му исходному показателю, p – число собственных состояний. Если собственные состояния определяются по матрице начальных вторых моментов, то $\bar{x}_i = 0$

Значения i главной компоненты для различных состояний экономических объектов объединяются в вектор z_i , из которых образуется матрица главных компонент Z , которая определяется как

$$Z = XV. \quad (6)$$

Коэффициенты матрицы X в случае вычисления собственных состояний с помощью ковариационной матрицы определяются по формуле $x_{ki} = x_{ki}^0 - \bar{x}_i$. В случае, если при вычислении собственных состояний применяется матрица начальных вторых моментов, коэффициенты матрицы X равны коэффициентам матрицы X^0 .

Таким образом, состояние любого экономического объекта представляется взвешенной комбинацией собственных состояний. Это означает, что экономический объект описывается уже не набором исходных показателей, а набором главных компонент, но каждая главная компонента отражает уже не отдельный исходный показатель, а группу исходных показателей (собственное состояние объекта).

Если собственные векторы вычисляются по ковариационной матрице, то собственные значения показывают изменчивость собственного состояния в общем состоянии экономического объекта и численно равны дисперсии, которая аккумулирует собственное состояние. Собственные состояния обладают рядом полезных свойств.

Свойство 1. При описании состояния экономического объекта в виде взвешенной суммы собственных состояний, главные компоненты собственных состояний являются статистически независимыми.

Доказательство. Статистическая независимость означает, что любое приращение главной компоненты j -го собственного состояния не приводит к изменениям главных компонент, соответствующих другим собственным состояниям. Такая формулировка независимости главных компонент означает, что матрица $Z^T Z$ является диагональной. Нетрудно убедиться, что это действительно так:

$$Z^T Z = (XV)^T XV = V^T X^T XV = V^T AV = \Lambda, \quad (7)$$

где Λ – диагональная матрица.

Свойство 2. Изменение состояния объекта, связанное с изменением главной компоненты j -го собственного состояния, может происходить только пропорционально коэффициентам j -го собственного состояния.

Доказательство. Состояние любого объекта представляется в виде суммы независимых собственных состояний. Пусть главная компонента j -го собственного состояния получает единичное приращение. В соответствии со свойством такое изменение j -й главной компоненты не вызывает изменения остальных главных компонент.

Пусть j -я главная компонента получает приращение Δz , тогда изменение исходных показателей описывается как

$$x_k^0 + \Delta x = \bar{x} + \sum_{i=1}^p v_i z_{ik} + v_j \Delta z. \quad (8)$$

Вычитая из уравнения (8) соотношение (4), получим

$$\Delta x = v_j \Delta z. \quad (9)$$

Из уравнения (9) видно, что если изменение состояния объекта обусловлено изменением j -й главной компоненты, то показатели, описывающие состояние объекта, могут изменяться только пропорционально коэффициентам j -го собственного состояния.

Анализ собственных состояний объекта строится на проверке того, удовлетворяет ли собственное состояние требованиям эффективности. Собственные состояния, которые не удовлетворяют этим требованиям, отбрасываются, оставшиеся собственные состояния используются для построения эталонной модели. Эталонная модель используется для описания эталонной деятельности объектов. Показатели эталонной деятельности объектов X^{et} определяются по формуле

$$X^{et} = V^{et} Z^{et}, \quad (10)$$

где V^{et} – матрица весовых коэффициентов эталонных собственных состояний, Z^{et} – матрица главных компонент эталонных собственных состояний.

Эффективность объекта будем оценивать по величине отклонений базовых показателей объекта от их эталонных значений. Для этого вычисляется штрафная функция по каждому базовому показателю. Базовые показатели выбираются из числа исходных показателей.

Если отрицательная величина отклонений фактических значений j -го базового показателя k -го объекта от эталонных значений является признаком неэффективности, то штрафная функция равна

$$f_{kj} = \begin{cases} |(x_{kj} - x_{kj}^{et}) / x_{kj}^{et}| & \text{если } (x_{kj} - x_{kj}^{et}) < 0 \\ 0 & \text{если } (x_{kj} - x_{kj}^{et}) > 0 \end{cases} \quad (11)$$

И наоборот, если положительная величина отклонений фактических значений j -го базового показателя k -го объекта от эталонных значений является признаком неэффективности, то в этом случае штрафная функция определяется как

$$f_{kj} = \begin{cases} |(x_{kj} - x_{kj}^{et}) / x_{kj}^{et}| & \text{если } (x_{kj} - x_{kj}^{et}) > 0 \\ 0 & \text{если } (x_{kj} - x_{kj}^{et}) < 0 \end{cases} \quad (12)$$

Штрафные функции входных показателей определяются по формуле (11), а для выходных показателей они вычисляются по формуле (12). Для каждого объекта определяется среднеквадратическое значение штрафных функций

$$\bar{f}_k = \sqrt{\frac{1}{r} \sum_{j=1}^r f_{kj}^2}, \quad (13)$$

где r – число показателей с ненулевыми штрафными функциями.

Коэффициент эффективности k -го объекта может быть получен по формуле

$$R_k = 1 - \bar{f}_k \quad (14)$$

АНАЛИЗ ЭФФЕКТИВНОСТИ РАБОТЫ ЭНЕРГОСБЫТОВЫХ КОМПАНИЙ

Анализ проводится для семи филиалов энергосбытового предприятия. В качестве основных показателей, описывающих деятельность филиалов, выделяются 15 показателей: y_1^0 – количество контролеров на линии, y_2^0 – количество сотрудников, осуществляющих обработку документов, y_3^0 – общее количество сотрудников, y_4^0 – количество расчетных точек учета, x_1^0 – количество договоров с гражданами-потребителями, x_2^0 – количество договоров с потребителями-юридическими лицами, x_3^0 – количество договоров с потребителями-бюджетными организациями, x_4^0 – среднее количество уведомлений, направляемых потребителям, находящимся на оперативном контроле, x_5^0 – среднее количество ежедневно обрабатываемых документов при разноске оплаты, x_6^0 – среднее количество полезного отпуска электроэнергии, x_7^0 – количество снятых показаний учета граждан-потребителей.

Анализируется эффективность работы семи филиалов: Челябинский филиал (ЧФ), Метэнергосбыт (МЭС), Центральный филиал (ЦФ), Златоустовский

филиал (ЗФ), Магнитогорский филиал (МФ), Троицкий филиал (ТФ), Кыштымский филиал (КФ).

1. На первой стадии исследований оценка эффективности работы филиалов выполняется методом АСФ. Для этого используется входная модель с постоянной отдачей от масштаба CRS (constant-return to scale). Целью входной модели является минимизация затрат без уменьшения объемов производства предприятий. В случае постоянной отдачи масштаба, выходные показатели изменяются пропорционально входным факторам. Для проведения расчетов используется программа DEAFrontier [11]. Показатели y_1^0, y_2^0, y_3^0 характеризуют трудовой потенциал филиалов и вместе с показателем y_4^0 образуют группу входных показателей, показатели $x_1^0, x_2^0, x_3^0, x_4^0, x_5^0, x_6^0, x_7^0$ описывают результаты деятельности филиалов. Эффективность работы филиалов оценивается по четырем вариантам выходных показателей, каждому из которых соответствует три варианта входных показателей. В табл. 1 представлены результаты расчета эффективности деятельности филиалов для 12 вариантов входных и выходных данных.

Как видно из табл. 1, эффективность деятельности филиалов при использовании АСФ зависит от числа входных и выходных показателей. Использование АСФ позволяет определить относительное положение филиала в пространстве входных и выходных параметров на основе расчета относительной результативности каждого филиала. Недостатком АСФ является то, что при большом числе показателей качество оценки результативности филиалов снижается, так как практически все филиалы оцениваются как эффективные.

2. На второй стадии исследований эффективность филиалов оценивается методом собственных состояний. Метод собственных состояний, в отличие от метода АСФ, для оценки деятельности филиалов использует эталонную модель, которая строится исходя из требований, предъявляемых к работе филиалов. В рамках этих требований формулируются правила изменения показателей деятельности филиалов, которая будет в дальнейшем считаться «эталонной».

Требования к эталонной работе филиалов формулируются следующим образом:

1) увеличение (сокращение) численности персонала y_3^0 должно приводить к росту (уменьшению) количества договоров с гражданами-потребителями x_1^0 , числа договоров с потребителями-юридическими лицами x_2^0 , среднего количества уведомлений, направляемых потребителям, находящимся на оперативном контроле x_4^0 , величины среднего количества ежедневно обрабатываемых документов x_5^0 ;

2) увеличение (сокращение) числа сотрудников, осуществляющих обработку документов, y_2^0 , должно приводить к росту (уменьшению) среднего коли-

чества ежедневно обрабатываемых документов при разности оплаты x_5^0 .

На изменение других показателей никаких ограничений в рамках эталонной деятельности филиалов не накладывается. Отметим, что эти требования от-

личаются от условий, для которых формируется входная модель АСФ.

В табл. 2 представлены коэффициенты первых шести собственных состояний, полученных по ковариационной матрице.

Таблица 1

Оценка эффективности деятельности филиалов методом АСФ

Выход	Вход	ЧФ	МЭС	ЦФ	ЗФ	МФ	ТФ	КФ
$x_1^0, x_2^0,$	y_3^0	1,000	0,770	0,864	0,987	0,571	0,738	0,961
	y_1^0, y_3^0	1,000	1,000	0,864	0,987	0,571	0,738	0,961
	y_1^0, y_2^0, y_3^0	1,000	1,000	0,864	0,987	0,571	0,738	0,961
$x_1^0, x_2^0,$ x_5^0	y_3^0	1,000	0,770	0,873	1,000	0,757	0,738	0,962
	y_1^0, y_3^0	1,000	1,000	0,920	1,000	1,000	0,738	0,962
	y_1^0, y_2^0, y_3^0	1,000	1,000	0,920	1,000	1,000	0,738	0,962
$x_1^0, x_2^0, x_3^0,$ x_4^0, x_5^0	y_3^0	1,000	0,813	0,961	1,000	0,952	0,747	0,972
	y_1^0, y_3^0	1,000	1,000	1,000	1,000	1,000	0,780	1,000
	y_1^0, y_2^0, y_3^0	1,000	1,000	1,000	1,000	1,000	0,925	1,000
$x_1^0, x_2^0, x_3^0,$ $x_4^0, x_5^0, x_6^0,$ x_7^0	y_3^0	1,000	1,000	1,000	1,000	0,952	0,747	0,982
	y_1^0, y_3^0	1,000	1,000	1,000	1,000	1,000	0,805	1,000
	y_1^0, y_2^0, y_3^0	1,000	1,000	1,000	1,000	1,000	0,925	1,000

Таблица 2

Коэффициенты собственных состояний

Показатель	Номер собственного состояния					
	1	2	3	4	5	6
y_1^0	0,0147	0,08375	-0,0003	0,16393	-0,3423	0,30942
y_2^0	0,0051	0,00701	-0,0193	0,08971	-0,0155	-0,0917
y_3^0	0,0385	0,39917	0,22988	0,74451	-0,1327	-0,16
y_4^0	0,9846	-0,1344	0,10964	-0,018	-0,0084	0,00298
x_1^0	0,0342	0,62692	0,36618	-0,554	-0,1979	-0,3408
x_2^0	0,0013	0,01373	0,00942	-0,0061	-0,0122	-0,0145
x_3^0	0,0238	0,09545	-0,0003	0,3017	-0,1595	-0,3093
x_4^0	0,0024	0,01848	0,0082	-0,0075	-0,0276	0,08974
x_5^0	0,1615	0,48712	-0,8519	-0,0203	0,08886	-0,0197
x_6^0	0,0266	0,36395	0,18003	-0,0212	-0,0204	0,80596
x_7^0	0,0173	0,20764	0,20592	0,10763	0,88944	-0,0146

Как видно из табл.2, коэффициенты первого собственного состояния имеют одинаковый знак, поэтому все показатели при изменении первой главной компоненты будут одновременно либо увеличиваться, либо уменьшаться. Это означает, что первое собственное состояние удовлетворяет условиям эталонной деятельности. Коэффициенты второго собственного состояния для показателей y_3^0 , x_1^0 , x_2^0 , x_4^0 , x_6^0 и y_2^0 , x_5^0 также имеют одинаковый знак, поэтому изменение этих показателей в рамках второго собственного состояния будет приводить либо к росту одновременно всех показателей, либо к их уменьшению. Таким образом, второе собственное состояние также может использоваться при построении эталонной модели.

Коэффициенты показателей y_3^0 , x_1^0 , x_2^0 , x_4^0 , x_6^0 третьего собственного состояния имеют знак плюс, а коэффициенты показателей y_2^0 , x_5^0 – знак минус. Увеличение численности персонала y_3^0 будет сопровождаться увеличением показателей деятельности филиалов x_1^0 , x_2^0 , x_4^0 , x_6^0 , но при этом будет уменьшаться количество сотрудников, осуществляющих обработку документов, y_2^0 , и число ежедневно обраба-

тываемых документов при разноске оплаты x_5^0 . Таким образом, третье собственное состояние также подходит для построения эталонной модели. Анализ остальных собственных состояний показывает, что изменение показателей в рамках этих собственных состояний не соответствует условиям эталонной деятельности филиалов. В табл. 2 жирным шрифтом выделены собственные состояния, которые используются при построении эталонной модели.

В табл. 3 представлены коэффициенты эффективности, полученные с использованием эталонной модели, состоящей из первых трех собственных состояний. Как видно из таблицы, коэффициенты эффективности с увеличением количества показателей уменьшаются, что связано с тем, что растет число признаков неэффективности, которые определяются при сравнении эталонных и фактических значений показателей. Анализ результатов, представленных в табл. 3, показывает, что наименее эффективными филиалами являются МЭС, МФ и ТФ, а наиболее эффективными – ЧФ, ЦФ, ЗФ и КФ. Причины неэффективности деятельности филиалов можно определить путем анализа значений штрафных функций. В табл. 4 представлены значения штрафных функций для различных показателей. Проверка означает отсутствие штрафа, т.е. штрафная функция равна нулю.

Таблица 3

Оценка эффективности деятельности филиалов методом собственных состояний

Выход	Вход	ЧФ	МЭС	ЦФ	ЗФ	МФ	ТФ	КФ
$x_1^0, x_2^0,$	y_3^0	1,000	0,926	0,954	1,000	0,886	0,898	1,000
	y_1^0, y_3^0	1,000	0,926	0,954	0,986	0,886	0,774	1,000
	y_1^0, y_2^0, y_3^0	1,000	0,926	0,953	0,986	0,867	0,778	1,000
$x_1^0, x_2^0,$ x_5^0	y_3^0	1,000	0,926	0,954	1,000	0,886	0,886	1,000
	y_1^0, y_3^0	1,000	0,926	0,954	0,986	0,886	0,793	1,000
	y_1^0, y_2^0, y_3^0	1,000	0,926	0,953	0,986	0,867	0,792	1,000
$x_1^0, x_2^0, x_3^0,$ x_4^0, x_5^0	y_3^0	0,973	0,695	0,956	0,972	0,813	0,886	0,883
	y_1^0, y_3^0	0,973	0,695	0,956	0,978	0,813	0,793	0,883
	y_1^0, y_2^0, y_3^0	0,973	0,695	0,955	0,978	0,815	0,792	0,883
$x_1^0, x_2^0, x_3^0,$ $x_4^0, x_5^0, x_6^0,$ x_7^0	y_3^0	0,973	0,695	0,956	0,929	0,828	0,840	0,853
	y_1^0, y_3^0	0,973	0,695	0,956	0,941	0,828	0,783	0,853
	y_1^0, y_2^0, y_3^0	0,973	0,695	0,955	0,941	0,827	0,785	0,853

Коэффициенты штрафных функций показателей

	ЧФ	МЭС	ЦФ	ЗФ	МФ	ТФ	КФ
y_1^0	-	-	-	0,01	-	0,36	-
y_2^0	-	-	0,05	-	0,18	0,21	-
y_3^0	-	-	0,04	-	0,13	0,11	-
x_1^0	-	-	0,06	-	0,14	0,09	-
x_2^0	-	0,07	0,04	-	0,05	-	-
x_3^0	-	0,42	-	-	-	-	0,02
x_4^0	0,03	-	0,04	-	0,32	-	0,16
x_5^0	-	-	-	-	-	0,14	-
x_6^0	-	-	-	-	0,09	-	0,16
x_7^0	-	-	-	-	-	0,25	0,18

Как видно из табл.4, признаки неэффективности филиала МЭС выявлены для работ, описываемых показателем x_3^0 (количество договоров с потребителями-бюджетными организациями). Неэффективность филиала МФ проявляется в деятельности, описываемой показателями y_2^0 (количество сотрудников, осуществляющих обработку документов) и x_4^0 (среднемесячное количество уведомлений, направляемых потребителям, находящимся на оперативном контроле). Филиал ТФ проявляет признаки неэффективности в деятельности, описываемой показателями y_1^0 (количество контролеров на линии), y_2^0 (количество сотрудников, осуществляющих обработку документов), x_7^0 (количество снятых показаний учета граждан-потребителей).

На рис. 1 показано сравнение фактической численности персонала филиалов и численности сотрудников, полученной по эталонной модели. Если фактическая численность персонала превышает количество сотрудников, необходимых для эталонной деятельности филиала, то это является признаком неэффективности их работы. На рис. 1 зоны неэффективности закрашены серым цветом.

Сравнение фактического количества договоров с гражданами-потребителями с числом договоров, соответствующих эталонной деятельности филиалов, демонстрируется на рис. 2. Их сравнение позволяет выявить признаки неэффективности деятельности филиалов, которые возникают, когда фактическое

число договоров меньше эталонного. Зоны неэффективности закрашены на рисунке серым цветом.

Недостатком эталонной модели, построенной с использованием собственных состояний, вычисленных с помощью ковариационной матрицы, является то, что собственные состояния описывают отклонения показателей от их средних значений. Если средние значения по какой-либо причине меняются, эталонная модель нуждается в корректировке. Поэтому интерес приставляет способ построения эталонной модели, не использующий центрирование данных.

3. На третьей стадии исследований оценка эффективности работы филиалов выполняется методом собственных состояний, но собственные состояния вычисляются по матрице начальных вторых моментов (данные не центрируются). В табл. 5 представлены коэффициенты первых шести собственных состояний, вычисленных по матрице начальных вторых моментов. Анализ результатов, представленных в табл. 5, показывает, что в рамках первого второго и третьего собственного состояния такие показатели, как численность персонала y_3^0 , количество договоров с гражданами-потребителями x_1^0 , число договоров с потребителями-юридическими лицами x_2^0 , среднемесячное количество уведомлений, направляемых потребителям, находящимся на оперативном контроле x_4^0 , величина среднемесячного отпуска электроэнергии x_6^0 либо увеличиваются, либо уменьшаются одновременно.

Коэффициенты собственных состояний

Показатель	Номер собственного состояния					
	1	2	3	4	5	6
y_1^0	0,017	0,043	0,073	0,130	-0,230	0,478
y_2^0	0,005	-0,006	0,013	0,098	-0,001	-0,049
y_3^0	0,066	0,451	0,230	0,769	0,003	-0,083
y_4^0	0,986	-0,021	-0,161	-0,027	-0,010	0,002
x_1^0	0,061	0,590	0,388	-0,478	-0,348	-0,363
x_2^0	0,002	0,014	0,008	-0,007	-0,012	0,000
x_3^0	0,029	0,070	0,078	0,330	-0,099	-0,113
x_4^0	0,002	0,011	0,014	-0,030	-0,009	0,121
x_5^0	0,125	-0,531	0,832	0,024	0,068	-0,054
x_6^0	0,039	0,303	0,243	-0,175	0,080	0,769
x_7^0	0,029	0,259	0,091	-0,106	0,897	-0,094

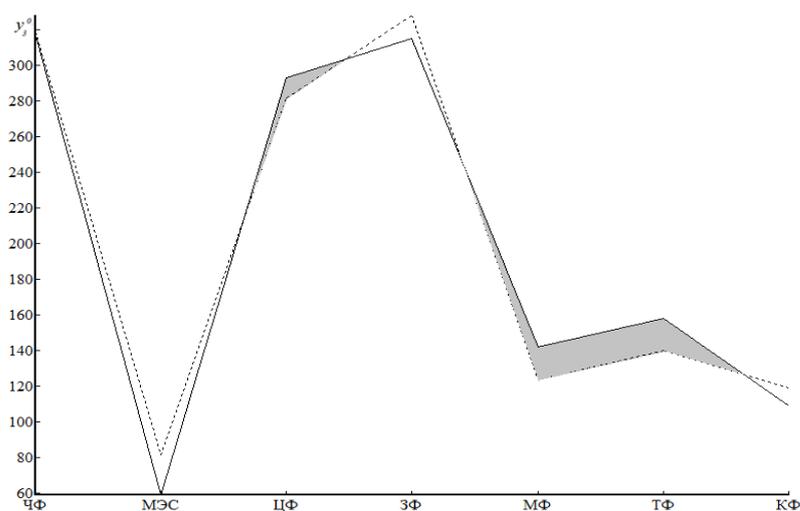


Рис. 1. Численность сотрудников филиалов: сплошная линия – фактические значения, штриховая линия – эталонные значения

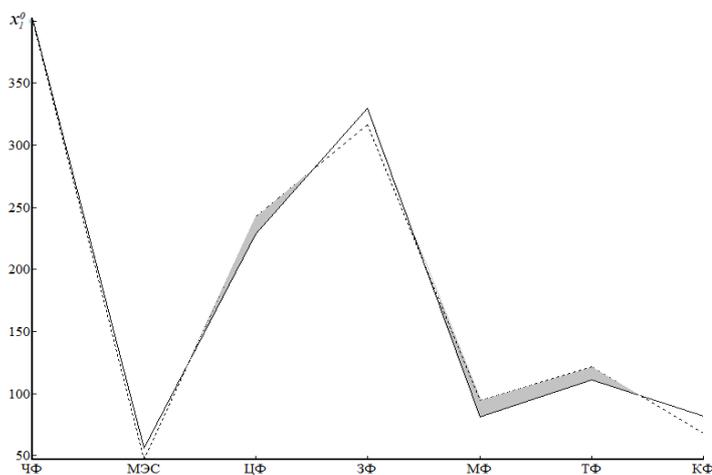


Рис. 2. Количество договоров с гражданами-потребителями: сплошная линия – фактические значения, штриховая линия – эталонные значения

В рамках первого и третьего собственного состояния рост численности персонала y_3^0 сопровождается увеличением числа сотрудников, осуществляющих обработку документов, y_2^0 , но при этом растет количество обрабатываемых документов при разности оплаты x_8^0 . Таким образом, первое и третье собственное состояние полностью соответствуют условиям эталонной деятельности филиалов. В рамках второго собственного состояния увеличение численности персонала y_3^0 сопровождается уменьшением числа сотрудников, осуществляющих обработку документов, y_2^0 , и сокращением количества обрабатываемых документов при разности оплаты x_8^0 . Поэтому второе собственное состояние также удовлетворяет условиям эталонной деятельности филиалов. Анализ четвертого, пятого и шестого собственных состояний показывает, что они не удовлетворяют условиям эталонной деятельности. Таким образом, эталонная модель формируется из первого, второго и третьего собственных состояний.

В табл. 6 представлены коэффициенты эффективности R_k для 12 вариантов входных и выходных показателей, полученные путем сравнения эталонной и фактической деятельности филиалов. Отклонения показателей фактической деятельности от эталонной можно интерпретировать как источник информации об эффективной или неэффективной работе филиала.

Как видно из табл.6, увеличение числа показателей не приводит к снижению качества оценки дея-

тельности филиалов. Коэффициенты эффективности для вариантов с большим числом показателей немного снижаются по сравнению с вариантами с небольшим числом показателей. Это связано с тем, что при увеличении числа показателей растет число источников неэффективности, которое может быть выявлено при анализе показателей.

Анализ результатов показывает, что наименее эффективными филиалами являются МЭС, МФ и ТФ, наиболее эффективными – ЧФ, ЦФ, ЗФ и КФ. Эффективность деятельности филиалов определяется через штрафные функции показателей. Значения штрафных функций представлены в табл. 7. В основном они совпадают со штрафными функциями, полученными с помощью собственных состояний, вычисленных по ковариационной матрице. Неэффективность филиала МЭС также связана с работой, описываемой показателем x_3^0 (количество договоров с потребителями-бюджетными организациями). Деятельность филиала МФ также имеет большие штрафы в показателях y_2^0 (количество сотрудников, осуществляющих обработку документов) и x_4^0 (среднемесячное количество уведомлений, направляемых потребителям, находящимся на оперативном контроле). Филиал ТФ проявляет признаки неэффективности в деятельности, описываемой показателями y_1^0 (количество контролеров на линии), y_2^0 (количество сотрудников, осуществляющих обработку документов), x_7^0 (количество снятых показаний учета граждан-потребителей).

Таблица 6

Коэффициенты эффективности работы филиалов, полученные методом собственных состояний

Выход	Вход	ЧФ	МЭС	ЦФ	ЗФ	МФ	ТФ	КФ
x_1^0, x_2^0	y_3^0	1,000	0,859	0,975	1,000	0,829	0,903	1,000
	y_1^0, y_3^0	1,000	0,859	0,975	1,000	0,829	0,882	0,948
	y_1^0, y_2^0, y_3^0	1,000	0,681	0,975	0,987	0,729	0,882	0,944
x_1^0, x_2^0, x_5^0	y_3^0	0,980	0,681	0,975	0,957	0,750	0,836	0,862
	y_1^0, y_3^0	1,000	0,859	0,975	1,000	0,829	0,767	1,000
	y_1^0, y_2^0, y_3^0	1,000	0,859	0,975	1,000	0,829	0,784	0,948
$x_1^0, x_2^0, x_3^0, x_4^0, x_5^0$	y_3^0	1,000	0,681	0,975	0,987	0,729	0,784	0,944
	y_1^0, y_3^0	0,980	0,681	0,975	0,957	0,750	0,776	0,862
	y_1^0, y_2^0, y_3^0	1,000	0,859	0,975	1,000	0,807	0,780	1,000
$x_1^0, x_2^0, x_3^0, x_4^0, x_5^0, x_6^0, x_7^0$	y_3^0	1,000	0,859	0,975	1,000	0,807	0,792	0,948
	y_1^0, y_3^0	1,000	0,681	0,975	0,987	0,733	0,792	0,944
	y_1^0, y_2^0, y_3^0	0,980	0,681	0,975	0,957	0,751	0,783	0,862

На рис. 3 показано сравнение фактической численности персонала филиалов и численности сотрудников, полученной по эталонной модели, а на рис. 4 – сравнение фактического количества договоров с гражда-

нами-потребителями с числом договоров, соответствующих эталонной деятельности филиалов. На рисунках зоны неэффективности закрашены серым цветом.

Таблица 7

Коэффициенты штрафных функций показателей

	ЧФ	МЭС	ЦФ	ЗФ	МФ	ТФ	КФ
y_1^0	-	-	-	-	-	0,38	-
y_2^0	-	-	0,02	-	0,25	0,17	-
y_3^0	-	-	0,02	-	0,17	0,11	-
x_1^0	-	-	0,03	-	0,21	0,08	-
x_2^0	-	0,14	0,02	-	0,12	-	-
x_3^0	-	0,43	-	0,01	-	-	0,06
x_4^0	-	-	-	-	0,45	-	-
x_5^0	-	-	-	-	-	0,15	0,05
x_6^0	-	-	-	0,01	0,13	-	0,09
x_7^0	0,02	-	-	0,07	-	0,26	0,25

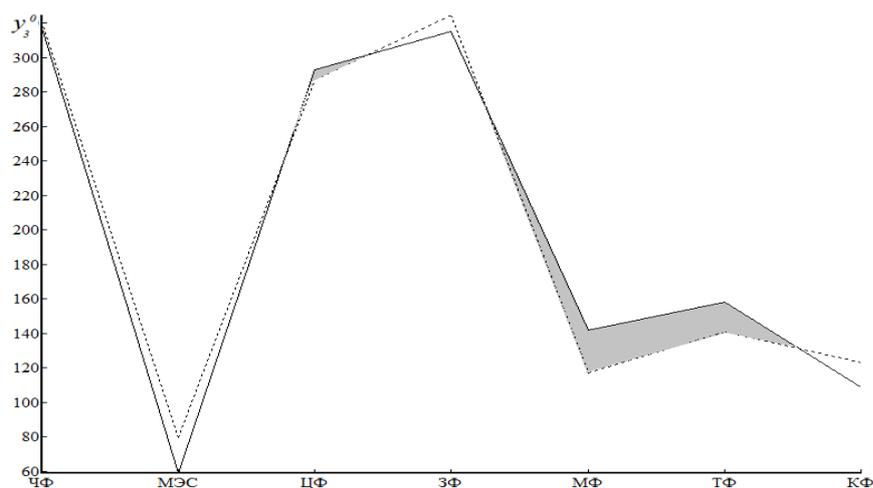


Рис. 3. Численность сотрудников филиалов: сплошная линия – фактические значения, штриховая линия – эталонные значения

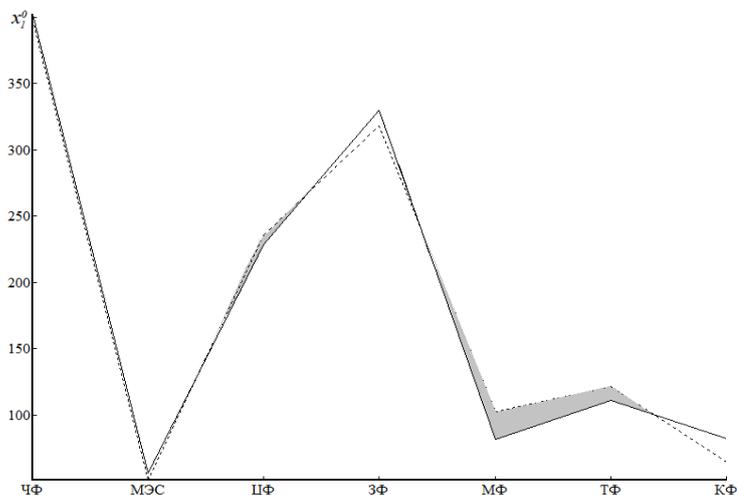


Рис. 4. Количество договоров с гражданами-потребителями: сплошная линия – фактические значения, штриховая линия – эталонные значения

Сравнение коэффициентов эффективности, полученных методом собственных состояний с использованием ковариационной матрицы и матрицы начальных вторых моментов, показывает их хорошее совпадение. Сопоставление результатов, полученных методом АСФ и методом собственных состояний, демонстрирует близость оценок при небольшом количестве показателей.

ЗАКЛЮЧЕНИЕ

Рассмотрена задача оценки эффективности функционирования предприятий. Для решения задачи предложен новый метод – метод собственных состояний. Рассмотрены основные соотношения метода собственных состояний, предложены формулы для вычисления коэффициентов эффективности экономических объектов. Возможности метода демонстрируются для задачи анализа эффективности деятельности филиалов энергосбытовой компании. Для решения задачи используется как метод собственных состояний, так и метод АСФ. Исследования методом АСФ показали, что при большом числе показателей граница эффективности размывается так, что практически все филиалы оцениваются как эффективные. Оценка деятельности филиалов методом собственных состояний выполняется с использованием эталонной модели. Рассмотрены различные способы построения эталонной модели. Получены коэффициенты эффективности филиалов, а также показаны источники неэффективности деятельности филиалов.

СПИСОК ЛИТЕРАТУРЫ

1. Charnes A., Cooper W.W., Rhodes E. Measuring the Efficiency of Decision Making Units // European Journal of Operational Research. – 1978. – Vol. 2. – P. 429–444.
2. Cooper, W.W., Seiford L., M. Tone K. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References, and DEA-Solver Software. – Boston: Kluwer Academic Publishers, 2000. – 318 p.
3. Кривоножко В.Е., Лычев А.В. Анализ деятельности сложных социально-экономических систем. – М.: Издательский отдел факультета ВМиК МГУ им. М.В. Ломоносова; МАКС Пресс, 2010. – 208 с.

4. Sowlati T., Paradi J.C. Establishing the «practical frontier» in data envelopment analysis // Omega. – 2004. – Vol. 32. – P. 261–272.
5. Adler N., Golany B. Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe // European Journal of Operational Research. – 2001. – Vol.132. – P. 260–273.
6. Adler N., Yazhemy E. Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction // European Journal of Operations Research. – 2010. – Vol. 202(1). –P. 273–284.
7. Tulkens H. On FDH efficiency analysis: some methodological issues and applications to retail banking, courts and urban transport // Journal of Productivity Analysis. – 1993. – Vol. 4. – P. 183–210.
8. Athanassopoulos A.D., Storbeck J.E. Non-parametric models for spatial efficiency. // Journal of Productivity Analysis. – 1995. – Vol.6. – P. 225–45.
9. Мокеев В.В. Решение проблемы собственных значений в задачах многофакторного анализа экономических систем // Экономика и математические методы. – 2010. – № 4. – С. 82–90.
10. Мокеев В.В., Плужников В.Г. Анализ главных компонент как средство повышения эффективности управленческих решений в предпринимательских структурах // Вестник Южно-Уральского государственного университета: Экономика и менеджмент. –2011. – № 41(258), вып. 20. – С. 149–154.
11. Zhu J. Running the DEA Frontier Software. – URL: www.deafrontier.com (дата обращения: 12.03.2014).

Материал поступил в редакцию 16.05.14.

Сведения об авторе

МОКЕЕВ Владимир Викторович – доктор технических наук, зав. кафедрой информационных систем Южно-Уральского государственного университета (Национальный исследовательский университет), г. Челябинск.
e-mail: mokeyev@mail.ru

Т.В. Азарнова, П.В. Полухин

Расширение функциональных возможностей фаззинга веб-приложений на основе динамических сетей Байеса

Представлены теоретические и практические результаты по интеграции математического аппарата динамических байесовских сетей в процесс фаззинга межсайтового скриптинга современных веб-приложений. Практическая апробация построенных на основе сетей Байеса моделей, методов и алгоритмов показывает, что их применение позволяет значительно повысить эргономичность тестирования межсайтового скриптинга. Приводится построенная динамическая байесовская сеть процесса фаззинга межсайтового скриптинга и описывается адаптация основных инструментальных средств байесовских сетей применительно к процессу фаззинга.

Ключевые слова: фаззинг, межсайтовый скриптинг, веб-приложения, веб-браузер, динамическая байесовская сеть, приближенные алгоритмы вероятностного вывода, фильтрации и прогнозирования для динамических байесовских сетей

ВВЕДЕНИЕ

Современные веб-приложения являются эффективными инновационными технологиями с широкими функциональными возможностями и проработанной системой защиты. Но, тем не менее, проблема обнаружения уязвимостей приложений не только не решена, но и динамично развивается вместе с развитием их функциональных возможностей. Десятки новых технологий и программных платформ позволяют значительно упростить механизмы создания веб-приложений, разработку программного обеспечения, функционирующего в Интернете, создаются небольшие компании разработчиков. Одновременно с развитием механизмов создания веб-приложений, должны развиваться эффективные эргономичные инструменты тестирования приложений на предмет наличия уязвимостей.

Одним из эффективных современных механизмов выявления и локализации возможных уязвимостей является фаззинг. Сущность фаззинга заключается в подаче на вход исследуемого объекта заведомо некорректных данных с целью наступления события сбоя или ошибки [7]. Фаззинг веб-приложений способствует выявлению уязвимостей в веб-приложениях и компонентах инфраструктуры (сервер приложений, веб-сервер, сервер баз данных и др.), в которые встроено приложение, а также клиентских приложений, применяемых пользователями для интерактивного взаимодействия с элементами интерфейса веб-приложения (веб-браузеры, java-апплеты, flash и др.). В научном мире фаззинг имеет много общего с анализом граничных значений (Boundary Value Analysis – BVA), который характеризуется диапазоном конкретных допустимых значений входных данных и разрешает

убедиться, насколько эффективно метод исключений позволяет отфильтровать нежелательные данные из всего диапазона входных значений. Однако фаззинг в некотором случае расширяет BVA, так как концентрируется не только на граничных, но и на промежуточных значениях, что дает возможность оценить весь диапазон наборов входных данных, которые могут вызвать непредсказуемое и небезопасное поведение.

Моделирование процесса фаззинга в виде байесовской сети и проведение на базе данной модели процедур вероятностного вывода, фильтрации и прогнозирования позволяет структурировать информацию фаззинга и проводить адаптивное регулирование алгоритмов дальнейшего тестирования на основе структурированной информации. Применение подобного инструментария обучения и регулирования направлено на повышение результативности и эффективности процесса фаззинга, а также на снижение временных и ресурсных затрат на его адаптацию к тестированию разнородных веб-приложений. В настоящей статье рассматривается применение описанных инструментальных средств для фаззинга межсайтового скриптинга веб-приложений.

КЛАССИФИКАЦИЯ ОСНОВНЫХ ВИДОВ ФАЗЗИНГА МЕЖСАЙТОВОГО СКРИПТИНГА ВЕБ-ПРИЛОЖЕНИЙ

Стратегия анализа, классификации OWASP (международной организации обеспечения безопасности веб-приложений) дают четкую градацию и классификацию уязвимостей, опираясь на масштабность их распространения и степень важности. По оценкам OWASP достаточно большие масштабы распространения и высокую степень важности для веб-

приложений имеет межсайтовый скриптинг (XSS) – выполнение недокументированного (вредоносного) кода внутри веб-браузера. XSS-уязвимости возникают в ситуации, когда веб-приложение получает данные извне и выводит их без фильтрации и экранирования (например – поиск). XSS может открыть возможность исполнять скрипты внутри веб-браузера пользователя, позволяя при этом перехватить его идентификатор сессии, изменить контент веб-страницы и перенаправить пользователей на вредоносные ресурсы с целью их заражения. Веб-браузер пользователя выполняет синтаксический анализ и компиляцию скриптов без анализа их функциональной составляющей, поэтому вывод любой динамической информации на веб-страницу может быть потенциально небезопасным и подверженным XSS-уязвимостям [1]. Успешная реализация такой атаки позволит внедрять коды скриптов и ссылок на веб-страницу, отображаемую в браузере пользователя. Стоит отметить, что злоумышленник может использовать различные механизмы кодирования вредоносного кода, например Unicode, чтобы запрос выглядел как можно менее подозрительным для пользователя.

На сегодняшний день существует множество сценариев проведения межсайтового скриптинга, при этом дело не ограничивается вставкой «< >» символов, поэтому применение механизмов фильтрации только с таким символом не приведет к положительному успеху и не сможет в полной мере защитить приложение от атаки с применением XSS-уязвимости. Как правило, XSS использует JavaScript для проведения деструктивного воздействия, хотя сама природа межсайтового скриптинга не ограничивается лишь этим скриптовым языком. Вследствие чего, XSS может быть реализован с помощью любой клиентской технологии и языка программирования, которые выполняются в браузере, в том числе ActiveX(OLE), Flash, Silverlight, Java и т.д.

Современные особенности тестирования XSS-уязвимостей, показывают, что сложность композиции межсайтового скриптинга напрямую зависит от тех компонентов веб-приложения, которые должны подвергнуться тестированию.

Межсайтовые скриптинги разделяются по направлениям вредоносного воздействия. К первому направлению относятся – хранимые XSS. Этот тип известен еще как HTML-инъекции. Они используют в качестве контейнера любое хранилище данных (СУБД, файловую систему, кэш). При запросе пользователем веб-страницы вредоносный код извлекается и вставляется в HTML-разметку документа и затем выполняется в браузере пользователя. Реализация такого типа межсайтового скриптинга может повлиять на безопасность работы каждого пользователя веб-приложения. Хранимым XSS-уязвимостям наиболее подвержены различные форумы и чаты, так как позволяют использовать теги форматирования внутри передаваемых сообщений. Такие XSS принято также называть устойчивыми (persistent), так как они будут выполняться каждый раз при осуществлении доступа к веб-ресурсу [2, 5].

Ко второму направлению относятся отраженные XSS, которые используют альтернативные способы

доставки вредоносного контента пользователю (сообщения электронной почты, сторонние веб-ресурсы). Такая разновидность межсайтового скриптинга является одной из наиболее распространенных проблем безопасности современных веб-приложений. Реализация отраженных XSS, в первую очередь, нацелена на эксплуатацию уязвимостей веб-приложений, в которых данные должны заполняться пользователем и отправляться на сервер для обработки и дальнейшего получения результатов, затем возвращаться в браузер конечного пользователя для отображения результатов запроса или действия. Под успешной эксплуатацией отраженного XSS будем понимать процесс успешной отправки кода XSS в веб-приложение с возможностью его дальнейшего отображения внутри HTML-страницы. При этом, если код не кодируется в HTML-сущности, то XSS-код будет успешно интерпретирован и исполнен браузером пользователя. Наиболее характерный способ эксплуатации межсайтового скриптинга – внедрение вредоносной ссылки (кнопки) на страницу, которая будет подгружена другому пользователю. Когда пользователь нажимает на кнопку или ссылку, то уязвимое веб-приложение подгружает вредоносное содержимое в браузер пользователя. Такие XSS принято называть неустойчивыми (non-persistent), так как они существуют лишь на сеанс работы пользователя и веб-приложения.

К третьему направлению относятся DOM XSS. Они используют объектную модель документа (DOM-модель) веб-страницы в качестве контейнера XSS. DOM является спецификацией, утвержденной World Wide Web Consortium (W3C), которая определяет объектную модель XML, HTML-структуры. DOM XSS появляется не в результате ошибок серверного кода, а как следствие некорректной обработки пользовательских данных из кода JavaScript. Эффективность этого типа XSS-атаки через DOM возможны благодаря недостаточной обработке на уровне JavaScript таких объектов как: document.URL, document.location, document.referrer и некоторых других. Принципиальным отличием этого типа XSS-уязвимостей является тот факт, что данные не встраиваются в HTML-код и не требуют получения полезной нагрузки с веб-сервера, а лишь интерпретируются DOM-моделью веб-страницы.

Детальный анализ приведенных направлений показывает, что XSS-уязвимости представляют собой динамически развивающийся вид уязвимостей, распространению которых способствует стремительный рост интерактивности веб-приложений и повышение функциональности веб-браузеров. Развитие уязвимостей стимулирует адаптацию и развитие методологии фазинга межсайтового скриптинга веб-приложений. Следует отметить, что под фазингом в научном исследовании нами будет пониматься метод обнаружения уязвимостей в программном обеспечении, заключающийся в подаче на вход исследуемого объекта заведомо некорректных данных, чтобы вызывать события сбоя или ошибки. Из сформулированного определения видно, что количество тестовых данных, используемых для целевого объекта, может быть бесконечным и определяется конкретным разработчи-

ком инструмента на основе фаззинга. Фаззинг в той или иной степени позволяет спрогнозировать наличие ошибки, проанализировав, какие именно входные данные их могут вызвать.

В фаззинге, на этапе генерации тестовых данных, применяются два типа методов порождающий и мутационный, каждый из которых имеет как положительные, так и отрицательные стороны. Порождающий метод, реализуется путем генерации случайных независимых тестов. Практические исследования показывают, что этот метод обладает высокой производительностью, но недостаточно эффективен, особенно для тестирования сложных приложений. В методе возникает много проблем с распознаванием и восстановлением входных данных, которые привели к выявлению уязвимостей; для решения этих проблем исследователями был предложен мутационный метод.

Мутационный метод отличается от порождающего тем, что позволяет изменять и приспосабливать входные данные с учетом специфики функционирования целевого приложения, метод более сложен в реализации, так как требует от разработчика некоторых предварительных исследований, например, понимания структуры протокола (создает грамматику, описывающую работы спецификации протокола). Сложность реализации алгоритма порождения мутированных данных неизбежно приводит к увеличению времени, отведенного на тестирование, расходу ресурсов системы и процессорного времени. Мутационный метод позволяет выявить аномалии – ранее недокументированные уязвимости, что является очень актуальным особенно на этапе выпуска приложения.

Современные приложения являются сложными и интерактивными, что усложняет применение существующих методов фаззинга. Одно из возможных решений проблемы – комбинирование описанных выше двух методов в единый гибридный, способный повысить уровень обнаружения общих уязвимостей и аномалий [2]. Предложенный в настоящей работе математический инструментальный процесс фаззинга XSS-уязвимостей веб-приложений, базирующийся на теории динамических байесовских сетей, реализован в рамках описанного направления исследований.

ДИНАМИЧЕСКАЯ БАЙЕСОВСКАЯ СЕТЬ ФАЗЗИНГА МЕЖСАЙТОВОГО СКРИПТИНГА ВЕБ-ПРИЛОЖЕНИЙ

Для моделирования процесса фаззинга XSS-уязвимостей веб-приложений используются динамические байесовские сети, но, вначале кратко остановимся на описании статических байесовских сетей, из которых, как из временных срезов, состоит динамическая байесовская сеть. Байесовские сети представляют собой модели вероятностных и причинно-следственных отношений между переменными в статистическом информационном моделировании. В байесовских сетях могут сочетаться эмпирические частоты появления различных значений переменных, субъективные оценки «ожиданий» и теоретические представления о математических вероятностях тех или иных следствий из априорной информации [8]. Это является важным практическим преимуществом и отличительной особенностью методики байесов-

ских сетей от других методик информационного моделирования.

Статическая байесовская сеть представляет собой ориентированный граф без циклов, вершинами которого являются дискретные и (или) непрерывные случайные величины. Каждый узел графа представляет собой случайную переменную x_i , $i = \overline{1, n}$, отражающую признак или атрибут исследуемого объекта (системы, процесса), а ребра (дуги) – зависимости между ними. Если ребро направлено от вершины y к вершине x , то вершина y является родительской по отношению к x , а x – дочерней для y . Множество родительских вершин для вершины x обозначается $Parents(x)$. Принято, что в графе байесовской сети родительские вершины располагаются на более высоком уровне, чем дочерние. Граф байесовской сети описывает ограничения зависимости переменных друг от друга; считается, что после задания родительских вершин, каждая вершина становится условно независимой от вершин, лежащих на более высоких уровнях. Его структура и условные распределения узлов однозначно описывают совместное распределение всех переменных, что позволяет решать задачу классификации как определение значения переменной класса, для которого условная вероятность этой задачи при заданных значениях признаков будет максимальной. Каждая вершина x_i , $i = \overline{1, n}$ характеризуется распределением условных вероятностей $P(x_i | Parents(x_i))$, которое определяется влиянием родительских вершин на дочерние, при этом каждое распределение выражается в виде таблицы условных вероятностей – сокращенно СРТ (Conditional Probability Table). Любая рассматриваемая строка в таблице СРТ содержит условную вероятность каждого значения вершины для каждого условия выборки, которое определяет условную вероятность. При этом под условием выборки понимается одна из возможных комбинаций значений родительских вершин по отношению к наблюдаемой вершине. Каждая строка в сумме должна давать единицу, поскольку элементы строки представляют собой исчерпывающее число исходов данной переменной. Важно отметить, что если переменные имеют булевское значение, то после определения вероятности истинного значения P , вероятность ложного будет равна $1-P$, поэтому второе значение в таблице СРТ можно не указывать.

Структура графа и условные распределения узлов однозначно описывают совместное распределение всех переменных:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | Parents(x_i)). \quad (1)$$

Таким образом, совместное распределение является корректным отображением рассматриваемой области, но только при условии, что вершина в этой области является условно независимой от предшествующих в конкретном наборе вершин, после того как определено множество всех родителей. Поэтому, учитывая данное свойство для составления сети Байеса с правильной структурой, необходимо для конкретной проблемной области выбрать для каждой

вершины родительские вершины. Логически следует, что множество родительских вершин должно включать вершины из множества X_1, X_2, \dots, X_{i-1} , которые напрямую оказывают воздействие на i -ю вершину.

Используя совместное распределение, можно находить значения ненаблюдаемых переменных, максимизирующие их условные вероятности при фиксированных значениях наблюдаемых переменных. Байесовские сети позволяют решать задачи вероятностного вывода – задачи вычисления распределения апостериорных вероятностей $P(X|E)$ для множества переменных запроса $X = \{X_1, X_2, \dots, X_n\}$, если дано некоторое событие e , при котором присвоены определенные значения множеству переменных свидетельства $E = \{E_1, E_2, \dots, E_m\}$, $E_1 = e_1, E_2 = e_2, \dots, E_m = e_m$. При этом свидетельства логически можно разделить на две категории: прямое свидетельство для вершины, состояние которой зафиксировано и неизменно в соответствии с произошедшим событием, и косвенное свидетельство, которое позволяет осуществлять процесс изменения предыдущих вероятностных характеристик вершины. Множество всех вершин байесовской сети считается объединением трех множеств вершин $X \cup E \cup Y$, где $Y = \{Y_1, Y_2, \dots, Y_p\}$ – множество скрытых переменных. Апостериорное распределение вероятностей $P(X|E)$ вычисляется по формуле

$$P(X|E) = \alpha \sum_Y P(X, E, Y) \quad (\alpha - \text{нормирующий множитель}).$$

Для вычисления $P(X|E)$ разработан целый ряд точных алгоритмов, к которым относятся алгоритмы перебора, устранения переменной и кластеризации. Однако для байесовских сетей со сложной структурой и большим числом узлов точные алгоритмы вероятностного вывода трудоемки и требуют значительных временных затрат. Исходя из этого, на практике целесообразно использовать приближенные алгоритмы логического вывода. Наиболее распространенными алгоритмами приближенного логического вывода являются алгоритмы Монте-Карло, обеспечивающие получение приближенных результатов, точность которых пропорциональна количеству сформированных выборок. Характерный алгоритм из группы алгоритмов Монте-Карло – метод непосредственной выборки с исключением. Метод генерируется на основе априорного распределения вероятностей, которое определяется исходя из структуры рассматриваемой сети, а затем исключаются те выборки, которые не удовлетворяют свидетельству $E = \{E_1, E_2, \dots, E_m\}$. Для каждой переменной сети формируется выборка, причем распределения вероятностей, на основе которых формируется выборка, напрямую зависят от вероятностных значений родительских переменных $\text{Parents}(X_i)$. Обозначим: N_E – общее число выборок для свидетельства E , $N_E(X)$ – частота конкретного события для переменных запроса $X = \{X_1, X_2, \dots, X_n\}$. С учетом изложен-

ного, для обозначения каждого частично заданного события X может быть получена согласованная оценка вероятности $P(X|E) = \frac{N_E(X)}{N_E}$, это означает,

что вероятность события может быть оценена за счет вычисления отношения количества выборок частично заданного события к общему числу всех событий [6, 7]. Что касается набора выборок, согласованных со свидетельством, то их значение экспоненциально уменьшается по мере увеличения переменных свидетельства – главным недостатком данного алгоритма. В качестве второго алгоритма рассмотрим алгоритм оценки веса с учетом правдоподобия, который позволяет решить проблемы, связанные с алгоритмом непосредственной выборки, за счет отбрасывания событий, несогласующихся со свидетельством, при этом значения переменных свидетельства $E = \{E_1, E_2, \dots, E_m\}$ фиксируются и формируются выборки только для переменных $X = \{X_1, \dots, X_n\}$ и $Y = \{Y_1, \dots, Y_p\}$. Обозначим $N(X, Y, E)$ – количество полученных выборок.

Правдоподобие вычисляется перемножением условных вероятностей для каждой переменной свидетельства, если заданы ее родительские переменные:

$$w(X, Y, E) = \prod_{i=1}^m P(E_i | \text{Parents}(E_i)). \quad (2)$$

Не все события равноправны, поэтому при вычислении результата каждое событие взвешивается с учетом правдоподобия того, насколько оно согласовано со свидетельством.

Результирующая вероятность вычисляется следующим образом:

$$P(X|E) = \alpha \sum_Y N(X, Y, E) w(X, Y, E). \quad (3)$$

Этот алгоритм является более эффективным по сравнению с алгоритмом непосредственной выборки, так как позволяет использовать все сгенерированные выборки, однако рост переменных свидетельств может приводить к постепенному снижению производительности.

В качестве третьего алгоритма рассмотрим алгоритм Монте-Карло с применением цепи Маркова. Его отличительная особенность заключается в том, что каждое новое событие на основе случайного изменения формируется в предшествующее событие. Выборка складывается в соответствии с Марковским покрытием переменной X_i , которое состоит из ее родительских переменных, дочерних переменных и родительских переменных для дочерних переменных.

Перейдем к описанию динамических сетей Байеса, которые представляют собой последовательность байесовских сетей [5, 6]. Каждый элемент последовательности (временной срез) описывает состояние байесовской сети на определенный момент времени, начиная с момента времени $t=0$. Обозначим каждый момент времени t через X_t, E_t , соответственно множество ненаблюдаемых (состояний) и наблюдаемых (свидетельств) переменных. Каждый временной срез дина-

мической байесовской сети может иметь любое количество переменных состояния X_t и переменных свидетельства E_t . За начало отчета появления переменных свидетельств будем брать момент времени $t=1$. Символами $X_{t_1:t_2}, E_{t_1:t_2}$ обозначим последовательности переменных на временных срезах с момента времени t_1 до момента времени t_2 . Динамическая байесовская сеть задается таблицей условных вероятностей для каждой переменной временного среза, для нулевого момента времени $P(X_0)$, и не изменяется при переходе от среза к срезу.

При анализе динамических байесовских сетей возникает ряд ограничений, обусловленных тем, что каждый узел сети может быть связан с бесконечным числом узлов предыдущих временных срезов. Для преодоления этих ограничений будем использовать предположение о марковости, ограничивающее зависимость текущего состояния от конечного числа предыдущих состояний. Предполагается, что текущее состояние динамической сети Байеса зависит непосредственно от предыдущего состояния, это утверждение характерно для Марковских процессов первого порядка $P(X_t|X_{0:t-1}) = P(X_t|X_{t-1})$. Вероятностное распределение $P(X_t|X_{t-1})$ называется моделью перехода, а вероятностное распределение $P(E_t|X_t)$ – моделью восприятия [8].

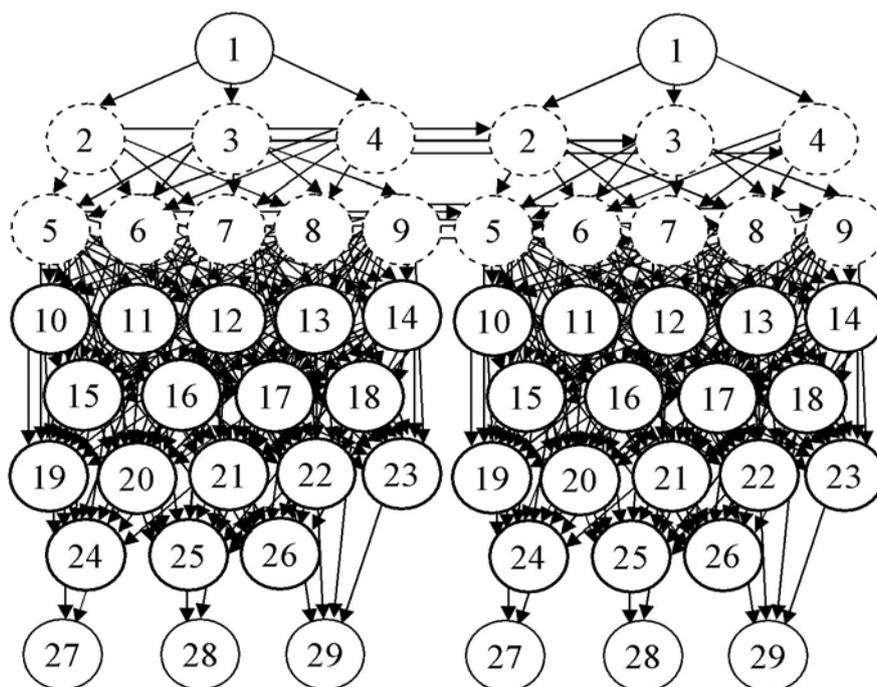
С учетом введенных предположений и вероятностных распределений полное совместное распределение вероятностей задается следующим образом:

$$P(X_0, X_1, \dots, X_t, E_1, \dots, E_t) = P(X_0) \prod_{i=1}^t P(X_i|X_{i-1}) P(E_i|X_i) \quad (4)$$

На рисунке приведен фрагмент (два временных среза) динамической байесовской сети процесса фазинга межсайтового скриптинга. Узлы в первом временном срезе представленной динамической байесовской сети не имеют никаких связанных параметров, однако каждый узел второго среза имеет соответствующую условную вероятность, определяемую как $P(X_t|X_{t-1})$, где $t > 1$. Если родительские узлы присутствуют как в первом, так и во втором временном срезе, то сеть представляет собой Марковский процесс первого порядка. Стоит отметить, что если существует дуга из узла X_{t-1} в X_t , то такой узел называется постоянным, а если дуга существует только в момент временного среза, то она является временной.

Таблица содержит список обозначений для данной байесовской сети.

Каждый временной срез, приведенный на рисунке, интерпретируется как эксперимент. В нашем исследовании будут рассматриваться два подхода к понятию эксперимент.



Фрагмент динамической байесовской сети процесса фазинга межсайтового скриптинга (XSS)

Характеристика узлов динамической байесовской сети

Узел	Характеристика
1	Определение программных компонентов на стороне клиента.
2, 3, 4	XSS:отраженные, хранимые, DOM
5, 6, 7, 8, 9	Наиболее характерные компоненты веб-страницы, подверженные XSS
10, 11, 12, 13, 14	Механизмы обходы алгоритмов фильтрации, как веб-приложения, так и веб-браузера пользователя
15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26	Уязвимости, связанные с XSS: обход NAT, чтение критичных данных и выполнения запросов от имени пользователя, реализация кейлоггера, фишинг, анализ DNS серверов, анализ локальной сети, выявление уязвимых компонентов браузера, DoS и DDoS, используя браузер пользователя, заражение вирусами
27, 28, 29	Механизмы контроля доступа, конфиденциальности и целостности

В рамках первого подхода под экспериментом будем понимать этап исследования тестирования веб-приложения как уникального приложения, имеющего единичный (неповторяющийся) набор программных компонентов. На каждом временном срезе тестирования определяются вероятностные характеристики веб-приложения – наблюдаемые переменные и переменные свидетельства, накапливается информация о наличии уязвимостей. Вероятностное распределение для нулевого момента времени $P(X_0)$, модель перехода и модель восприятия строятся на основе статистики по этапам тестирования веб-приложений. При этом подразумевается, что модель перехода и модель восприятия являются стационарными (одинаковыми для любого момента времени t).

Второй подход предполагает, что эксперимент – это процедура полного тестирования одного приложения из группы однородных веб-приложений (имеющих сходный набор программных компонентов). Для группирования веб-приложений в однородные классы мы предлагаем специальную многокритериальную оценку, учитывающую функциональный набор используемых в приложении веб-технологий и программных компонентов.

На построенной динамической байесовской сети нами решаются следующие задачи вероятностного вывода.

Задача фильтрации – вычисление апостериорных вероятностей $P(X_t|E_{1:t})$ переменных текущего состояния X_t при условии наличия всех свидетельств $E_{1:t}$. Логически решение задачи фильтрации можно разделить на два этапа:

1) вычисление распределения вероятности из t в $t+1$, затем оно обновляется на основе получения ново-

го свидетельства E_{t+1} . Вычисление осуществляется по следующему правилу:

$$\begin{aligned}
 P(X_{t+1}|E_{1:t+1}) &= P(X_{t+1}|E_{1:t}, E_{t+1}) = \\
 &= \alpha P(E_{t+1}|X_{t+1}) \sum_{X_t} P(X_{t+1}|X_t) P(X_t|E_{1:t}), \quad (5)
 \end{aligned}$$

где α представляет собой нормализованную константу, применяемую для того, чтобы значение вероятности в сумме было равно 1; $(X_{t+1}|X_t)$ – модель перехода; $P(X_t|E_{1:t})$ – одношаговое предсказание, обновление которого осуществляется за счет получения нового свидетельства. $P(X_{t+1}|E_{1:t+1})$ можно также получить непосредственно из модели восприятия.

В рамках первого подхода, используя механизмы фильтрации, можно оценить вероятность наличия XSS-уязвимостей веб-приложений на основе свидетельств, полученных из предыдущих тестирований;

2) вычисление распределения вероятности наличия уязвимостей следующего тестирования веб-приложения (при наличии свидетельств текущего тестирования и свидетельств тестирования однородных приложений в предшествующие временные срезы).

Задача предсказания – вычисление распределения апостериорных вероятностей $P(X_{t+k}|E_{1:t})$ значений переменных в будущем состоянии, если даны все свидетельства, полученные к данному моменту времени. Исходя из этого, задача предсказания может быть представлена в виде задачи фильтрации (при отсутствии получения новых свидетельств). Учитывая, что фильтрация уже включает в себя одношаговое предсказание, задача предсказания решается рекурсивно:

$$P(X_{t+k+1}|E_{1:t}) = \sum_{X_{t+k}} P(X_{t+k+1}|X_{t+k}) P(X_{t+k}|E_{1:t}). \quad (6)$$

Прогнозирование позволяет принять решение о присутствии уязвимости при тестировании последующих приложений из однородной группы.

Задача сглаживания – вычисление апостериорных вероятностей значений переменных $P(X_k|E_{1:t})$ прошлого состояния, если даны свидетельства вплоть до нынешнего:

$$P(X_k|E_{1:t}) = \alpha P(X_k|E_{1:k})P(E_{k+1:t}|X_k),$$

$$P(E_{k+1:t}|X_k) = \sum_{X_{k+1}} P(E_{k+1}|X_{k+1})P(E_{k+2:t}|X_{k+1})P(X_{k+1}|X_k). \quad (7)$$

Эта задача позволяет нам провести оценку присутствия уязвимостей состояния в прошлом за счет получения всех значений переменных свидетельств до текущего момента времени. Для динамической байесовской сети фаззинга межсайтового скриптинга задача сглаживания сводится к определению вероятности того, что можно будет получить аутентификационные данные пользователя, если есть результаты выполнения всех механизмов проведения межсайтового скриптинга до текущего момента времени. Такой подход получил название сглаживание с фиксированной задержкой, однако, в случае автономного состояния будет рационально называть его – сглаживание с фиксированным интервалом.

Решение изложенных задач вероятностного вывода позволяет оценить, какие конкретно переходы между состояниями имели место, и какое именно состояние привело к той или иной модели восприятия, фактические оценки могут быть использованы для обновления моделей восприятия и переходов. При этом рекурсивное обновление моделей происходит до тех пор, пока все вычисления не сойдутся.

Для решения задач фильтрации, прогнозирования и сглаживания в рамках нашего исследования используются алгоритмы приближенного вероятностного вывода на основе подхода взвешивания с учетом правдоподобия. Для адаптации математического аппарата алгоритма взвешивания с учетом правдоподобия к описанной динамической байесовской сети осуществляется ее развертывание. В методе взвешивания с учетом правдоподобия порождаются выборки переменных состояния, которые взвешиваются с учетом соответствия их переменным свидетельствам, находящимся под наблюдением. Очередная выборка формируется изолированно (независимо), итеративно проходят переменные состояния развернутой сети. Большинство выборок имеют низкие веса, вследствие чего в составе взвешенной оценки преобладает незначительная доля выборок, согласующихся со свидетельством, поэтому при увеличении числа переменных свидетельств в развернутой сети возрастает пространственная и временная сложность. При этом, как и алгоритмы точного вероятностного вывода, алгоритм взвешивания с учетом правдоподобия может быть применен лишь к развернутой динамической байесовской сети. Для решения этих проблем мы предлагаем использовать алгоритм фильтрации частиц, который позволяет на основе наблюдений отбрасывать выборки с наименьшим весом, одновременно увеличивая число выборок с наибольшим ве-

сом. В алгоритме фильтрации частиц в начальный момент времени из вероятностного распределения $P(X_0)$ формируется N выборок. На последующих шагах, при переходе от временного среза t к временному срезу $t+1$, для обновления множества выборок берется текущее значение X_t и строится модель перехода $P(X_{t+1}|X_t)$. Общее число выборок, достигающих X_{t+1} из X_t , определяется следующим выражением: $N(X_{t+1}|E_{1:t}) = \sum_{X_t} P(X_{t+1}|X_t)N(X_t|E_{1:t})$,

где $N(X_t|E_{1:t})$ – количество выборок для состояния X_t после получения свидетельств $E_{1:t}$. Обобщенно алгоритм фильтрации частиц можно представить следующим образом:

Particle_Filtering (e,N,dbn)

inputs: e – полученное свидетельство, – количество выборок, которые должны сопровождаться, dbn – динамическая сеть Байеса с распределением априорных вероятностей, моделью перехода и моделью восприятия $P(E_1|X_1)$

static: – вектор выборок размерностью, и из
local variable: W – вектор весов размерностью N
for i = 1 to N do

S[i] ←— выборка из

S[i] ←— $P(e|X_0 = S[i]), P(X_1|X_0 = S[i])$

S[i] ←— weighted – sample – with – replace(N,S,W)

return S

Каждая выборка взвешивается с учетом правдоподобия, переменной свидетельства присваивается вес $P(E_{t+1}|X_{t+1})$, суммарный вес выборок равен $w(X_{t+1}|E_{1:t+1}) = P(E_{t+1}|X_{t+1})P(X_{t+1}|E_{1:t})$. Выборки, которые имеют малый вес, отбрасываются. Из новой популяции выборок формируется N выборок, имеющих наибольшие веса.

ЗАКЛЮЧЕНИЕ

Технология фаззинга – это эффективный, апробированный инструмент поиска XSS-уязвимостей. Предложенный в нашей статье метод моделирования процесса фаззинга межсайтового скриптинга на основе динамических сетей Байеса реализует системный подход к анализу и локализации уязвимостей рассматриваемого типа. Решение задач вероятностного вывода (фильтрации, прогнозирования, предсказания) позволяет получить оценку апостериорных вероятностей случайных переменных, связанных с фактом присутствия уязвимостей межсайтового скриптинга. Рассмотренные модели, методы и алгоритмы можно применять как инструментальные средства мутационного фаззинга, которые существенно расширяют его возможности, способствуют своевременной локализации уязвимостей, блокированию потенциального вредоносного воздействия, накоплению статистической информации о распро-

страненных векторах атак для последующего тестирования веб-приложений со сходным набором программных компонентов, а также открытию новых векторов вредоносного воздействия. В рамках нашего исследования разработано оригинальное программное обеспечение (фаззер), проведены его апробация и сравнение как качественных, так и функциональных показателей с уже существующими программными продуктами, предназначенными для тестирования веб-приложений. Результаты практического применения программного инструмента подтверждают эффективность предложенных алгоритмов, расширяют диапазон выявления уязвимостей межсайтового скриптинга и позволяют снизить временные и ресурсные затраты на проведения тестирования.

СПИСОК ЛИТЕРАТУРЫ

1. Godefroid P., Levin M. Automated whitebox fuzz testing. – USA: NDSS, 2008. – 350 p.
2. Godefroid P., de Halleux Peli. Automating Software Testing Using Program Analysis // IEEE SOFTWARE. – 2008. – №5 – P. 30 – 37.
3. Korb A., Nicholson A. Bayesian Artificial Intelligence. – UK: Chapman & Hall CRC Press, 2004. – 244 p.
4. Murphy K. Machine Learning: A Probabilistic Perspective. – USA: MIT Press, 2012. – 1067 p.
5. Zalewski M. The Tangled Web. A Guide to Securing Modern Web Applications. – USA: No Starch Press, 2012. – 477 p.

6. Азарнова Т., Полухин П. Применение динамических байесовских сетей для повышения эффективности процесса фаззинга SQL-инъекций веб-приложений // Системы управления и информационные технологии. – 2014. – №1.1(55). – С. 106-112.
7. Бабинов В. Байесовские сети доверия и некоторые аспекты идентификации систем. – СПб.: УТЭОСС-2012. – URL: <http://uteoss2012.ipu.ru/procdngs/0353.pdf>
8. Тулупьев А., Николенко С., Сироткин А. Байесовские сети, логико-вероятностный подход. – СПб.: Наука, 2006. – 728 с.

Материал поступил в редакцию 25.05.14.

Сведения об авторах

АЗАРНОВА Татьяна Васильевна – доктор технических наук, профессор кафедры математических методов исследования операций, факультета прикладной математики и механики Воронежского государственного университета
e-mail: ivdas92@mail.ru

ПОЛУХИН Павел Валерьевич – аспирант кафедры математических методов исследования операций, факультета прикладной математики и механики Воронежского государственного университета
e-mail: alfa_force@mail.ru

С. Л. Мишланова, Н. И. Береснева, Я. В. Мишланов

Измерение информации: количественный и качественный аспекты*

Рассматривается поиск способов определения понятий «информация» и «количество информации». Представлены основные подходы, позволяющие определить количество информации. Особое внимание уделяется противоречию между содержанием понятия "информация" в обыденном сознании – как содержание, независимо от формы, и определением «количество информации» для нужд вычислительной и телекоммуникационной техники, когда учитывается преимущественно форма сообщения, независимо от содержания. Выявляется философская тенденция движения знания в сторону содержательного компонента.

Ключевые слова: информация, количество информации, искусственный интеллект, форма и содержание, информационное и идеальное

Картина постижения проблемы измерения информации демонстрирует общий философско-семиотический принцип: исходный пункт познания – явления, несущие в себе сущность. Изучая явление, мы проникаем в сущность, проявляющуюся в нем, а исходя из знания сущности, познание снова движется к объяснению многообразия явлений. Если предмет познания выступает система взаимодействующих объектов, когда предмет или процесс не доступен непосредственному восприятию, то выделить существенные свойства непросто. Взаимодействуя друг с другом, объекты "порождают" множество явлений, в каждом из которых проявляется одна из сторон сущности. Для вскрытия реальной сущности сложной системы приходится изучить как можно больше ее проявлений.

Переход развитых стран в постиндустриальную стадию, возникновение глобального коммуникационного пространства [1] с новой силой поставили вопрос об определении и способах измерения информации. Необходимость четкого определения информации и способа выявления ее количества обусловлена тем, что общее количество информации, которым оперирует человечество, приближается к зеттабиту [2].

Современное состояние понимания термина *информация* в научном сознании А. Соколов охарактеризовал как «невообразимый хаос» [3]. К. Саган, возражая против переоценки информационных критериев подхода к проблеме, поставил вопрос: «сколько *бит* информации содержится в соотношении Эйнштейна $E=mc^2$?» Непосредственная информационная оценка этого соотношения несопоставима с действительным значением [4].

В настоящей статье мы рассматриваем трансформации понятия *информация*, связанные с развитием техники. Значительные успехи в этом направлении достигнуты в разработках по созданию систем искусственного интеллекта [5], однако дальнейшее развитие информационных технологий ограничено используемыми подходами. Представим основные взгляды на построение систем искусственного интеллекта и их критику.

Термин *информация* произошел от латинского глагола «*informato*» – придавать вид, форму; обучать, воспитывать; строить, составлять; мыслить, воображать [6]. К XX в. значение понятия трансформировалось в «осведомленность», «сведения», «сообщение» [3], а в настоящее время толкуется в зависимости от сферы употребления. В словаре С. Ожегова зафиксировано определение: «1. Сведения об окружающем мире и протекающих в нем процессах, воспринимаемые человеком или специальным устройством. 2. Сообщения, осведомляющие о положении дел, о состоянии чего-н.» [7]. Юридическая точка зрения на понятие отражена в Федеральном законе РФ от 27.07.2006 г.: «Информация – сведения ... независимо от формы их представления» [8]. Если на основе представленных определений рассмотреть соотношение формы и содержания понятия, то получится, что содержание сведений является более важным по отношению к форме.

Математическая теория связи. В XIX и начале XX вв. развиваются средства связи: телеграф, телефон, радио и телевидение. Проблема определения количества информации заключалась, в основном, в обеспечении надежности передачи сообщений по каналу связи. В работах Г. Найквиста, Р. Хартли и К. Шеннона [9–12] были заложены основы математической теории информации, имеющей дело с сообщениями, которые рассматриваются как последовательность символов, отвлеченно от значения: совпадение

* Исследование выполнялось при финансовой поддержке Российского гуманитарного научного фонда (проекты РГНФ-цр № 14-13-59007, РГНФ-цр №14-16-59007, № 14-06-31143 мол_а)

исходного сообщения и сообщения, прошедшего через канал связи, формальное, с точностью до символа, означало надежность канала. Таким образом, предложенный этими исследователями способ определения количества информации не соответствует ее обыденному пониманию.

Комбинаторный, вероятностный и алгоритмический подходы к определению количества информации. А. Колмогоров выделил три подхода к определению количества информации: комбинаторный, вероятностный и алгоритмический [13]. Подход, предложенный К. Шенноном, он назвал вероятностным и выделил комбинаторный подход, несмотря на его сходство с вероятностным.

В 1960-х гг. А. Колмогоров и, независимо от него, Р. Саломонофф, предложили алгоритмический подход. В нем предлагается взять за основу количество информации в объекте x относительно объекта y как более содержательное понятие, чем просто информация, содержащаяся в объекте x . Количество информации, содержащееся в объекте x относительно объекта y предполагается равным длине самой короткой программы, которая позволяет получить объект x из объекта y .

Алгебраический подход, предложенный В. Гоппой в конце 1980-х гг. [14], предполагает определение количества информации на основе применения теории групп и симметрии, позволяет учесть симметрию расположения букв в сообщении и, таким образом, упорядочить формально его структуру, при этом содержание не учитывается.

Теория интегрированной информации была разработана в 2000-х гг. Г. Тонони [15] для объяснения природы сознания, которое он понимал как «интегрированная информация» физической системы [15]. Основными понятиями теории являются система, обладающая состоянием и механизмом, информация и интеграция. Информация вычисляется как уменьшение неопределенности при переходе из предыдущего состояния в текущее. Для определения интеграции система разбивается на части с минимальной собственной информацией, связность этих частей друг с другом определяет интеграцию. Далее вводится понятие комплекса – части системы, при увеличении которой связность не увеличивается (в исследованиях по физиологии высшей нервной деятельности показано, что именно такие комплексы нейронов являются физиологической основой отдельных понятий (теория когнитива К.Анохина [16, 17])). При определении количества информации система рассматривается отвлеченно от содержания – как набор абстрактных структур. Применяя расчет количества информации к конкретной системе и интерпретируя вычисленное значение в соответствии с заданной системой, можно получить конкретные результаты. Количество информации не отражает природу системы.

Меры простоты. В 1950-х гг. Н. Гудмен ставит задачу оценить сложность знания [18], в котором выделяются логический и содержательный аспекты. Знания не различаются логически, сложность же определяется содержательно. В содержательном аспекте выделяется базис – набор предикатов, харак-

теризующихся симметричностью, транзитивностью, рефлексивностью. Сложность одноместных предикатов принимается за единицу. Сложность предикатов с большим числом мест увеличивается с ростом числа этих мест. Если предикат оказывается симметричным или рефлексивным, то сложность его считается меньшей [18].

Семантическая информация. Для семантической концепции определения информации наиболее важным оказывается анализ содержательных характеристик. При этом семантическая информация высказывания исключает некоторые альтернативы, выражаемые средствами языка: чем больше альтернатив исключает высказывание, тем более оно семантически информативно [9]. Однако проблемы определения мер информации не могут решаться на чисто логических основаниях. Я. Хинтикка в 1968 г. предложил решение проблемы: на основе различения "поверхностной" и "глубинной" информации, обращаясь друг в друга в процессе познания. Это позволило объяснить, каким образом логические и математические доказательства дают приращение поверхностной информации (знания), измерять его возможно даже в случаях частичного выполнения этих доказательств. Однако сказать то же самое о глубинной информации подход Хинтикки не позволил [9].

Ю. Шрейдер [9] определил зависимость информации в сообщении от степени развития тезауруса получателя (запаса знаний). Е. Вентцель проиллюстрировала эту зависимость примером: учебник по высшей математике несет нулевую информацию для ребенка, максимальную информацию извлечет из него студент. Но вычислить эту зависимость непротиворечивым образом не удалось.

Подходы к измерению ценности информации разрабатываются в теории принятия решений, теории игр и т.п. По А. Харкевичу [9], величину ценности информации можно измерить посредством приращения вероятности достижения цели после получения этой информации. Таким образом, вводится понятие субъективной вероятности.

Проблема информации в философии. Н. Винер определяет информацию как: «обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств» [19]. Также информация определяется как мера организации. В понятие *информация* у Винера входят как содержание, так и «мера организации», не зависящая от природы системы.

У. Эшби определяет информацию как разнообразие [20]. Разнообразие – это либо число различных элементов, либо логарифм этого числа, т. е. разнообразие зависит не от природы, а только от числа элементов.

С 1960-х гг. проблема определения понятия *информация* привлекла внимание отечественных философов. А. Урсул выдвинул такое определение: «Информация есть отраженное многообразие». Она включает в себя не все содержание отражения, а только аспект, связанный с многообразием. «Отраженное однообразие» – ничто. Таким образом, информация при таком определении совпадает с отражением [3].

Другого подхода к определению информации придерживались А.Берг, В.Глушков, А.Ершов, В.Сифоров, отождествлявшие информацию и ее организацию [3]. Другая крайность в определении информации проявляется при попытке практиков определить информацию как содержание сигнала или сообщения [3] – информация оказывается тождественной сигналу [15].

По В.Афанасьеву, информация "представляет собой знания, сообщения, сведения о социальной форме движения материи и о всех ее других формах в той мере, в какой они используются обществом, человеком, вовлечены в орбиту общественной жизни" (цит. по [3]) – она представляет собой социальную коммуникацию [3]. Информацию можно определить как формальную структуру, схваченную знанием [4], при этом она отличается от идеального, знания, отражения, организации – это формальная величина, не выражающая, как таковая, никаких качеств.

С позиций современных подходов к проблемам научной философии, реальный мир – последовательность бесконечного множества форм материи – ступеней единого закономерного мирового процесса. Качества, присущие более высоким уровням существования материи, не сводимы к более низким уровням. Поэтому сравнивать информацию, относящуюся к разным формам существования материи – некорректно. В работе [4] выдвигается гипотеза о том, что возможно ввести коэффициенты сложности для каждого уровня организации материи, которые позволили бы переводить бит информации одного уровня в бит информации другого.

Проблема информации в искусственном интеллекте – научном направлении, которое ставит и решает задачи аппаратного или программного моделирования видов человеческой деятельности, традиционно считающихся интеллектуальными [21]. В 1951 г. М. Минский и Д. Эдвардс создали первый нейросетевой компьютер [22]. 1950-1960-е гг. – период энтузиазма и больших ожиданий – А. Ньюэллом, Г. Саймоном создана программа Logic Theorist, доказавшая большинство теорем «Principia Mathematica» Б. Рассела и А. Уайтхеда, которые в 1975 г. выдвинули гипотезу, в которой утверждалось, что физическая символьная система имеет необходимые и достаточные средства для выполнения базовых интеллектуальных действий [23]. Г. Гелертнер создал программу Geometry Theorem Prover, доказывающую сложные теоремы. А. Самюэл разработал программу, способную играть в шашки лучше, чем ее создатель [22]. Были успешно решены задачи, относящиеся к частным предметным областям. Решение достигалось манипулированием абстрактными структурами, символами. Успех способствовал вере, что рассуждений на уровне символов достаточно для решения задач искусственного интеллекта.

В 1966-1973 гг. при стремлении решать задачи в более широких предметных областях проявились сложности. Во-первых, связанные с ростом объема необходимых знаний о предметной области. Так, в 1960-х гг. были попытки создать программы машинного перевода, считая, что для перевода достаточно учесть грамматику и произвести простую замену

слов [22]. А во-вторых, связанные с ростом количества возможных вариантов действий при расширении предметной области. Так, оптимизм, с которым встречали сообщения о доказательстве теорем с помощью метода резолюции, угас, когда исследователи не смогли доказать теоремы, содержащие более десятка фактов [22]. В области генетических алгоритмов также возникли сложности: считалось, что необходимо проверять случайные мутации и оставлять те, которые кажутся полезными, но "никаких признаков прогресса не было обнаружено" [18].

С 1969 по 1979 гг. развивались системы, основанные на накопленных знаниях. Успешной оказалась система Dendral, позволяющая определить молекулу по результатам масс-спектрометрии. Такие системы потребовали развития способов представления знаний. Так, М. Минским была предложена фреймовая модель [22].

С 1995 г. развивается подход, основанный на создании целостных интеллектуальных агентов, например, когнитивных архитектур Soar, ACT-R и Clarion [22, 24].

Перечисленные подходы к построению систем искусственного интеллекта опираются на манипулирование абстрактными структурами. Ограниченность этих подходов была предметом дискуссий. Дж. Серл в 1990 г. говорит об ограниченности искусственного интеллекта, основанного на оперировании символами, апеллируя к невозможности понимания естественного языка [25]. Однако в следующей статье в этом журнале (П. Черчланд) высказывается противоположный взгляд [26]. Проблема решается с точки зрения философии: «информационное описание действительности, несмотря на всю свою важность для науки и техники, имеет ограниченное значение и нуждается в качественной интерпретации, которая дается на основе внеинформационных факторов» [4, 27–30]. Представленные здесь подходы имеют ограниченную связь с миром, в то время как человек – потребитель информации, состоит с ним в сложных практических отношениях. Так, в системах ConceptNet 5 [5] и Watson [31] заложены знания из Википедии, словарей и тезауруса WordNet. Но дальнейшее развитие исследований ограничено: ConceptNet 5, свидетельствуют разработчики, испытывает трудности не в объеме знаний, а в области «здравого смысла».

Таким образом, рассмотренные основные подходы к определению понятий *информация* и *количество информации* показывают, что уже в первых теориях возникло противоречие между общеупотребительным понятием (где содержание сообщения важнее, чем форма) и значением термина в теории информации (где количество информации определялось формой, абстрагированной от содержания).

В области вычислительной и телекоммуникационной техники количество информации определяется безотносительно к содержанию сообщений, однако такое определение количества информации неполно. При сравнении количества информации, определенного для разных уровней организации материи, необходимо учитывать различия в уровнях – например, с помощью коэффициентов сложности. Задача опреде-

ления значений коэффициентов сложности не решена. Хотя можно сделать вывод об ограниченности методов искусственного интеллекта, основанных на оперировании абстрактными структурами.

СПИСОК ЛИТЕРАТУРЫ

1. Варакин Л. Глобальное информационное общество. – М.: Междунар. акад. связи, 2001. – 43 с.
2. Floridi L. Information: A Very Short Introduction. – Oxford: Oxford University Press, 2010. – 130 p.
3. Соколов А. Общая теория социальной коммуникации. – СПб.: Изд-во В.Михайлова, 2002. – 461 с.
4. Орлов В. Основы философии. Ч. 1. Общая философия. Вып. 1. – Пермь: Пермский ун-т, 2001. – 216 с.
5. Galatzer-Levy J. Computer as Smart as a 4-Year-Old? Researchers IQ Test New Artificial Intelligence System. – URL: <http://news.uic.edu/a-computer-as-smart-as-a-four-year-old> (DateViews 26/08/2013)
6. Gaffiot F. Dictionnaire latin français. – Paris.: Hachette, 1934.
7. Ожегов С., Шведова Н. Толковый словарь русского языка. – М.: Азъ, 1992.
8. Об информации, информационных технологиях и защите информации: федеральный закон РФ от 27.07.2006 #149-ФЗ (в ред. 02.07.2013).
9. Усов В. Философские проблемы информатики. – Челябинск: Изд. центр ЮУрГУ, 2010. – 26 с.
10. Hartley R. Transmission of information // Bell System Technical Journal. – 1928. – Vol. 7, № 3. – P. 535-563.
11. Fisher R. Theory of statistical estimation // Proceedings of the Cambridge philosophical Society. – 1925. – Vol. 22. – P.700-725.
12. Shannon C. Communication Theory of Secrecy Systems // Bell System Technical Journal. – 1949. – Vol. 49. – P. 656-715.
13. Колмогоров А. Три подхода к определению количества информации. – М.: Наука, 1965. – С. 4-11.
14. Гоппа В. Введение в алгебраическую теорию информации. – М.: Наука, 1995. – 112 с.
15. Tononi G. Consciousness as integrated information: a provisional manifesto // The Biological Bulletin. – 2008. – Vol. 215 (3). – P. 216–242.
16. Анохин К. Внутри вавилонской библиотеки мозга. Ч.1. – URL: http://tvkultura.ru/anons/show/video_id/292927/brand_id/20898/ (DateViews 30.04.2013).
17. Анохин К. Внутри вавилонской библиотеки мозга. Ч.2. – URL: http://tvkultura.ru/anons/show/video_id/292929/brand_id/20898/ (DateViews 30.04.2013).
18. Мамчур Е., Овчинников Н., Уемов А. Принцип простоты и меры сложности. – М.: Наука, 1989. – 304 с.
19. Wiener N. The Human Use of Human Beings: Cybernetics and Society. – Boston: Houghton Mifflin Co, 1950.
20. Ashby W. Ross. An Introduction to Cybernetics. – London: Chapman & Hall, 1956.
21. Аверкин А., Гаазе-Рапопорт М., Поспелов Д. Толковый словарь по искусственному интеллекту. – М.: Радио и связь, 1992. – 256 с.
22. Russel S., Norvig P. Artificial Intelligence: A Modern Approach. Second Edition. Pearson Education, 2003. – 1112 p.
23. Newell A., Simon H. Computer Science as Empirical Inquiry: Symbols and Search // Communications of the ACM. – 1976. – Vol. 19 (3). – P. 113–126.
24. Sun R. Cognition and Multiagent interaction. – Cambridge University Press, 2005. – 449 p.
25. Searle J. Is the Brain's Mind a Computer Program? // Scientific American. – 1990. – Jan. vol. 262. – p. 26–31.
26. Churchland P., Churchland P. Could a machine think? // Scientific American. – 1990. – Jan. vol. 262. – p. 32–39.
27. Dermott D. Artificial Intelligence and Consciousness // Cambridge Handbook of Consciousness. – Cambridge; New York: Cambridge University Press, 2007.
28. Pfeifer R., Bongard J. How the Body Shapes the Way We Think: a new view of intelligence. – MIT Press, 2006.
29. Chomsky N. On Nature and Language. – Cambridge University Press, 2002. – 218 p.
30. Feldman A. From Molecule to Metaphor: A Neural Theory of Language. – MIT Press, 2008. – 384 p.
31. Ferrucci D. Introduction to “This is Watson” // IBM Journal of Research and Development. – 2012. – Vol.56, № 3.4. – P. 1-15.

Материал поступил в редакцию 09.10.13.

Сведения об авторах

МИШЛАНОВА Светлана Леонидовна – зав. кафедрой лингводидактики Пермского государственного национального исследовательского университета, доктор филологических наук, профессор
e-mail: mishlanovas@mail.ru

БЕРЕСНЕВА Наталья Ириковна – декан философско-социологического факультета Пермского государственного национального исследовательского университета, доктор философских наук
e-mail: beresnev@mail.ru

МИШЛАНОВ Ярослав Витальевич – аспирант кафедры информационных систем и телекоммуникаций Пермского государственного национального исследовательского университета
e-mail: sconymare@yandex.ru

Ю.М. Кувшинская

Согласование сказуемого с именной группой, включающей слова *сколько, столько, много, немало, несколько**

Анализируется вариативность числовых форм сказуемого при согласовании с малоисследованными именными группами – сочетанием слов «сколько», «столько», «много», «немало», «несколько» с существительным. Рассматривается влияние грамматических свойств количественного слова, прежде всего его способности к склонению, на выбор числа сказуемого. Анализируются факторы контекста, способствующие постановке сказуемого во множественном числе в тех предложениях, где статистически преобладают формы единственного числа сказуемого.

Ключевые слова: *вариативность грамматических форм, предикативное согласование, числовые обороты, корпусные методы в лингвистике*

В русском языке, как и в ряде других славянских языков, наблюдается вариативность числовых форм сказуемого при согласовании с подлежащим, выраженным именной группой с количественным значением. В зависимости от того, как именно выражено подлежащее и каково его значение, тенденции в выборе числа (и рода) сказуемого оказываются различны вплоть до противоположности. Так, при согласовании с именной группой, включающей количественное числительное (*два, пять, сто*), сказуемое с большей вероятностью ставится во множественном числе (*Десять организаций перечислили средства для помощи детскому центру*), при согласовании с именной группой, имеющей неопределенно-количественное значение и включающей слова *более/свыше/около + количественное числительное* или слова *несколько, мало, много* и др., преобладают формы единственного числа сказуемого (*На конкурс было представлено немало работ*) [1, с. 28–29; 2, с. 558–559; 3, с. 58–78; 4, с. 259–263; 5 и др.] Различия в выборе числовой формы сказуемого определяются различными причинами, в значительной мере морфологическими и синтаксическими характеристиками именной группы [2, 3 и др.], а также определенностью/неопределенностью количественного значения [6, с. 472–477, 503–504; 7, с. 241–247; 8, с. 34–35; 5].

В настоящей работе предпринимается попытка описать современные тенденции согласования сказуемого с именной группой, имеющей значение неопределенного количества и включающей слова *сколько, столько, много, несколько, немало*. Эти группы предложений исследованы менее других, по-

скольку в большей части из них мало проявляется вариативность числовых форм сказуемого; сказуемое преимущественно употребляется в единственном числе. Автор рассматривает современные статистические данные, касающиеся выбора числовой формы сказуемого в предложениях со *сколько, столько, много, несколько, немало*. Анализируются различия в частотности выбора той или иной формы числа сказуемого в предложениях с разными количественными словами. Обсуждаются основные условия, способствующие постановке сказуемого во множественном числе в таких предложениях, в которых в целом преобладают формы единственного числа.

Исследование проводилось на базе Национального корпуса русского языка (НКРЯ), использовались данные Основного корпуса, выборка была произведена за 2000–2013 гг. В общей сложности нами рассмотрено 1296 примеров, из которых примеров с подлежащим, включающим слово *сколько* – 267; *столько* – 179; *много* – 174; *немало* – 252; *несколько* – 424.

СОВРЕМЕННЫЕ ТЕНДЕНЦИИ В СОГЛАСОВАНИИ СКАЗУЕМОГО С ИМЕННЫМИ ГРУППАМИ, ВКЛЮЧАЮЩИМИ СЛОВА *СКОЛЬКО, СТОЛЬКО, МНОГО, НЕСКОЛЬКО, НЕМАЛО*.

В справочниках и исследованиях отмечается, что при предикативном согласовании со *сколько, столько, много, немало* сказуемое преимущественно принимает форму ед.ч. ср.р. [2, с. 558а; 4, с. 262–263; 7, с. 246]. Однако в издании работы Д.Э. Розенталя 2010 г. указывается, что все чаще в предложениях с рассматриваемыми подлежащими сказуемое ставится

* Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ в 2014 г.

во мн.ч. Данные НКРЯ подтверждают это: в современной речи частотность форм мн.ч. выросла по сравнению с серединой – второй половиной XX в. В то же время динамика роста употреблений сказуемого во мн.ч. неодинакова в предложениях с разными подлежащими:

- Предложения со словами *сколько, столько*

Если для периода середины XX в. (начала 1960-х гг.) А.Е. Супрун указывает, что в предложениях со словами *сколько, столько* соотношение форм ед. и мн.ч. сказуемого равно соответственно 94% : 6%. [2, с. 558а], то к первому десятилетию XXI в., согласно НКРЯ, частотность форм мн.ч. увеличилась, причем в предложениях со *сколько*, значительно – до 19%; в предложениях со *столько* – до 8%.

- Предложения со словом *много*

Не столь резкий, но все же ощутимый рост употреблений сказуемого во мн.ч. наблюдается и в предложениях со словом *много*: если в начале 1960-х гг., по данным Супруна, форма мн.ч. сказуемого встречается лишь в 1% предложений [Там же], то в современном материале – в 3%.

В исследованиях [2, 9] приводятся статистические данные о выборе формы сказуемого в рассматриваемых предложениях начиная с XVIII в. Супрун отмечает отсутствие употреблений в XVIII в. сказуемого во мн.ч. в предложениях со словами *сколько, столько, много* и не приводит данных о согласовании с этими словами для XIX в. [Там же 2, с. 558а].

- Предложения со словом *немало*

Поскольку авторы, занимавшиеся статистическим анализом вариативности, прежде всего А.Е. Супрун, Л.К. Граудина и ее соавторы, не дают статистики для предложений со словом *немало*, была рассмотрена выборка за 1950–1960 гг. из НКРЯ. Во всех примерах, в которых подлежащее включало слово *немало*, сказуемое ставилось в ед.ч. Однако в выборке из НКРЯ за первое десятилетие XXI в. появляются употребления мн.ч. сказуемого – они составляют 4%.

- Предложения со словом *несколько*

В предложениях со словом *несколько* особенно заметен рост частотности употреблений сказуемого

во мн.ч. По данным НКРЯ за последнее десятилетие, сказуемое почти с равной вероятностью принимает форму как ед., так и мн.ч., однако форма ед.ч. употребляется все же немного чаще. Вероятность выбора форм ед. или мн.ч. сказуемого в предложениях с *несколько* – 56% : 44%. Для середины – второй половины XX в. А.Е. Супрун дает соотношение форм ед. и мн.ч. сказуемого как 64% : 36%, а Л.К. Граудина, В.А. Ицкович и Л.П. Катлинская, проводившие статистический анализ газетных материалов в 1960-х – начале 1970-х гг., – как 74,57% : 25,43% [1, с. 28; 2, с. 558а]. Таким образом, по сравнению с серединой XX века, в современной речи сказуемое в этих предложениях ставится во мн.ч. в полтора или почти два раза чаще. Более того, есть основания полагать, что тенденция к выбору формы мн.ч. сказуемого в этих предложениях достаточно сильна и может привести к коренному изменению соотношения форм числа при согласовании сказуемого с именной группой, включающей слово *несколько*.

Исследование вариативности сказуемого при подлежащем с *несколько* в разных речевых сферах показало, что в публицистической сфере формы мн.ч. сказуемого встречаются немного чаще, чем формы ед.ч. Соотношение форм ед. и мн.ч. – 49% (126 примеров): 51% (129 примеров). Этот факт существен, поскольку сфера СМИ представляет собой пространство наиболее динамичного развития языка и оказывает в условиях информационного общества определяющее влияние на современную русскую речь [10, с. 95; 11, с. 185]. В результате есть основания полагать, что согласование сказуемого с *несколько* в форме мн.ч. станет более частотным. В художественной речи формы ед. и мн.ч. равноправны. Преобладание форм ед.ч. сказуемого в предложениях с *несколько* характерно для официально-деловой и особенно для учебно-научной речи, а также для электронной коммуникации [12, с. 102–103].

Как видно из табл. 1, вероятность выбора формы мн.ч. сказуемого снижается, а частотность форм ед.ч. повышается в последовательности:

несколько – *сколько* – *столько* – *много* – *немало*.

Таблица 1

Выбор формы числа сказуемого в предложениях со словами *сколько, столько, много, несколько* по данным НКРЯ¹

	Несколько	Сколько	Столько	Много	Немало
Единственное	55% (233)	81% (218)	92% (166)	95% (156)	95,5% (234)
Множественное	45% (190)	19% (49)	8% (14)	5% (8)	4,5% (11)
Всего	100% (423)	100% (267)	100% (180)	100% (164)	100% (245)

¹ Из рассмотрения были исключены предложения с субстантивированным употреблением *сколько, столько, много и немало*, т.е. без зависимых слов (*За это время сделано немало*) (о различии субстантивного употребления этих слов и употребления в качестве числительного см. [13, с. 292–297]), по причине отсутствия вариативности в предложениях с таким подлежащим.

Тенденции изменения согласования сказуемого в числе с середины XX по начало XXI вв. (ед.ч. : мн.ч.)

	Несколько	Сколько, Столько	Много	Немало
1960–1970-е гг.	64% : 36% (Супрун) 74,57% : 25,43% (Граудина, Ицкович, Катлинская)	94% : 6% (Супрун)	99% : 1% (Супрун)	100% : 0%
2000–2013 гг.	56% : 44%	81% : 19% (сколько) 92% : 8% (столько)	95% : 5%	95,5% : 4,5%

Таблица 3

Выбор формы числа сказуемого в предложениях со словами *сколько, столько, много, немало* – «зона вариативности»

	Сколько	Столько	Много	Немало
Единственное	76% (158)	90% (117)	93% (105)	95% (207)
Множественное	24% (49)	10% (13)	7% (8)	5% (11)
Всего	100% (207)	100% (130)	100% (113)	100% (218)

Несомненно, рост частотности форм мн.ч. отражает общую тенденцию, наблюдаемую в большинстве предложений с количественным подлежащим [4, с. 262–263]: сказуемое все чаще согласуется не грамматически, а семантически, и принимает форму мн.ч. При этом для большинства рассмотренных предложений мы не можем прогнозировать дальнейший рост количества употреблений форм мн.ч. сказуемого (за исключением предложений с *несколько*), поскольку, как показывают диахронические исследования развития согласования сказуемого с количественным подлежащим разных типов, на протяжении последних трех веков вариативность сказуемого в этих предложениях меняется непоследовательно [9]; кроме того, существуют такие условия контекста, которые не допускают постановки сказуемого во мн.ч.

Условия, ограничивающие вариативность

Необходимо отметить, что именные группы со словами *сколько, столько, много, немало* могут включать в себя существительное как мн., так и ед.ч.:

– «Нигде не *останавливалось столько народа*, как перед картинною лавочкою на Щукином дворе. [Леонид Юзефович. Дом свиданий (2001)].

После декабря 1989 года, как говорят в Румынии, много воды утекло по Дунаю. [Николай Морозов. Декабрьские события 1989 года в Румынии: революция или путч? // «Неприкосновенный запас» (2009)].

В этих случаях сказуемое принимает только форму ед.ч., выбор формы мн.ч. невозможен. Учитывая этот факт, для дальнейшего исследования вариативности числовых форм сказуемого необходимо ограничить выборку теми предложениями, в которых может наблюдаться вариативное согласование, ис-

ключив предложения с существительным ед.ч. в составе именной группы.

Уточненное соотношение форм ед. и мн.ч. сказуемого в предложениях с *сколько, столько, много, немало* (за исключением примеров с существительным ед.ч., зависимым от количественного слова), дано в табл. 3. Предложения с *несколько* исключены из рассмотрения, потому что употребление существительного в ед.ч. в именной группе с *несколько* невозможно.

Как нетрудно заметить, в «зоне вариативности» предложений со словами *сколько, столько, много, немало* вероятность выбора формы мн.ч. сказуемого в действительности выше, чем это представлялось на основе общих статистических данных. В предложениях со словом *сколько* форма мн.ч. сказуемого используется чаще, чем в других. При этом форма ед.ч. преобладает в предложениях с *немало* и чуть реже встречается в предложениях с *много*.

Что же касается предложений с *несколько*, то условием, ограничивающим в них вариативность, как и в предложениях с количественными числительными (*Два студента пришли на экзамен*), оказывается употребление определения мн.ч. им.п. в препозиции к подлежащему (*Последние несколько серий фильма вышли в январе*). В нашей выборке из 423 примеров всего в 14 в препозиции к подлежащему стояло определение. Исключив эти предложения, в которых нет вариативности, мы получаем следующее соотношение форм ед. и мн.ч. сказуемого: 58% (233 примера) : 42% (166 примеров). Таким образом, в «зоне вариативности» в целом вероятность выбора ед.ч. сказуемого оказалась немного выше.

Тем не менее, частотность выбора формы мн.ч. сказуемого в предложениях с *несколько* остается заметно более высокой, чем в предложениях со *сколько, столько* и др.

Значение формы единственного числа сказуемого

Прежде чем перейти к анализу зависимости выбора числовой формы сказуемого от грамматических свойств количественного слова, необходимо оговорить значение формы ед.ч. сказуемого в рассматриваемых предложениях.

Традиционно принято говорить о согласовании грамматическом (в ед.ч.) [4, с. 257] или условно-грамматическом [14, с. 176], syntactic agreement [15] – и смысловом, семантическом (во мн.ч.) [4, с. 257; 14, с. 176], semantic agreement [15]. Однако во многих случаях – в том числе при согласовании сказуемого с именной группой, включающей *сколько, столько, много, несколько, немало*, – невозможно говорить о полноценном грамматическом согласовании, т.к. вершина именной группы не имеет форм рода и числа, с которыми может согласовываться сказуемое. В этих случаях постановка сказуемого во мн.ч. отражает семантические свойства подлежащего (значение множества предметов), а форма ед.ч. ср.р., по мнению А.Е. Супруна, а также Е.С. Скобликовой и G. Corbett, является условной, нейтральной, дефолтной [15, с. 64–80; 14, с. 176; 16, с. 13]. Она указывает на неохарактеризованность подлежащего по числу и роду. Поэтому мы будем рассматривать употребление мн.ч. сказуемого как проявление семантического согласования, а ед.ч. ср.р. – как нейтральную, дефолтную форму.

ВЛИЯНИЕ ГРАММАТИЧЕСКИХ СВОЙСТВ КОЛИЧЕСТВЕННОГО СЛОВА НА ВЫБОР ЧИСЛА СКАЗУЕМОГО

При том что для всех рассматриваемых в данной работе предложений характерно согласование сказуемого в ед.ч., очевидно, что соотношение частотности выбора ед. и мн.ч. в предложениях с разным количественным компонентом неодинаково. Очевидно также, что выбор формы сказуемого зависит от грамматических и семантических свойств подлежащего [14, 9]. Супрун убедительно показал, что различия в грамматических свойствах числительных, в том числе различия в согласовании сказуемого с числовым оборотом, во многом обусловлены происхождением числительного: числительные, которые в древности были прилагательными (*два, три, четыре*), согласовывались с существительным, принимая форму мн.ч. Именно эти числительные в составе подлежащего, как правило, определяют выбор формы мн.ч. сказуемого. Числительные, в древности представлявшие собой существительные (*пять, десять, сто*), управляя зависимым существительным, ставились в ед.ч., и в современном языке при согласовании с именной группой, включающей такое числительное, сказуемое чаще, чем в случае числительных *два, три, четыре*, принимает форму ед.ч. [2, с. 524–528]. Развивая эту мысль, Corbett выделил в согласовательных свойствах количественных и местоименных числительных признаки, близкие к признакам существительных и прилагательных, и показал, что количественные слова различаются поведением при согласовании – поведением, подобным существительному (noun-like behavior) или по-

добным прилагательному (adjective-like behavior). Он отмечает слова, близкие к существительным (*миллион, тысяча, ряд, большинство*) (в отечественной традиции эти слова считаются существительными с количественным значением), и слова, близкие к прилагательным (*один, некоторые, два, три, четыре*), а также те, которые занимают среднее положение, проявляют признаки и прилагательного, и существительного (*пять, несколько*) [3, с. 71, 80]. Существенно то, что количественные слова можно выстроить в ряд, в котором постепенно нарастает noun-like behavior, – от слова *один* до слова *миллион*. Наблюдения над тем, как влияют на согласование грамматические особенности количественных слов, были подкреплены диахроническим исследованием динамики развития согласования [9]. Согласно А. Krasovitsky, G. Corbett и их соавторам, сказуемое ставится во мн.ч. и проявляет тенденцию ко все более частому выбору мн.ч. при согласовании с именной группой, включающей числительное, которое ведет себя как прилагательные (*два, три, четыре*). Сказуемое чаще ставится в ед.ч. и не проявляет устойчивой тенденции ко все более частому выбору формы мн.ч. сказуемого при согласовании с количественным словом, в той или иной мере проявляющим поведение существительного (*тысяча, ряд, большинство, несколько, пять, сто*) [9, с. 117–118].

В ряду количественных слов, способных в той или иной степени вести себя как прилагательное или как существительное, рассматриваемые нами *сколько, столько, много, несколько, немало* занимают среднее положение и должны вести себя в значительной степени как существительные, а значит, преимущественно обуславливать выбор ед.ч. сказуемого. Действительно, сказуемое значительно чаще согласуется с такими подлежащими в ед.ч. Однако есть и примеры согласования во мн.ч. Наиболее проблематичным представляется то, что, согласно данным НКРЯ, несмотря на общность noun-like behavior этих слов, мы наблюдаем достаточно заметные различия в согласовании сказуемого с каждым из подлежащих.

Представляется, что эти различия лежат также в области грамматических свойств количественного слова. Рассмотрим применительно к этим словам синтаксические признаки, которые выделяет Corbett, описывая noun-like behavior: согласование с существительным в числе, возможность генитивного или номинативного определения к существительному, согласование по одушевленности [3, с. 58–72].

Все анализируемые слова управляют существительным, если стоят в именительном (или винительном) падеже. Ни одна из именных групп с этими словами не допускает определения в именительном падеже (**несколько старшие студентов, *много новые столов*. Ср.: *три новые двери, пять старших студентов*). Эти признаки свидетельствуют о близости грамматического поведения рассматриваемых слов к существительным. Сложнее обстоит дело с выражением одушевленности, поскольку не все эти слова склоняются, а те, которые могут склоняться, не всегда частотны в косвенных падежах. В то же время у склоняемых слов наблюдается изменение по адъективному типу, т.е. проявляются свойства прилагательного.

Рассмотрим последовательно эти количественные слова.

Несколько склоняется и согласуется с существительным в косвенных падежах (у *нескольких участников, на нескольких страницах*). Употребление *несколько* в косвенных падежах менее частотно, чем в именительном и винительном, в то же время нельзя сказать, что формы косвенных падежей слова *несколько* редки: НКРЯ дает 186 028 вхождений, из которых 24 375 (13,1%) – в косвенных падежах.

Местоименные числительные *столько* и *сколько*, в именительном падеже управляя существительным родительного падежа и проявляя себя как существительные, в косвенных падежах согласуются с существительным по числу, падежу и одушевленности (у *стольких участников, о скольких авторах*), т.е. ведут себя как прилагательные. Однако частотность употребления этих местоимений в косвенных падежах в русской речи невелика. Даже приблизительный подсчет на базе НКРЯ (Основной подкорпус) показывает, что там обнаруживается 35 148 употреблений «столько», из которых лишь 188 (т.е. 0,53%) вхождений – в косвенных падежах; 72567 употреблений «сколько», из которых 929 (1,28%) вхождений – в косвенных падежах. Сравнивая, например, со словом *несколько*, обнаруживаем, что частотность форм косвенных падежей слова *сколько* более чем в 10 раз, а слова *столько* – почти в 25 раз ниже, чем у слова *несколько*.

Итак, *сколько* и *столько* используются в современной русской речи преимущественно в именительном и винительном падежах, хотя имеют формы косвенных падежей², управляют существительным в именной группе². В результате для носителей языка форма именительного/винительного падежа становится основной, склоняемость этих количественных слов оказывается под сомнением. Кроме того, форма именительного/винительного падежа заканчивается на –о, как существительное среднего рода. Как следствие, сказуемое при именной группе, включающей слова *сколько, столько*, чаще ставится в форму ед.ч. ср.р., которую приходится рассматривать как нейтрализованную, дефолтную.

Способностью к склонению и частотностью форм косвенных падежей объясняется также и тот факт, что в предложениях со словом *сколько* формы мн.ч. сказуемого встречаются заметно чаще, чем в предложениях со словом *столько*.

Много обладает общей падежной парадигмой с прилагательным *многие* [18, § 1379; 19, с. 515–516], но, видимо, воспринимается преимущественно как несклоняемая форма. Дело в том, что между числительным *много* и прилагательным *многие* есть различия в значении: *много* обозначает «неопределенно большое количество кого-, чего-л.»; *многие* – «составляющие неопределенно большое число, значительную часть кого-, чего-л.» [19], иначе говоря, неопределенно большое подмножество. Употребления

в косвенных падежах семантически соответствуют скорее прилагательному, означая избранную часть предметов или лиц. А.А. Зализняк отмечает, что в косвенных падежах *много* употребляется лишь с теми существительными, которые называют исчисляемые предметы [20, с. 515–516], (в то время как форма *много* именительного и винительного падежа нередко сочетается с существительными ед.ч.):

Со многими известными спортсменами... (со многими из известных, подмножество);

Для многих известных спортсменов (для многих из тех, кто известен, подмножество).

Ср.: *С многочисленными известными спортсменами...* (с неопределенным большим количеством, неопределенное множество).

Ср. именительный падеж:

На Олимпиаду приехало много спортсменов (неопределенно большое количество, множество);

На Олимпиаду приехали многие известные спортсмены (многие из известных, подмножество).

Таким образом, несогласуемая форма *много* оказывается основной для слова со значением неопределенно большого количества, не предполагающего выбора из множества. С этой формой, управляющей зависимым существительным, и согласуется сказуемое, также принимая нейтральную (дефолтную) форму ед.ч. ср.р.

Что же касается слова *немало*, то оно не обладает склонением и не согласуется с существительным. МАС характеризует *немало* как наречие в значении числа. (При необходимости употребления именной группы с *немало* в косвенном падеже вместо *немало* используется *немалое число*, наречие заменяется сочетанием прилагательного и существительного: *немало студентов – *у немалых студентов – у немалого числа студентов*). Поэтому нейтральное, дефолтное согласование здесь оказывается вполне естественным.

Вероятность выбора ед.ч. сказуемого и допустимость форм мн.ч. оказывается в прямой зависимости от способности количественного слова к склонению, точнее, от возможности его согласования с существительным. Если именная группа (подлежащее) включает количественное слово, согласуемое с существительным в косвенных падежах, то, очевидно, такое слово, как и числительные *два, три, четыре*, отчасти ведет себя как прилагательное и не может полноценно определять согласование сказуемого (ср. [9, с. 118]). В результате выбор числа сказуемого нередко оказывается семантически – ориентированным на множественность, выраженную существительным. Это мы видим в предложениях с *несколько*.

По этой же причине предложения с *несколько* обладают потенциалом увеличения частотности выбора формы мн.ч. сказуемого.

Если именная группа включает в себя количественное слово, которое в косвенных падежах согласуется с существительным, но употребляется в такой позиции редко, то это слово в меньшей степени проявляет свойства прилагательного, воспринимается подобно существительному, причем, вероятно, несклоняемому, и в большей степени влияет на форму сказуемого, диктуя постановку сказуемого в ед.ч. В

² Нельзя не отметить, что это отражает общую тенденцию к утрате склонения и к противопоставленности форм прямых и косвенных падежей, что отмечается большинством исследователей (см. например [17] и др.).

то же время в предложениях с этими словами сохраняется возможность семантического согласования и выбора формы мн.ч. сказуемого. Это мы видим в предложениях со словами *сколько, столько*.

Если слово практически не используется в косвенных падежах или вовсе не склоняется, то оно полностью теряет черты прилагательного, согласуемого слова и проявляет главным образом черты несклоняемого существительного. В таких случаях сказуемое, согласуясь с ним, принимает форму ед.ч. Примером этого служат предложения с *много и немало*.

Таким образом, при согласовании с именной группой, включающей *сколько, столько, много, несколько, немало*, четко прослеживается закономерность, в силу которой склонность к выбору мн.ч. сказуемого в большей степени проявляется в предложениях, где количественное слово в именной группе ведет себя как согласуемое (как прилагательное) и в меньшей – в предложениях, где количественное слово проявляет признаки несклоняемого существительного.

В результате высказанная Corbett идея иерархии количественных слов с точки зрения их грамматических свойств, влияющих на согласование сказуемого, подтверждается. При этом ключевым свойством, определяющим место количественного слова в иерархии и вероятность семантического или дефолтного согласования, оказывается именно способность к склонению и, соответственно, согласованию с существительным в косвенном падеже.

Однако остается вопрос о том, в силу каких причин в современной русской речи стал возможным выбор формы мн.ч. сказуемого даже при подлежащем, включающем несклоняемое слово (например, *немало*).

ФАКТОРЫ, ОПРЕДЕЛЯЮЩИЕ ВЫБОР МНОЖЕСТВЕННОГО ЧИСЛА СКАЗУЕМОГО В ПРЕДЛОЖЕНИЯХ СО СЛОВОМ НЕМАЛО

В выборке из НКРЯ за 2000–2013 гг. встретилось 252 предложения с подлежащим, включающим *немало*. Из них в 11 сказуемое стояло в форме мн.ч. Исходя из факторов, выделенных в науке и обычно влияющих на выбор числа сказуемого в предложениях с количественным подлежащим, попытаемся описать особенности этих 11 предложений на фоне всей выборки.

1. Несомненно, постановке сказуемого во мн.ч. способствует фактор одушевленности существительного, входящего в именную группу:

а) *Кроме того, немало ребят посещают городские творческие кружки и секции.* [Возраст, золотом

отмеченный (2004) // «Уральский автомобиль» (Миасс), 2004.02.05].

Данные о влиянии одушевленности/неодушевленности существительного в составе подлежащего на выбор числа сказуемого в предложениях с *немало* представлены в табл. 4.

Как видно из таблицы, в предложениях, в которых сказуемое стоит в ед.ч., подлежащее почти всегда оказывается существительным неодушевленным, а для предложений со сказуемым мн.ч. в качестве подлежащего характерно одушевленное существительное. В то же время не исключено согласование сказуемого в ед. ч. с одушевленным существительным и сказуемого во мн.ч. с неодушевленным существительным.

Приведем примеры постановки сказуемого во мн.ч. при неодушевленном подлежащем:

б) *Не случайно, наверно, уже в начале правления Путина немало важных международных встреч второго Президента России прошли и будут, видимо, проходить и дальше в Санкт-Петербурге.* [Борис Грищенко. Посторонний в Кремле (2004)].

в) *Немало станций и вестибюлей одеты в мрамор светлых оттенков из месторождения близ села Коелга (Урал) и крымских месторождений Блюк-Янкой, Чергунь, Кадыковское (например, станции «Курская», «Охотный ряд»).* [Т. Зимица. Подземный музей под Москвой // «Наука и жизнь», 2008].

В этих примерах есть дополнительные условия, поддерживающие выбор формы мн.ч. сказуемого: однородные сказуемые в примере а, однородные подлежащие – в примере б. В то же время даже эти условия не исключают постановки сказуемого в ед.ч. как в предложениях с *немало*, так и в других предложениях с количественным подлежащим [5, 21].

Таким образом, при несомненном влиянии одушевленности и других факторов, способствующих выбору мн.ч. сказуемого, эти факторы не полностью определяют выбор сказуемого и, в частности, употребление столь редких форм мн.ч. сказуемого.

2. Другой существенный фактор, влияющий на выбор числа, – порядок слов. Взаимодействие одушевленности и порядка слов Corbett считает условием, в значительной мере определяющим выбор сказуемого при согласовании с количественным подлежащим [15, с. 11–12].

Порядок слов несомненно оказывает влияние на возможность постановки сказуемого во мн.ч. в предложениях с *немало*: в большинстве предложений со сказуемым мн.ч. наблюдается прямой порядок слов, в то время как в предложениях со сказуемым ед.ч. преобладает инверсия (см. табл. 5).

Таблица 4

Одушевленность и выбор числа сказуемого³:

	Неодушевленное	Одушевленное	Всего
Единственное	98% (150)	88% (57)	207
Множественное	2% (3)	12% (8)	11
Всего	100% (153)	100% (65)	218

³ Данные приводятся за вычетом тех примеров, в которых заведомо нет вариативности.

В то же время порядок слов, как и одушевленность, не предопределяет выбор числа сказуемого полностью. Отметим, что в двух из трех предложений, в которых сказуемое стоит во мн.ч. при инверсии, наблюдаются другие факторы, способствующие выбору формы мн.ч. сказуемого: одушевленность – во всех примерах, однородные сказуемые – в примере *г*, составное глагольное сказуемое в примере *д*. В то же время в примере *е* сказуемое по своим лексическим свойствам вероятнее могло бы стоять в ед.ч. [6, с. 461–468; 4, с. 261; 22].

г) *В этом направлении думают и небезуспешно работают немало людей в нашей стране, и это питает сдержанную надежду на то, что научное книгоиздание может обрести необходимую масштабность и значимость.* [«Барометр» зашкалило на аномалиях (2002) // «Витрина читающей России», 2002.09.13].

д) *Третьи свои победы в Олимпиадах сумели одержать немало спортсменов...* [Евгений Гик, Екатерина Гупало. Пекин–2008 // «Наука и жизнь», 2009].

е) *А не так давно Совет по национальной стратегии, в который входят немало маститых российских ученых, выступил с сенсационным предупреждением: «Страна на пороге ползучего олигархического переворота».* [Павел Воцанов. Проект «враги народа» (2003) // «Новая газета», 2003.01.02].

В результате мы видим, что возможность постановки сказуемого во мн.ч. в значительной мере обусловлена и поддержана факторами, которые обычно способствуют выбору формы мн.ч. сказуемого в предложениях с количественным подлежащим. Однако ни один из этих факторов не является исключительной причиной выбора формы мн.ч.

Обращаясь к взаимодействию одушевленности и порядка слов в предложениях с *немало*, мы находим следующую картину (см. табл. 6).

При расчете за 100% было принято общее количество предложений с *немало* в выборке.

В результате мы видим, что предложения со сказуемым ед.ч. можно расположить, в зависимости от сочетания факторов, в такой последовательности (по убыванию частотности):

Неодуш. сущ. в составе подлежащего, инверсия – одуш. сущ. в составе подлежащего, инверсия – неодуш. сущ. в составе подлежащего, прямой порядок слов – одуш. сущ. в составе подлежащего, прямой порядок слов.

Как видно по этой шкале, для предложений со сказуемым ед.ч. наиболее значимым фактором оказывается инверсия.

Предложения со сказуемым мн.ч. располагаются так:

Одуш. сущ. в составе подлежащего, прямой порядок слов – одуш. сущ. в составе подлежащего, инверсия – неодуш. сущ. в составе подлежащего, прямой порядок слов.

Для предложений со сказуемым во мн.ч. наиболее значима одушевленность.

Итак, в предложениях с *немало* наблюдается приблизительное соответствие иерархии, предложенной Corbett. Сказуемое ставится во мн.ч. прежде всего в предложениях, в которых подлежащее включает одушевленное существительное и порядок слов является прямым; в ед.ч. – при неодушевленности существительного в именной группе и инверсии. Однако на выбор формы мн.ч. одушевленность оказывает большее влияние, чем порядок слов. На выбор ед.ч., напротив, в большей степени влияет порядок слов.

Это наблюдение коррелирует с выводами многих исследователей о том, что для семантического согласования (во мн.ч.) существенны определенные семантические признаки подлежащего и сказуемого, прежде всего указание на активность, способность к активному самостоятельному действию [6, с. 505; 4, с. 259; 23, с. 372; 22]. Напротив, для дефолтного согласования в некотором смысле удобнее определен порядок слов – инверсия, поскольку свойства подлежащего в этом случае в меньшей степени влияют на формы сказуемого [21, с. 136–138].

Таблица 5

Инверсия и выбор числа сказуемого в предложениях с *немало*

	Инверсия	Прямой порядок слов
Единственное	98% (195)	60% (12)
Множественное	2% (3)	40% (8)
Всего	100% (198)	100% (20)

Таблица 6

Влияние одушевленности и порядка слов на выбор числовой формы сказуемого в предложениях с *немало*

	Неодушевленное подлежащее		Одушевленное подлежащее	
	Инверсия	Прямой порядок	Инверсия	Прямой порядок
Единственное	64% (139)	5% (10)	26% (56)	0,9% (2)
Множественное	0	1% (3)	1% (3)	2% (5)

Другие факторы, которые обычно способствуют выбору формы мн.ч. сказуемого, в предложениях с *немало* проявляются не настолько последовательно, чтобы можно было говорить об их определяющем влиянии.

В целом приходится констатировать, что возможность выбора формы мн.ч. сказуемого в предложениях с *немало*, для которых характерна форма ед.ч. сказуемого (эта форма в прошлом, видимо, была единственно допустимой), в значительной степени – но не абсолютно – обусловлена рядом факторов, в первую очередь одушевленностью существительного в именной группе и прямым порядком слов. В то же время, очевидно, должны быть и другие – и, по-видимому, решающие – причины, в силу которых в предложениях, где в прошлом сказуемое ставилось лишь в форме ед.ч., стала возможной форма мн.ч. Представляется, что эти причины лежат в области коммуникативной организации предложения [6, с. 490; 24; 21, с. 143–146], однако, это требует специального исследования.

ЗАКЛЮЧЕНИЕ

Таким образом, в предложениях с подлежащим, включающим слова *сколько, столько, много, несколько, немало*, за последние полвека наблюдается небольшая тенденция к постановке сказуемого во мн.ч. Эта тенденция особенно заметна, если выделить в выборке примеров «зону вариативности», исключив те случаи, когда выбор числовой формы (обычно ед.ч.) предопределен.

Вероятность семантического согласования сказуемого с подлежащим оказалась тем выше, чем более количественное слово проявляет способность к словоизменению, а конкретнее – к склонению (имеет формы косвенных падежей), т.е. обнаруживает adjective-like behavior. Наименее частотны формы мн.ч. сказуемого в предложениях с *немало* – несклоняемым словом. Примечательно то, что все рассмотренные количественные слова, при общих признаках noun-like behavior, все-таки проявляют некоторые черты поведения прилагательного, причем в разной степени.

Выбору формы мн.ч. сказуемого в наименее благоприятной для этого ситуации – в предложениях с несклоняемым количественным словом – заметно способствуют одушевленность существительного в именной группе, а также – в меньшей степени – прямой порядок слов, однако решающее значение, как представляется, здесь должна иметь коммуникативная организация высказывания.

СПИСОК ЛИТЕРАТУРЫ

1. Граудина Л.К., Ицкович В.А., Катлинская Л.П. Грамматическая правильность русской речи. Опыт частотно-стилистического словаря вариантов. – М.: Наука, 1976.
2. Супрун А.Е. Славянские числительные (становление числительных как особой части речи). Дис. ... д-ра филол. наук. – Л., 1965.

3. Corbett G.G. Predicate agreement in Russian. – Birmingham, 1979.
4. Розенталь Д.Э. Справочник по правописанию и литературной правке. – М.: Айрис-Пресс, 2010.
5. Кувшинская Ю.М. Согласование сказуемого с подлежащим, выраженным счетным оборотом // Проблемы русской стилистики по данным НКРЯ. – URL: <http://studiorum.ruscorpora.ru/stylistics/oboroty.html>.
6. Скобликова Е.С. Согласование и управление как способы синтаксической организации слов в русском языке. Дис. ... д-ра филол. наук. – Куйбышев, 1969.
7. Бельчиков Ю.А. Практическая стилистика современного русского языка. – М.: АСТ-Пресс, 2012.
8. Никунласси А. О числе сказуемого при количественном подлежащем в русском и финском языках // Семантические проблемы русского языка. – Таллинн, 2002. – С. 33–45.
9. Krasovitsky A., Baerman M., Brown D., Corbett G.G., Williams P. Predicate agreement in Russian: a corpus-based approach // Diachronic Slavonic Syntax. Gradual Changes in Focus. – München–Berlin–Wien, 2009. – P. 109–121.
10. Солганик Г.Я. О современной языковой ситуации // Стилистика сегодня и завтра: медиатекст в прагматическом, риторическом и лингвокультурологическом аспектах. – М.: МГУ, 2010.
11. Бельчиков Ю.А. Нормативно-стилистическая база языка средств массовой информации // Язык массовой и межличностной коммуникации. – М.: Медиамир, 2007. – С. 185–207.
12. Кувшинская Ю.М. Тенденции развития вариативности в согласовании сказуемого с подлежащим, включающим числительное «несколько» // Материалы международного научного семинара «Вопросы русской исторической грамматики и славяноведения: к 175-летию со дня рождения Ватрослава Ягича (1838–1923)». Петрозаводск, 19–20 сентября 2013 г. – Петрозаводск: Изд-во ПетрГУ, 2013. – С. 99–104.
13. Булыгина Т.В., Шмелев А.Д. Числительные в русском языке: лексикографические лакуны // Слово в тексте и в словаре. Сб. статей к семидесятилетию академика Ю.Д. Апресяна. – М.: Языки русской культуры, 2000. – С. 291–305.
14. Скобликова Е.С. Согласование и управление в русском языке. 2-е изд-е. – М.: URSS, 2005.
15. Corbett G.G. Agreement in Slavic. – Bloomington, 1998.
16. Супрун А.Е. Славянские числительные (становление числительных как особой части речи). Автореферат дис. ... д-ра филол. наук. – Л., 1965.
17. Старикова Г.Н. Современные числительные в аспекте языковой динамики // Вестник Томского государственного университета. Сер. «Филология». – 2011. – № 4 (16). – С. 34–39
18. Русская грамматика. – М.: Наука, 1980.

19. Словарь русского языка: в 4 т. / РАН, Ин-т лингвистич. исследований. Под ред. А. П. Евгеньевой. 4-е изд., стер. – М.: Рус. яз.; Полиграфресурсы, 1999.
20. Зализняк А.А. Грамматический словарь русского языка. Словоизменение. – М.: Русский язык, 1980.
21. Кувшинская Ю.М. Согласование сказуемого с подлежащим, выраженным именной группой с количественным значением (по данным НКРЯ за 2000–2010 гг.) // Русский язык в научном освещении. – 2013. – № 26. – С. 112–150.
22. Robblee K.E. Individuation and Russian agreement // Slavic and East European Journal. – 1993. – Vol. 37. – P. 423–441.
23. Голуб И.Б. Новый справочник по русскому языку и практической стилистике. – М.: Эксмо, 2008.
24. Соколова Е.В. О взаимосвязи коммуникативного задания высказывания и колебания координации подлежащего и сказуемого в формах числа // Вестник СПбГУ. Сер. 2. – 1998. – Вып. 2, № 9. – С. 36–42.

Материал поступил в редакцию 10.06.14.

Сведения об авторе

КУВШИНСКАЯ Юлия Михайловна – кандидат филологических наук, доцент факультета филологии НИУ ВШЭ, Москва
e-mail: kjulia4@yandex.ru