

стоятельность выбранных моделей численного моделирования при проектировании накатных мелко-модульных зубчатых передач из условия обеспечения заданных показателей качества.

В результате проделанной работы была подтверждена адекватность проведенных экспериментов и имитационного моделирования процесса накатки мелко-модульных зубчатых колес, а также разработана методика измерения геометрии мелко-модульных зубчатых колес на основе использования цифрового оптического микроскопа и обработки изображения на ЭВМ.

Библиографические ссылки

1. Карпов Ю. Имитационное моделирование систем. Введение в моделирование с AnyLogic 5. СПб. : БХВ-Петербург, 2005.

2. Барбарич М. В., Хоруженко М. В. Накатывание цилиндрических зубчатых колес. М. : Машиностроение, 1970.

3. Вавилов Д. В., Иптышев А. А., Усаков В. И. Моделирование накатывания мелко-модульных цилиндрических зубчатых передач с заданными показателями качества // Вестник СибГАУ. Вып. 4(21). 2008. С. 67–70.

4. Автоматизированный программный модуль для экспорта геометрии зубчатых колес в CAD-среду SolidWorks : программа для ЭВМ : свидетельство о гос. регистрации № 2009613284 / А. П. Смирнов, Д. В. Вавилов, А. А. Иптышев, Д. Б. Елисеев. № 2009612213 ; заявл. 13.05.2009.

5. Степнов М. Н. Статистические методы обработки результатов механических испытаний. М. : Машиностроение, 1985.

D. V. Vavilov, A. A. Iptyshev, A. P. Smirnov, D. V. Eliseev, M. M. Kolegova

METHOD OF ROLLED FINE PITCH SPUR GEAR ANALYSIS BY DIGITAL MICROSCOPE

In the article the authors present a method of comparative analysis of rolled fine pitch spur gear, special needs drives, made by numeric simulation and natural experiment.

Keywords: simulation modeling, rolled fine pitch spur gear, digital microscope.

© Вавилов Д. В., Иптышев А. А., Смирнов А. П., Елисеев Д. В., Колегова М. М., 2011

УДК 621.31:51

П. В. Валь

КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ ГОРНОГО ПРЕДПРИЯТИЯ С ИСПОЛЬЗОВАНИЕМ ОДНОФАКТОРНЫХ МЕТОДОВ

Проведен анализ качества краткосрочного прогнозирования электропотребления предприятия горной промышленности с использованием популярных однофакторных методов прогнозирования (сезонной модели АРПСС, метода Хольта–Винтерса, а также наивного метода).

Ключевые слова: электропотребление, прогнозирование, модель АРПСС, метод Хольта–Винтерса, наивный метод.

Устойчивое развитие горной промышленности в значительной мере зависит от повышения ее конкурентоспособности за счет снижения издержек производства. В настоящее время для предприятий отрасли характерен опережающий рост уровня электропотребления по сравнению с ростом выпуска продукции. Это связано с изменением условий добычи и переработки полезных ископаемых, внедрением природоохранных программ и т. д. В последние годы также наблюдается устойчивый рост цен на электроэнергию. Таким образом, особую актуальность приобретает снижение затрат на электроэнергию, которая составляет значительную часть себестоимости продукции предприятий горной промышленности.

Для решения поставленной задачи можно выделить два основных направления:

- снижение электропотребления за счет повышения эффективности использования энергоресурсов;
- снижение удельной стоимости потребленной электроэнергии.

Одним из возможных путей снижения удельной стоимости потребленной электроэнергии является ее покупка на оптовом рынке электроэнергии и мощности (ОРЭМ). Однако правила функционирования ОРЭМ предусматривают жесткие требования к его участникам по прогнозированию планового почасового электропотребления. Так, участники ОРЭМ в сутки, предшествующие операционным, должны подавать администратору торговой системы почасовую заявку на электропотребление для каждого часа операционных суток. Неправильная подача заявок на электроэнергию может привести, во-первых, к значи-

тельным экономическим потерям, вызванным покупкой/продажей электроэнергии на балансирующем рынке (БР) и оплатой части отрицательного небаланса БР энергосистемы, и, во-вторых, к наложению санкций со стороны коммерческого оператора. Таким образом, для участников ОРЭМ необходимым условием их нормального функционирования является наличие методики прогнозирования почасового электропотребления, обеспечивающей приемлемое качество прогнозных расчетов.

К настоящему времени разработано большое число методов прогнозирования электропотребления, однако не существует универсального метода, который мог бы с одинаковым успехом применяться для различных типов объектов [1]. Это вызвано главным образом тем, что любой объект прогнозирования, будь то региональная энергосистема или конкретное промышленное предприятие, обладает уникальным характером электропотребления и сложными зависимостями между электропотреблением и влияющими на него факторами. В связи с этим возникает задача выбора методики прогнозирования электропотребления, обладающей удовлетворительным качеством прогноза.

На практике нередко ситуации, когда отсутствует достоверная информация о том, какие именно факторы влияют на процесс электропотребления, а также о количественном влиянии и каждого из них в отдельности, и их произвольных комбинаций. Возможным вариантом выхода из таких ситуаций является использование однофакторных методов прогнозирования (моделей на основе временных рядов), в которых определение прогнозных значений переменной осуществляется на основе прошлых и текущих значений этой же переменной.

Автором был проведен анализ качества краткосрочного прогнозирования электропотребления предприятия горной промышленности на примере ООО «Сорский горно-обогатительный комбинат» (Сорский ГОК) с использованием популярных однофакторных методов прогнозирования: сезонной модели АРПСС, метода Хольта–Винтерса, а также наивно-го метода.

Характеристика объекта исследования. Основной производственный цикл ООО «Сорский ГОК» включает в себя добычу молибденовой руды, ее обогащение, т. е. получение молибденового концентрата, и производство конечной продукции – ферромолибдена. На предприятии также существует ряд вспомогательных производств.

Электропотребление горных предприятий при добыче и обогащении руд имеет специфические особенности, вызванные его случайным характером, связанным с влиянием значительного числа горно-геологических, технологических, производственных, климатических и других факторов, и представляет собой сложный нестационарный процесс (рис. 1).

По приведенному на рис. 1 участку временного ряда можно судить об особенностях электропотребления ООО «Сорский ГОК». Отчетливо прослежива-

ется суточный цикл продолжительностью 24 ч. В пределах суток наблюдаются циклы продолжительностью 8 и 12 ч, вызванные смешанным характером работы предприятия в три смены по 8 ч и в две смены по 12 ч. Недельный цикл продолжительностью 168 ч практически не прослеживается из-за непрерывности рабочего цикла. Временной ряд также содержит провалы электропотребления, вызванные плановыми отключениями части электроприемников предприятия.

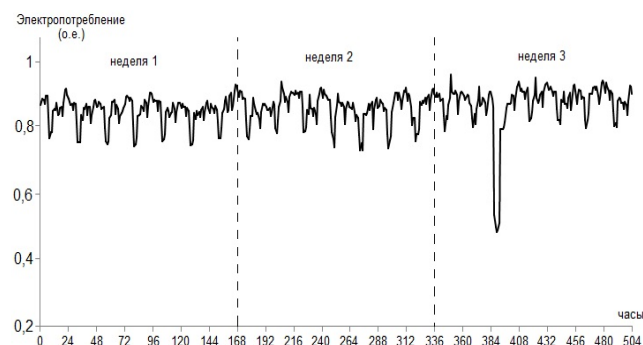


Рис. 1. Почасовые значения электропотребления с 5 по 25 января 2009 г.

Описание методов прогнозирования. *Сезонная модель авторегрессии интегрированного скользящего среднего* (АРПСС) (Autoregressive Integrated Moving Average, ARIMA), разработанная Дж. Боксом и Г. Дженкинсом [2], позволяет прогнозировать нестационарные циклические временные ряды.

Сезонная модель Бокса–Дженкинса может быть представлена в виде

$$\text{ARIMA}(p, d, q) \cdot (P_s, D_s, Q_s)_s, \quad (1)$$

где p – порядок составляющей авторегрессии; d – порядок разности дискретной производной; q – порядок скользящего среднего; P_s – порядок сезонной авторегрессии; D_s – порядок сезонной разности (сезонной производной); Q_s – порядок сезонного скользящего среднего; s – период сезонности.

Нестационарные ряды, содержащие сезонные колебания, приводят к стационарным с помощью взятия последовательных разностей между соседними элементами (несезонных разностей) и элементами, отстоящими друг от друга на значение периода сезонности s .

После взятия разностей ряд становится стационарным, т. е. удовлетворяющим условиям модели (1). Разностный ряд может быть представлен в виде

$$\Delta Y_t = \Delta^d Y_t + \Delta^{D_s} Y_t, \quad (2)$$

$$\begin{aligned} \Delta^d Y_t &= \alpha_1 \cdot Y_{t-1} + \dots + \alpha_p \cdot Y_{t-p} + \\ &+ \varepsilon_t - \beta_1 \cdot \varepsilon_{t-1} - \dots - \beta_q \cdot \varepsilon_{t-q}, \end{aligned} \quad (3)$$

$$\begin{aligned} \Delta^{D_s} Y_t &= \gamma_1 \cdot Y_{t-s} + \dots + \gamma_{P_s} \cdot Y_{t-P_s \cdot s} + \\ &+ \varepsilon_t - \delta_1 \cdot \varepsilon_{t-s} - \dots - \beta_{Q_s \cdot s} \cdot \varepsilon_{t-Q_s \cdot s}, \end{aligned} \quad (4)$$

где $\Delta^d Y_t$ – несезонная составляющая порядка p , d , q коэффициентов линейной регрессии α между элементами ряда Y_t и коэффициентов линейной регрессии β между значениями ошибок ε_t ; $\Delta^{D_s} Y_t$ – сезонная составляющая порядка P_s , D_s , Q_s коэффициентов линейной регрессии γ между элементами ряда Y_t и коэффициентов линейной регрессии δ между значениями ошибок ε_t .

Для возврата к исходному временному ряду и получения прогноза необходимо выполнить обратную операцию, т. е. проинтегрировать данные:

$$Y_t = Y_{t-1} + \Delta Y_t. \quad (5)$$

Метод Хольта–Винтерса (Holt–Winters) является усовершенствованием метода экспоненциального сглаживания временного ряда. Этот метод учитывает линейный тренд и мультипликативную сезонность [3; 4]. Его математическое описание может быть представлено в виде системы уравнений

$$\begin{cases} \Omega_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(\Omega_{t-1} - T_{t-1}), \\ T_t = \beta(\Omega_t - \Omega_{t-1}) + (1-\beta)T_{t-1}, \\ S_t = \gamma \frac{Y_t}{\Omega_t} + (1-\gamma)S_{t-s}, \\ \hat{Y}_{t+p} = (\Omega_t + pT_t)S_{t-s+p}, \end{cases} \quad (6)$$

где $\hat{Y}_{t+p}$ – прогноз на p отсчетов по времени вперед; Y_t – реальное значение на момент времени t ; s – период сезонности; Ω_t и T_t – сглаженные уровень и тренд; S_t – сезонный индекс суточного цикла; α , β и γ – параметры сглаживания.

Самым простым из возможных методов прогнозирования является *наивный метод* (naïve). Выражение для сезонного наивного метода имеет вид [4]

$$\hat{Y}_{t+p} = Y_{t-s+p}, \quad (7)$$

где $\hat{Y}_{t+p}$ – прогноз на p отсчетов по времени вперед.

Описание эксперимента. Анализ качества прогнозирования электропотребления ООО «Сорский ГОК» выбранными методами проводился с помощью пассивного статистического эксперимента. В ходе эксперимента был выбран временной ряд электропотребления с 15 декабря 2008 г. по 15 марта 2009 г., содержащий 2 184 почасовых значений (рис. 2).

Из второй части этого ряда с помощью генератора случайных чисел случайным образом были выбраны 10 сут, для которых впоследствии строился прогноз. Для сезонной модели АРПСС и метода Хольта–Винтерса оценивание неизвестных параметров и построение прогноза осуществлялось для различных периодов оснований прогноза (периодов предыстории) (от 240 до 1 200 значений с интервалом 120).

Идентификация сезонной модели АРПСС, т. е. определение порядков авторегрессии (p , P_s), скользяще-

го среднего (q , Q_s) и разности (d , D_s) в формуле (1), опиралась на исследование графиков рядов, автокорреляционных функций и частных автокорреляционных функций. Этап идентификации является основным в получении модели, от принятых на нем решений зависит ее адекватность и точность прогнозирования. По результатам идентификации было принято решение об использовании модели $ARIMA(1,0,0) \cdot (0,1,1)_{24}$.



Рис. 2. Временной ряд электропотребления, используемый в эксперименте

Для данной спецификации модели АРПСС неизвестными параметрами являются значения коэффициентов α_1 и δ_1 в формулах (3) и (4). Для их оценивания был использован квазиньютоновский алгоритм максимизации правдоподобия, заключающийся в нахождении таких параметров модели, при которых сумма квадратов остатков будет наименьшей. Получены следующие значения параметров модели АРПСС: $\alpha_1 = 0,81$, $\delta_1 = 0,64$.

Для прогнозирования с помощью метода Хольта–Винтерса необходимо оценить параметры сглаживания α , β и γ в системе (6). Оценивание этих параметров также производилось с помощью квазиньютоновского алгоритма минимизации квадратичной ошибки прогнозирования на расчетном множестве. В результате определены следующие значения параметров сглаживания модели Хольта–Винтерса: $\alpha = 0,82$, $\beta = 0,00$, $\gamma = 0,06$.

При прогнозировании наивным методом согласно формуле (7) в качестве прогнозных использовались значения электропотребления соответствующих часов предыдущих суток.

После определения параметров моделей был проведен анализ их адекватности.

Анализ адекватности моделей. Под *адекватностью модели* понимается степень совпадения свойств модели и моделируемого объекта. Причиной ошибочного прогноза является наличие источников регулярных и нерегулярных ошибок. К источникам регулярных ошибок относят неадекватный метод прогнозирования, недостоверные и недостаточные исходные данные и т. п. Причиной нерегулярных (случайных) ошибок является непредсказуемость развития объекта прогнозирования. Таким образом, в ходе анализа аде-

кватности прогнозирующей модели производится проверка остатков (ошибок):

$$e_t = Y_t - \hat{Y}_t,$$

где Y_t – фактическое значение; $\hat{Y}_t$ – прогнозное значение.

При корректном выполнении моделирования остатки должны являться результатом случайного рассеяния, а не доминирующего действия какого-либо неконтролируемого и неуправляемого воздействия [5].

Анализ остатков включает в себя их проверку на нормальность распределения, нулевое математическое ожидание, однородность дисперсий и отсутствие значимой сериальной автокорреляции.

Для предварительного графического анализа адекватности удобно перейти к стандартизированным остаткам:

$$se_t = \frac{e_t}{\sigma_e},$$

где σ_e – стандартное отклонение остатков.

По характеру поведения этих остатков можно судить о степени их принадлежности белому шуму (рис. 3). Кроме того, по результатам графического анализа строятся гистограмма распределения остатков и их автокорреляционная функция.

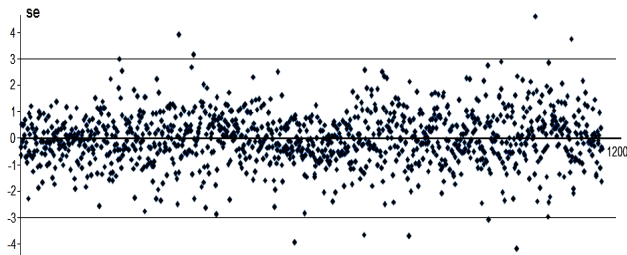


Рис. 3. Стандартизированные остатки сезонной модели АРПСС

Строгая проверка гипотезы о соответствии распределения остатков нормальному закону распределения осуществляется по критерию согласия Пирсона. Нулевая гипотеза о соответствии эмпирического распределения нормальному принимается, если величина статистики (критерия расхождения) χ^2 не превосходит квантиль закона распределения $\chi^2_{кр}$ заданного уровня значимости α с заданным числом степеней свободы:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \leq \chi^2_{кр},$$

где n_i – число элементов выборки из i -го интервала; p_i – предполагаемая вероятность попадания в i -й интервал; n – общее число элементов выборки; k – число интервалов.

Для нахождения автокорреляции остатков используется Q -тест Льюнга–Бокса. Нулевая гипотеза об отсутствии значимой сериальной автокорреляции принимается, если значение статистики Q не превосходит

квантиль распределения $\chi^2_{кр}$ заданного уровня значимости α с заданным числом степеней свободы:

$$Q = n(n+2) \sum_{k=1}^m \frac{\rho_k^2}{n-k} \leq \chi^2_{кр},$$

где ρ_k – автокорреляция k -го порядка; n – число наблюдений; m – число проверяемых лагов.

Оценка адекватности модели может быть разной: от полной адекватности до частичной адекватности или неадекватности. Однако при составлении модели следует искать разумный компромисс: модель должна иметь достаточную для решения задачи адекватность и не быть очень сложной. В описываемом исследовании в ходе анализа адекватности было установлено, что степень адекватности используемых моделей можно считать достаточной, несмотря на то, что часто эти модели не проходили строгих тестов адекватности.

Результаты исследования. Для сравнения различных альтернативных прогнозов необходим критерий оценки качества прогноза [6]. В качестве меры оценки точности прогноза использовался традиционный показатель MAPE (Mean Absolute Percentage Error – среднеабсолютная процентная ошибка):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \cdot 100 \%,$$

где Y_t – фактическое значение; $\hat{Y}_t$ – прогнозное значение; n – число наблюдений. Показатель MAPE характеризует относительную точность прогноза. При этом $MAPE > 0$.

Для оценки относительной степени смещения (постоянного занижения или завышения) прогноза использовался показатель MPE (Mean Percentage Error – средняя процентная ошибка):

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{Y_t - \hat{Y}_t}{Y_t} \cdot 100 \%.$$

При условии, что потери при прогнозировании, связанные с завышением фактического будущего значения, уравниваются его занижением, идеальный прогноз должен быть несмещенным, а MPE должна стремиться к нулю.

Результаты прогнозирования суточного почасового электропотребления на 26 февраля 2009 г. приведены ниже (рис. 4). Для рассматриваемых суток определены следующие значения ошибок MAPE (MPE): сезонная модель АРПСС – 2,91 % (–1,36 %); метод Хольта–Винтерса – 3,60 % (–3,24 %); наивный метод – 4,70 % (1,81 %).

Значения ошибок MAPE и MPE, полученные при проведении эксперимента, представлены в таблице.

В результате анализа качества прогнозирования электропотребления с использованием различных методов было установлено, что ни один из рассматриваемых методов не показал результата, значительно превосходящего результаты остальных методов. В среднем модель сезонной АРПСС показала несколько лучший результат, чем метод Хольта–Винтерса. Наивный метод в большинстве случаев давал наихудший результат.

Значения ошибок эксперимента

Метод	Показатель	Прогнозные сутки									
		14.03.09	26.02.09	23.02.09	12.02.09	24.02.09	7.02.09	3.02.09	15.03.09	8.02.09	7.03.09
Сезонная модель АРПСС	МАРЕ, %	4,59	2,94	2,00	3,98	1,98	1,87	3,40	2,09	2,44	3,23
	МРЕ, %	-4,40	-1,24	1,06	1,02	1,92	-0,97	3,36	1,28	1,48	-2,77
Хольта-Винтерса	МАРЕ, %	4,41	3,60	4,15	3,52	1,48	3,97	6,03	2,15	2,25	3,23
	МРЕ, %	-4,17	-3,04	3,98	-1,64	1,21	3,95	6,03	1,23	1,60	-2,98
Наивный	МАРЕ, %	4,14	4,70	3,87	5,83	2,93	5,65	3,52	4,55	3,45	4,70
	МРЕ, %	-3,64	1,81	0,88	1,06	2,54	-2,30	2,46	4,41	3,20	-3,97

Примечание. Для сезонной модели АРПСС и метода Хольта-Винтерса приведены средние значения ошибок.

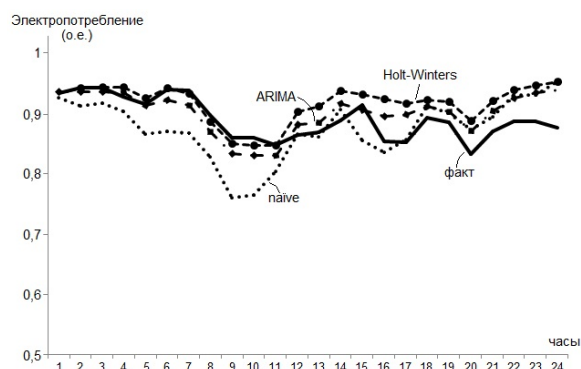


Рис. 4. Прогноз электропотребления ООО «Сорский ГОК» на 26 февраля 2009 г.

В сложившейся ситуации, когда нельзя выделить лучший метод, перспективным может быть применение для прогноза группы методов. В качестве результата можно использовать как среднее, так и взвешенное значение прогнозов, адаптивно подбирая веса для каждого метода. При этом менее точные методы вносят меньший вклад в общий прогноз. Прогноз, полученный группой методов, в среднем является более точным и надежным, чем прогноз, сделанный любым из методов группы в отдельности [6].

Результаты эксперимента показывают, что ошибка прогнозирования МАРЕ для различных дней изменяется в достаточно широком диапазоне. Такое изменение ошибки МАРЕ и высокая степень смещения прогнозов МРЕ свидетельствуют о том, что однофакторные методы прогнозирования не могут дать высокого качества прогнозирования для такого сложного объекта, как предприятие горной промышленности. Значительное изменение одного из влияющих факторов (горно-геологического, технологического, производственного или климатического) приводит к смещению прогноза, увеличению ошибки МАРЕ и, соответственно, к увеличению затрат на электроэнергию. Таким образом, однофакторные методы прогнозирования, такие как сезонная модель АРПСС и метод Хольта-Винтерса, могут применяться для прогнозирования сложного нестационарного процесса электропотребления горного предприятия, однако для более качествен-

ного прогнозирования необходимо использование многофакторных методов.

Перспективным направлением в области прогнозирования электропотребления является использование методов искусственного интеллекта, в первую очередь искусственных нейронных сетей и теории нечетких множеств. Модели на основе искусственных нейронных сетей и систем нечеткого вывода обладают рядом ценных свойств: способностью к аппроксимации сколь угодно сложных функций, воспроизведением сложных нелинейных зависимостей, способностью к обобщению, устойчивостью к помехам, отсутствием ограничений на характер входной информации и т. д. Однако использование любых достаточно сложных методов должно быть обосновано посредством анализа качества их прогнозирующих свойств по сравнению с более простыми методами. Одним из вариантов такого рода анализа является пассивный статистический эксперимент, рассмотренный в данной статье.

Библиографические ссылки

1. Шумилова Г. П., Готман Н. Э., Старцев Т. Б. Прогнозирование нагрузки ЭЭС на базе новых информационных технологий // Новые информационные технологии в задачах оперативного управления электроэнергетическими системами / Урал. отд-ние Рос. акад. наук. Екатеринбург, 2002. С. 127–156.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. В 2 т. М.: Мир, 1974.
3. Тихонов Э. Е. Методы прогнозирования в условиях рынка: учеб. пособие. Невинномысск, 2006.
4. Taylor J. W., de Menezes L. M., McSharry P. E. A comparison of univariate methods for forecasting electricity demand up to a day ahead // Intern. J. of Forecasting. 2006. Vol. 22. P. 1–16.
5. Гнатюк В. И. Закон оптимального построения технотензоров [Электронный ресурс]. URL: <http://gnatukvi.ru/ind.html> (дата обращения: 30.09.2010).
6. Ежов А. А., Шумский С. А. Нейрокомпьютеринг и его применение в экономике и бизнесе / под ред. проф. В. В. Харитоновой; Моск. инж.-физ. ин-т. М., 1998.

P. V. Val

SHORT-TERM FORECASTING OF MINING ENTERPRISE ENERGY DEMAND BY UNIVARIATE METHODS

In this paper the author performs analysis of short-term forecasting quality of mining enterprise energy demand, using the popular univariate forecasting methods (seasonal ARIMA, Holt-Winters method, and naive method).

Keywords: energy consumption, forecasting, ARIMA, Holt-Winters method, naive method.

© Валь П. В., 2011

УДК 62-506.1

П. В. Зеленков, Г. А. Сидорова

МОДИФИЦИРОВАННЫЙ АЛГОРИТМ HITS*

Показана проблема современных поисковых систем, связанная с ранжированием документов. Для решения данной проблемы в процессе поиска и обработки информации предлагается использовать модифицированный алгоритм HITS. Данный подход помогает решать проблемы поиска, определения релевантности найденной информации, а также производить ранжирование отклика системы.

Ключевые слова: HITS, ранжирование, обработка информации, поиск информации.

В настоящее время при создании и развитии технологий сбора и обработки информации основное внимание уделяется развитию существующих технологий, нацеленных на анализ баз данных поисковых сервисов сети Интернет, и развитию алгоритмов ранжирования [1; 2]. Однако если встает вопрос об организации подобных процедур в рамках локальных корпоративных систем, то возникает проблема в анализе информации и ее взаимосвязей на локальном уровне.

На сегодняшний день существует множество алгоритмов ранжирования информации в поисковых системах сети Интернет [3]. Один из самых распространенных – это алгоритм Клейнберга, для которого создано несколько модификаций. Наиболее значимым является метод HITS, который заключается в присвоении каждому документу в веб-множестве некоторых значений, которые называются весами документа. Существует два вида таких весов: a (authority) – вес авторитетного документа и h (hub) – вес хаб-документа. Авторитетный документ – это документ, соответствующий запросу пользователя, имеющий больший удельный вес среди документов данной тематики, т. е. большее число документов ссылается на данный документ. Хаб-документ – это документ, содержащий много ссылок на авторитетные документы. Соответственно, для каждой страницы рассчитывается не один, а два веса. Такой подход обусловлен наличием в Сети большого числа сообществ, т. е. наборов страниц близкой тематики, которые весьма сильно связаны друг с другом ссылками. Исходя из значе-

ний весов, происходит формирование множества поиска и его ранжирование по релевантности.

Такой подход очень удобен, так как позволяет находить больше документов, соответствующих заданной тематике. Однако у него есть и недостатки, которые естественным образом вытекают из достоинств: во множество найденных документов может попасть большое количество страниц с низким коэффициентом релевантности, которые, тем не менее, имеют много ссылок друг на друга, и именно им будут присвоены наивысшие ранги. Это явление называется смещением тематики (diffusion, drift). Обычно оно происходит в направлении более широкой предметной области (или лучше представленной в Сети). Для решения этой проблемы Клейнберг предложил использовать анализ содержимого страниц, но оценивать не отдельные страницы, а разные сообщества целиком.

Описание модифицированного метода. Модифицированный метод может быть полезен для поиска как в корпоративных информационно-управляющих системах, так и в локальных и глобальных сетях. Основа метода – избирательный поиск не по всему веб-пространству, а по документам, принадлежащим внутренней сети.

Его очень удобно использовать в организациях, специализирующихся на узкой тематике и имеющих обширную базу данных. Разработанная на базе этого метода поисковая система будет не только обрабатывать нужные документы, но и производить пополнение внутренней базы документами смежной тематики, найденными в Сети.

*Работа выполнена в рамках Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг.