Реляционные базы данных имеют неоспоримые преимущества, если хранимые данные табличного типа, если безопасность данных является приоритетом и предъявляются высокие требования по ACID.

Если данные плохо структурированы, и их объем потенциально существенно увеличится, следует использовать NoSQL базы данных, например, документо-ориентированную СУБД MongoDB, для которых появляются развитые средства визуализации, что повышает эффективность и качество разработки.

### БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ


1. MongoDB [Электронный ресурс]. URL: http://softtime.info/view/MongoDB (дата доступа: 15.02.2022).
2. Введение в JSON [Электронный ресурс]. URL: http://json.org/json-ru.html (дата доступа: 17.02.2022).
3. Королева Ю.А., Маслова В.О., Козлов В.К. Разработка концепции миграции данных между реляционными и нереляционными системами БД [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/razrabotka-kontseptsii-migratsii-dannyh-mezhdu-relyatsionnymi-i-nerelyatsionnymi-sistemami-bd/viewer (дата доступа: 05.11.2022).
4. NoSQL Database Planning & Modeling Tool Online |NoSQL DB Modeler, Entity Relationship Diagram [Электронный ресурс]. URL: https://nosqldbm.ru (дата доступа: 05.11.2022).
5. Руководство Swagger UI [Электронный ресурс]. URL: https://starkovden.github.io/swagger-ui-tutorial.html (дата доступа: 05.11.2022).


## МАТЕМАТИЧЕСКИЕ АСПЕКТЫ СОВРЕМЕННЫХ РЕЙТИНГОВЫХ СИСТЕМ

## MATHEMATICAL ASPECTS OF MODERN RATING SYSTEMS


***В.А. Нифагин***

***V.A. Nifagin***

Белорусский государственный университет
Минск, Беларусь
Belarusian State University
Minsk, Belarus
*e-mail: vladnifagin@bsu.by*



Разнообразные рейтинговые системы и интернет поисковики используют математические алгоритмы, включающие в качестве содержательного основания участника рейтинговой системы, т. е. объект, подлежащий оценке в рассматриваемой рейтинговой системе, оцифрованный параметр – ранг в виде неотрицательного числа, характеризующий один из аспектов (свойств, показателей полезности) участника рейтинговой системы, в заданном числовом диапазоне или их совокупность. А также целевую вектор-функцию (или функцию полезности, функцию предпочтения) – функционал $F$, в соответствии, со значением которого все участники рейтинговой системы упорядочиваются на основании учета значений всех их индикаторов.

A variety of rating systems and Internet search engines use mathematical algorithms including as a meaningful basis the participant for the rating system as a meaningful basis, i.e. a digitized parameter – rank in the form of a non-negative number characterizing one's aspects (properties, utility indicators) of the participant for the rating system, in a given numerical range or a range's set. Besides the target vector function (or utility function, preference function) is a functional $F$, in accordance with the value of which all participants are ordered on the indicator set.

*Ключевые слова:* рейтинг; поисковые системы, оцифровка показателей, нечеткая формализация; фунцинал полезности; унимодальные и полимодальные экстремальные задачи; нейронные сети.

*Keywords*: ranking; search engines; digitization of indicators; fuzzy formalization; utility functional; unimodal and polymodal extremal problems; neural networks.

Многие современные поисковые системы интернета содержат *PageRank* – один из алгоритмов ранжирования по внешним ссылкам [1]. Алгоритм применяется к коллекции документов, связанных гиперссылками, и назначает каждому из них некоторое численное значение (ранг), оценивающее его «значимость» или «авторитетность» среди остальных документов. Понятно, алгоритм может применяться не только к веб-страницам, но и к любому набору объектов, связанных между собой взаимными ссылками, т.е. к любому графу. Порядок ранжирования, например, в *Google* работает следующим образом. Сначала учитываются обычные текстовые критерии, такие как тэг *Title* и ключевые слова, а затем используется алгоритм *PageRank* для сортировки результатов, таким образом, чтобы более «важные» сайты получали наибольший ранг и поднимались вверх на странице результатов поиска. Главная задача заключается в том, чтобы найти критерий, выражающий важность страницы. В случае с *PageRank* таким критерием была выбрана теоретическая посещаемость страницы. Авторы исходили из того, что если страница $A$ ссылается на страницу $B$, то страница $A$ считает, что страница $B$ – «важная» страница. Логично предположить, что документ, имеющий большее число качественных внешних ссылок, содержит наиболее ценную информацию. Ранг $R$ страницы А рассчитывается по формуле

$$R(A) = (1-d) + d\sum_{k=1}^{n}(R(T_k)/N(T_k)) \qquad (1)$$

где $R(A)$ – ранг страницы $A$; $R(T_k)$ – ранг страницы $T_k$, которая ссылается на страницу А; $N(Tj)$ – количество исходящих ссылок на странице $A$; $d$ сглаживающий параметр, имеющий значение от 0 до 1 (т.н. коэффициент «затухания» передающегося ранга, $d \approx 0.85$). Модифицированная рейтинговая система *PageRank* включает оценку авторитетности, когда находятся несколько ресурсов с высоким уровнем доверия (seed pages) и с их учетом оцениваются остальные страницы. В новой системе применяется т.н. тематический *PageRank*, учитывающий только ссылки с тематически связанных страниц. Таким образом, новая формула расчета ранга

$$R'(p) = d\left(\sum_{q \in B_p} \left(R(q)/N_q + E(p)\right)\right) \qquad (2)$$

здесь $p$ – некоторый документ, $N_p$ – множество документов, на которые ссылается документ $p$; $B_p$ – множество документов, ссылающихся на $p$; $E(p)$ – параметр, предназначенный для документов и их групп, не имеющих общих связей с другими;

Простые рейтинговые системы [2] включают несколько цифровых индикаторов, определенных на компактных множествах и унимодальную целевую функцию ранга с ограничениями. Пусть $\mathbb{R}_n$ задана унимодальная функция $F(\cdot)$, и ее экстремум ищется на множестве $U$ $\mathbb{R}_n$ , в точках которого выполняются условия трех типов:

$$U = \{u \in \mathbb{R}_n \vee \mathbb{Z}_n : \ \varphi_j(u) = 0, j = 1, \dots, l, \ \psi_k(u) \ge 0, k = 1, \dots, s\} \qquad (3)$$

В БГУ для определения позиции в рейтинге сайтов факультетов использовалась логарифмическая нормировка каждому показателю, которые имели вид линейной формы нескольких числовых параметров с соответствующими весовыми коэффициентами (аналогичноWRWU).

$$R^a = \sum_{i=1}^{3} k_i \frac{\ln(P_i^a + 1)}{\ln(P_i^{\max} + 1)}$$

$k_1 = 0.5$ – *вебометрический весовой коэффициент*

$k_2 = 0.2$ – *наукометрический весовой коэффициент*

$k_3 = 0.2$ – *статистический весовой коэффициент*

Часто встречаются рейтинговые системы, у которых индикаторы $Y_i$ функции полезности $F$ сами являются результатом вычисления других функций, например, $G_i$ от совокупности других индикаторов, например, $Z_k$, что $Y_i = (G_i(Z_k(X_i)))$, $k = 1,\dots,n$, $j = 1,\dots,m$ (суперпозиция показателей). При этом индикаторы $Z_k$ раскрывают формирование индикатора $Y_i$. Обратная операция, когда несколько индикаторов $Y_i$ объединяются и преобразуются с помощью некоторой функции $S_l$, является процедурой агрегирования индикаторов. В случае, когда детализация (агрегирование) применяется не ко всем аргументам целевой функции, то ее результат решения экстремальной задачи может быть произвольным (за исключением линейных целевых функций), т.е. ранги после применения этих операций могут менять свою последовательность.

В продвинутых рейтинговых системах [3] формируется нечеткая реляционная модель, когда для каждого из $m$ объектов рейтинговой системы имеем вектор $n$ индикаторов (оценок) $Y_i = (y_1, y_2, \ldots, y_m)$ – или вектор альтернатив, которые подлежат оценке и ранжированию. Кроме того, $C = (c_1, c_2, \ldots, c_m)$ – вектор критериев, характеризующих эти альтернативы $y_i$ критерию $c_i$ представляются функцией принадлежности $f_{k_j}(y_i) \to [0, 1]$, т.е.:

$$f_{k_j}(y_i) : Y \times C \to [0, 1] \tag{4}$$

Целью сформулированной задачи является получение упорядоченного списка заключений каждого элемента экспертного множества, который получен по результатам выполненной оценки: $F : C \to Y'$, где $Y'$ – вектор участников, рассматриваемой задачи рейтинговых оценок.

Пусть $\{f_{k_{j1}}(y_i), f_{k_{j2}}(y_i), \ldots, f_{k_{jN}}(y_i)\}$, или $f_{kt}(y_i) = 1$, $j = 1, \ldots, m$, $t = 1, \ldots, T$. – функции, соответствующие альтернативам $y_i$ с частными критериями $c_{j1}, c_{j2}, \ldots, c_{jT}$ и $\{w_{j1}, w_{j2}, \ldots, w_{jT}\}$ – весовые коэффициенты, оценивающие относительную значимость частного критерия. Для любых $c_{j1}, c_{j2}, \ldots, c_{jt}$, принадлежащих к характеристикам критерия $C_j$, справедливо условие нормировки: $\sum_{t=1}^{T} w_{jt} = 1$. Постановка задачи теории нечетких множеств требует формализации частных критериев, что обуславливается необходимостью оперировать количественной и качественной информацией одновременно. Оцифровка качественной информации необходима из-за неопределенности в неполных и неточных данных. В качестве переменной модели используется лингвистическая переменная со значениями: несколько, мало, много, большинство и т.д. Лингвистическая переменная может выступать и в виде функции. Зададим вид функции принадлежности альтернативы $y_i$ к частным критериям $c_{j1}, c_{j2}, \ldots, c_{jT}$. Каждый элемент из $C_j$ имеет качественное описание в виде шкалы лингвистической переменной (нечеткое соответствие). Далее эксперты присваивают каждой градации индивидуальные оценки, которые принимают значения в интервале [0, 1]. Затем, интегрируя индивидуальные оценки, получают общую оценку градаций. Такой оценкой будет пересечение нечетких подмножеств индивидуальных оценок экспертов. Выявленное пересечение считается искомой количественной оценкой анализируемой ситуации. Заметим, что при сведении индивидуальных экспертных оценок в интегральную можно использовать правило пересечения нечетких множеств (выбирается минимальная из оценок функции принадлежности), а также применить некоторую новую операцию над такими множествами, в форме их согласованного выбора.

В этом случае интегральная оценка конструируется «наиболее удачным» экспертом. В каждой точке множества возможных альтернатив выбирается норма принадлежности к этой точке интегральной оценки. Норма, которая удалена от крайних значений оценок, и занимает некоторое «среднее» положение, будет считаться мерой оценки альтернатив.

*Целевая функция (функция полезности, функция предпочтения)* – правило $F$, по которому для объекта рейтинговой системы номер $i$ на основании учета значений всех $n$ его индикаторов приписывается некоторый вектор. Функция $F$ является вектор-функцией $n$ переменных (дискретных или непрерывных). Задача функции состоит в том, чтобы привести совокупность из $n$ индикаторов к одному (или нескольким) числам с целью их последующего сравнения с такими же числами, относящимися к другому объекту и ранжировать их по значимости.

В многокритериальной задаче с заданными критериями $F = (f_1, f_2,.., f_m)$, отвечающими конечному множеству возможных решений $Y$ оптимальным решением $Y^* \in Y$ является то, которое делает значение линейной свертки критериев $\sum_{i=1}^{m} \omega_i f_i\left(Y^*\right) = \max_{y \in Y} \sum_{i=1}^{m} \omega_i f_i\left(Y\right)$ максимальным. Компоненты вектора весов $(\omega_1, \omega_2, \ldots, \omega_m)^T > 0$ каждого объекта вычисляются как собственный вектор, который соответствует максимальному собственному значению $\lambda_{\max}$ матрицы $A$, где $A = (a_{ij})$ – матрица парных сравнений $a_{ij} = \dfrac{\omega_i}{\omega_j}$.

Наконец, восстребованными становятся методы, построения рейтинговых систем для решения актуальных задач на логических нейронных сетях. Рассмотрим пример построения типовой рейтинговой системы [4].

Пусть для данного типа объектов задан вектор индикаторов $(Y_i)$, $i = 1 ,\ldots, n$. Диапазон изменения каждого (вещественный) на основе экспертных оценок разбит, например, на $m$ промежутков, принадлежность которым говорит о неизменном рейтинге по данному индикатору интервале её значений:

$$Y_i \rightarrow \{[\, y_{i0}, y_{i1}), [\, y_{i1}, y_{i2}), [\, y_{i2}, y_{i3}), \ldots ,[\, y_{i(l-1)}, y_{il}),\}, \; i = 1 ,\ldots, n,$$

заметим, что $y_{ij}$ может быть равно 0 или $\infty$. Здесь учитывается, что промежуток $[y_{ik}, y_{i(k+1)})$ может соответствовать диапазону значений индикатора $Y_i$, промежуток $[y_{ir}, y_{i(r+1)})$ – временно допустимому значению этого индикатора а промежуток $[y_{is}, y_{i(s+1)})$ соответствует области значений индикатора, где измерения не проведены, либо их значения выше или ниже допустимых. Предположим, что для построенных промежутков, используемых совместно, как возможные ситуации на основе вектора индикаторов, эксперты установили значения рейтинга из некоторого нечеткого множества: рейтинг высокий, рейтинг средний, рейтинг низкий. То есть, любое возможное состояние объекта описывается тремя интервалами, каждый из которых соответствует своей качественной оценке, заданой одним из трёх значений рейтинга. Сформируем логическую нейронную рейтинговую сеть. Для этого предположим, что рецепторам соответствует $3n$ значений характеристик объектов системы, возбуждаемым извне, а нейронам – $3n$ вектор, возбуждение компонент, которого от связанных с ними рецепторов указывает на значение рейтинга. При этом возбуждение рецепторов интерпретируется как достоверность высказываний о принадлежности данных, а нейрон выполняет пороговую функцию активации. Рассмотрим пороговую функцию активации вида:

$$V_i = \begin{cases} \sum_{k=1}^{n} v_{jk},\ v_{jk} \geq h \\ 0,\ v_{jk} < h \end{cases}$$

где $V_i$ – возбуждение $i$-го нейрона $v_{jk}$, $k = 1,\ldots,n$; $j_k = 1, \ldots, 3n$; – возбуждения одного из 3 рецепторов, связанных с $i$-м нейроном. Порог $h$ выбирается экспериментально в процессе функционирования рейтинговой системы для ликвидациининия избыточных вычислений.

## БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ


1. Sarma A.D. Fast Distributed PageRank Computation // Theoretical Computer Science, Vol. 561. 2015, pp. 113-121.
2. Васильев Ф.П. Численные методы решения экстремальных задач. – М.: Наука, 1988.
3. Fitch Ratings: Definitions of ratings and other types of rating estimates. [Electronic resource.] URL: http://www.fitchratings.ru/dms/ fitch-russia/documents/fitch-ratings-definitions-ru/Ratings Definitions-ru.pdf (date of the access: June, 2022).
4. Бенджио И., Курвилль А., Гудфеллоу Я. Глубокое обучение. – М.: ДМК-Пресс, 2018.