

Раков Виктор Сергеевич

**ДИАЛОГОВЫЕ ГРАФИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ МНОГОКРИТЕРИАЛЬНОЙ
ОПТИМИЗАЦИИ**

Адрес статьи: www.gramota.net/materials/1/2011/9/11.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2011. № 9 (52). С. 37-41. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2011/9/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

$$\begin{aligned}
 TF-IDF(n_1, n_2) &= \sum_{t \in n_1 \cap n_2} V(t, n_1) V(t, n_2) \\
 V(t, n) &= \frac{V'(t, n)}{\sum_{t' \in n} V'(t', n)^2} \\
 V'(t, n) &= \log(TF(t, n) + 1) \log(IDF(t))
 \end{aligned}
 \tag{1}$$

В формуле, IDF делает вес слова обратно пропорциональным частоте появления в базе данных, а TF делает его пропорциональным частоте в записи. Таким образом, часто встречающимся словам назначаются небольшие значения веса, в то время как редко встречающимся - высокие. Базовая проблема поиска Top-K со степенью сходства TF-IDF была интенсивно изучена [3]. Таким образом, простым механизмом решения Batch-Top-K проблемы является запуск базового Top-K алгоритма для каждого сегмента во входном предложении. Однако, так как сегменты содержат совпадающие слова, можно достичь хороших результатов, группируя результаты вычисления.

Список литературы

1. Cafarella M. J., Etzioni O. A Search Engine for Natural Language Applications // WWW Conference. 2005. P. 442-452.
2. Chakrabarti S., Punera K., Subramanyam M. Accelerated Focused Crawling through Online Relevance Feedback // Ibidem. 2002.
3. Chaudhuri S., Ganjam K., Ganti V., Motwani R. Robust and Efficient Fuzzy Match for Online Data Cleaning // SIGMOD. 2003.
4. Cho J., Rajagopalan S. A Fast Regular Expression Indexing Engine // ICDE. 2002. P. 419-430.
5. DeRose P., Shen W., F. C. 0002, Lee Y., Burdick D., Doan A., Ramakrishnan R. DBLife: a Community Information Management Platform for the Database Research Community (Demo) // CIDR. 2007. P. 169-172.
6. Grishman R., Huttunen S., Yangarber R. Information Extraction for Enhanced Access to Disease Outbreak Reports // Journal of Biomedical Informatics. 2002. Vol. 35. P. 236-246.
7. <http://ru.wikipedia.org/wiki/TF-IDF>
8. Ipeirotis P. G., Agichtein E., Jain P., Gravano L. Towards a Query Optimizer for Text-Centric Tasks // ACM Transactions on Database Systems. 2007. Vol. 32.
9. Kim M.-S., Whang K.-Y., Lee J.-G., Lee M.-J. N-Gram/2l: a Space and Time Efficient Two-Level N-Gram Inverted Index Structure // VLDB'05: Proceedings of the 31st International Conference on Very Large Data Bases. 2005. P. 325-336.
10. Resnik P., Elkins A. The Linguist's Search Engine: an Overview (Demonstration) // ACL. 2005.

УДК 519.816

Виктор Сергеевич Раков
Братский государственный университет

ДИАЛОГОВЫЕ ГРАФИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ®

Решение многокритериальных задач представляется в виде множества допустимых решений, которое сводится к одному из его подмножеств, называемому подмножеством эффективных решений. Решение эффективно, если не существует такого же хорошего по всем критериям и строго лучшего хотя бы по одному из них. Множество допустимых эффективных решений может быть представлено ориентированным графом, в котором варианты решения вершины, а связи между вариантами - дуги графа. В таком многокритериальном графе, каждой вершине которого присвоена определенная векторная оценка (индикатор), может быть определен эффективный путь среди множества допустимых решений, при $C_1, C_2, \dots, C_i, \dots, C_k$ - различных критериях ($1 \leq i \leq k$) и Q_j индикаторах ($1 \leq j \leq m$).

Для многокритериальных задач типичным является определение такого эффективного пути подграфа, который состоял бы только из оптимальных индикаторов. Геометрическая интерпретация эффективного подграфа выражается траекторией, проходящей через все точки, отображающие оптимальные индикаторы, и характеризуется величиной

$$S_Q = \sum_{j=1} l_j(Q_j) \tag{1}$$

где $l_j(Q_j)$ - расстояние от начала координатной системы до j -й точки, отображающей Q_j индикатор. Алгоритм многокритериальной оптимизации включает:

- построение графа вариантов решений;
- разбиение графа на блоки;
- составление упорядоченного списка индикаторов;
- добавление фиктивного индикатора $Q_j = (O \dots O)T$, если имеется несколько начальных или конечных вершин;

- вычисление индикаторов путей;
- проведение анализа и выбора оптимальных индикаторов от нулевой ступени до конечной;
- определение эффективного пути или подграфа оптимальных индикаторов.

Для определения эффективного подграфа в многовариантной сетевой модели каждый индикатор Q_j представляется точкой на плоскости, например, в прямоугольных координатах t_j , C_j («продолжительность-стоимость») [5, с. 421]. Для решения задачи в диалоговом режиме на экран дисплея выводятся индикаторы, начиная с первой ступени. С помощью специальной программы определяют индикатор, удаленный на минимальное расстояние от начала координат. На экран выводят график гиперболы, проходящей через эту точку, анализируют полученное изображение, вычеркивают точки, расположенные правее от гиперболы, т.е. ограничивают общее число рассматриваемых вариантов. В дальнейших расчетах эти варианты и соответствующие им индикаторы не учитываются, а для последующего анализа на экране дисплея остаются выделенные индикаторы.

Рассматривая оставшиеся индикаторы, начиная с последней ступени, ведут поиск эффективного варианта сети. Интерактивная оптимизация позволяет значительно сократить число переборных всех вариантов, особенно в случае вероятностных сетей, где в качестве третьей характеристики может быть указана вероятность. При $k > 3$ следует воспользоваться алгоритмом объединения индикаторов, полученных из проекций. Предположим, что при проецировании $Q_j = (C_1, C_2, C_3, C_4)T$ на плоскость C_1, C_2, C_3 вычислены эффективные индикаторы всех путей, конечная вершина которых j имеет $C_1=2, C_2=3, C_3=4$, а при проецировании на плоскость C_2, C_3, C_4 получили индикаторы $C_2=1, C_3=4, C_4=5$. Очевидно, что индикатор $(2,1,4,5)T$ не обязательно будет оптимальным, однако путем многократного проецирования на различные плоскости можно сформировать достаточно приемлемый индикатор даже при $k > 3$.

Рассмотрим несколько графических способов преобразования двух критериев, например, социальной и экономической эффективности ($CЭ$ и $ЭЭ$) в общую эффективность $Э$. При способе прямых

$$Э = (CЭ + ЭЭ) / 2, \quad (2)$$

при способе дуг окружности.

$$Э = \sqrt{[(CЭ - 1)^2 - (ЭЭ - 1)^2] / 2} \quad (3)$$

Более эффективное преобразование получается при способе гипербол, где

$$Э = \sqrt{CЭ * ЭЭ} \quad (4)$$

При равных значениях общей эффективности лучшее решение то, которое отстоит от диагональной линии (оптимальной линии развития), соединяющей начало координат с точкой с координатами (I, I) на наименьшее расстояние. Если решения получаются равноценными, то заранее устанавливают, что лучшим считается решение с большим $CЭ$ или с большим $ЭЭ$. Если же решения получают равную оценку, то целесообразно их выводить на графический дисплей для дальнейших расчетов. В [1, с. 220] подробно рассматривается алгоритм оценки решений (Рис. 1) и приводятся схемы построения ограничений области решений на графиках общей эффективности. Алгоритм оценки разработан таким образом, что компьютер рассчитывает и сравнивает лишь те решения, которые лежат в определенной зоне поиска. В качестве принципов ограничений предлагается использовать наименьшую эффективность и наименьшую выравненность.

При этом если задано некоторое значение наименьшей общей эффективности, то компьютер «забывает» все решения, у которых общая эффективность будет ниже заданной. При введении критерия параллельных прямых решения ищутся лишь в полосе, прилегающей к линии оптимального развития. Невыравненные решения (т.е. решения с резко различными $CЭ$ и $ЭЭ$) не учитываются. При введении углового критерия исключаются решения с малыми значениями $CЭ$ и $ЭЭ$. На практике следует также использовать комбинации ограничений. Если при первом поиске в области, установленной с принятыми ограничениями, нет ни одного решения, то область ограничений постепенно расширяют.

В многокритериальной оптимизации весьма актуальна разработка таких моделей-индикаторов, которые обеспечивают наибольшее приближение человека к естественным (трехмерным) условиям восприятия зрительных образов, характеризующих пространственно-временную взаимосвязь процессов и явлений [3, с. 241]. В зависимости от числа отображаемых параметров множество процессов предлагается представить в виде множества непересекающихся параллелепипедов или прямоугольников-портретов (Рис. 2). В отличие от прямой линии в параллелепипеде значения параметров изображаемых процессов определяются тремя размерами: двух сторон основания и высоты (Рис. 3). Трехмерный портрет на плоскости представляется посредством трех двумерных портретов, которые аналогичны знаку ВАРЗАРА [1, с. 107], а N -мерный - с помощью M проекций на трехмерное пространство, где $M = C_3 N$. Если высоты прямоугольников на двумерном портрете приравнять нулю, то получается традиционный линейный график. С помощью портрета представляются также сетевые модели различных типов: «работы-дуги», «работы-вершины», обобщенные.

К основным достоинствам портретной сетевой модели следует отнести:

- возможность отображать сетевую модель в различных системах координат, например, «интенсивность потребления ресурсов время»;
- возможность отображения очередности выполнения работ, критического пути, распределения интенсивностей потребления ресурсов, продолжительностей и объемов работ, сроков их начал и окончаний, резервов времени и т.д.;
- возможность отображения событий работ (сторон прямоугольников, перпендикулярных временной оси);
- отсутствие необходимости изображения фиктивных работ (с нулевой продолжительностью);

- облегчение анализа сетевой модели при оценке объемов работ и сопоставлении их между собой площадями прямоугольников.

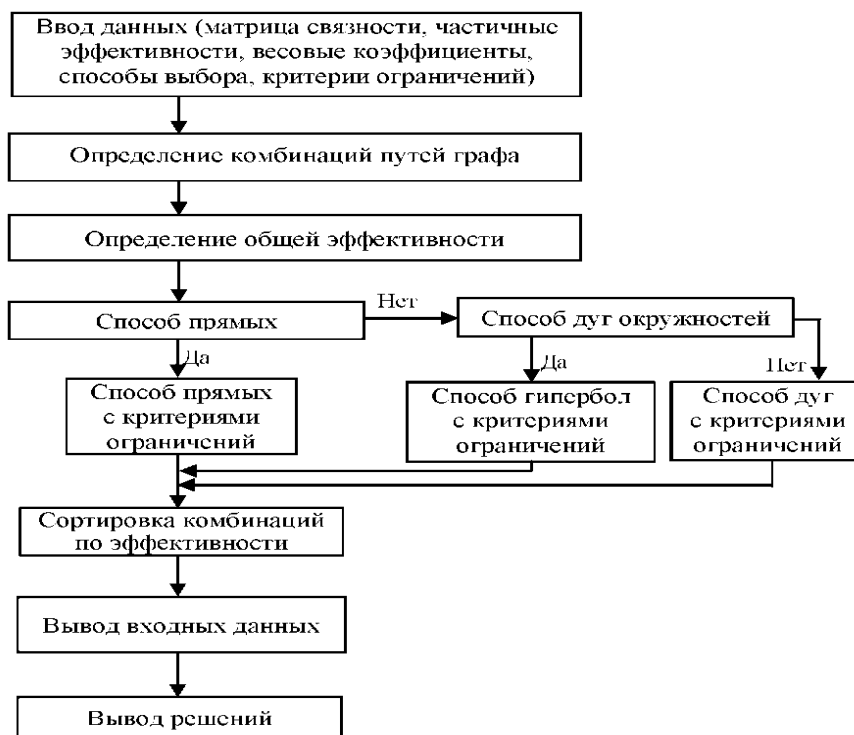


Рис. 1. Алгоритм оценки решений

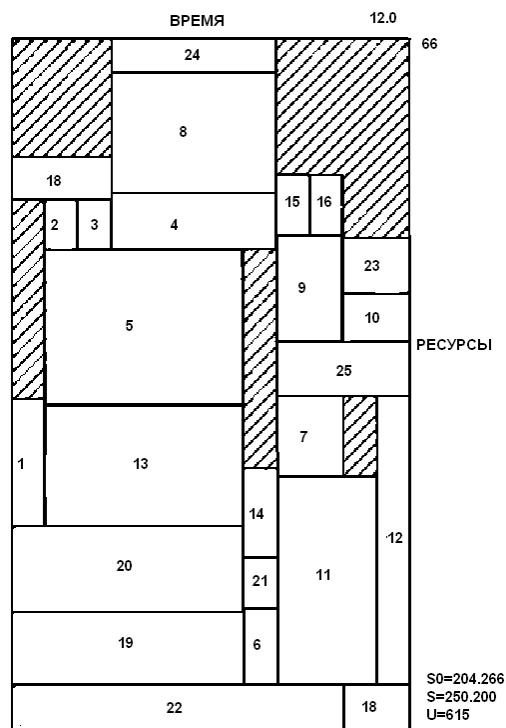


Рис. 2. Портретная модель

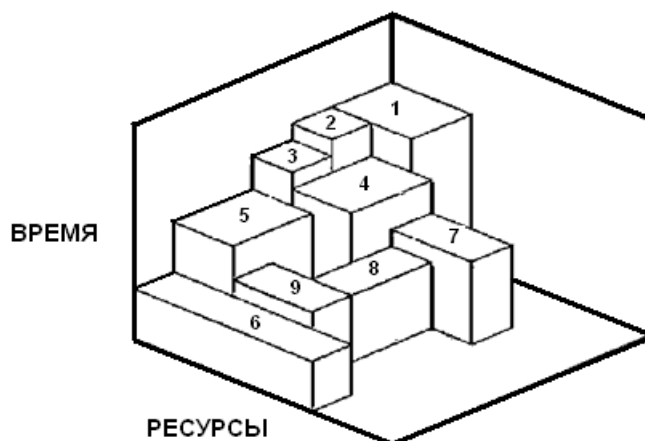


Рис. 3. Трехмерная портретная модель

Если представить N-мерный портрет посредством проекции его в трехмерное пространство «фронт работы - виды работ и исполнители (бригады, звенья) - время», то получим модель календарного плана, которая в модифицированной форме нашла применение на объектах гражданского, жилищного, сельского и промышленного строительства [2, с. 488].

Синтезирующий характер портретной формы делает индикатор особенно эффективным для применения в задачах оптимизации календарных планов и программ, контроля за ходом их выполнения. С помощью портрета получены геометрические образы оптимальных календарных планов и исследованы задачи и принципы оптимизации управления. Если отобразить оптимальный календарный план, например, по стоимости на плоскости координат «производная стоимостных затрат по времени - время», то получим портрет, в котором отсутствуют свободные (заштрихованные) участки между прямоугольниками.

Необходимые и достаточные условия оптимальности для комплексов работ (у которых функции стоимостных затрат от их продолжительностей - нелинейные) выводятся другим путем, отличным от аналитического способа, менее сложным и более наглядным. При этом получены уравнения, описывающие управление ходом работ при оптимальных стоимостных затратах. Целесообразно при решении оптимизационных задач использовать в качестве критерия, определяющего степень приближения к оптимальному или рациональному варианту, безразмерную величину U как отношение суммы площадей всех прямоугольников к площади конверта (обрамляющего портрет прямоугольника). Для оптимального безрезервного плана $U=1$, а стоимостная функция $C=(C^*/T)\tau$, $0 \leq \tau \leq T$, (5)

где T - продолжительность выполнения плана, а C^* - стоимостные затраты всех работ плана. Зависимость $C=C(\tau)$ в неявной форме отражает влияние условий выполнения работ, применяемой технологии, используемых ресурсов на продолжительность и на стоимость комплексов работ.

Величина $V=dC/d\tau$ показывает скорость изменения затрат в данный момент, которая определяется технологическими факторами, интенсивностью потребления различных ресурсов, а также влиянием случайных факторов (сбоями в снабжении, поломками машин, природными условиями и т.п.). Скорость изменения затрат в значительной мере связана с ходом выполнения работ. Для оптимального плана можно принять $V=const$. Однако в ходе работ из-за случайных факторов это условие, как правило, нарушается. Оптимальное, в смысле стоимостных затрат, управление ходом работ в классе гладких функций $C=C(\tau)$ обеспечивает минимизацию интеграла цены управления

$$I = \int_0^T \left(\frac{dC}{d\tau} \right)^2 d\tau \quad (6)$$

при выполнении граничных условий решения задачи: $C(0)=0$; $V(0)=0$; $C(T)=C^*$; $V(T)=0$; $a(\tau)=dV/d\tau=0$. Граничные условия совпадают с реальной ситуацией: в начальный момент $t=0$ величина стоимостных затрат $C=0$. В момент $\tau=T$ стоимостные затраты достигают максимальной величины $C=C^*$. Выявлены следующие закономерности в характере изменения функции скорости изменения затрат: положительная асимметрия, т.е. максимум кривой смещен влево относительно линии, делящей площадь под кривой на две равные части; траектория скорости круто поднимается при удалении от начала и полого спускается при приближении к окончанию выполнения работ [4, с. 193].

Рассматривая кривую оптимальных затрат, нетрудно обнаружить следующую закономерность: при приближении к точкам экстремума скорость изменения функции падает. Это означает, что при движении социально-экономических систем для того, чтобы минимизировать затраты (ресурсов природы и общества), необходимо на определенных этапах движения периодически ускорять изменения функции затрат. Переходить же из одного этапа в другой можно лишь тогда, когда будут исчерпаны все возможности предыдущего.

На практике удобно строить как оптимальную кривую затрат, так и наихудшую кривую (из данных статистического анализа).

В результате получается так называемая «банановая кривая». С учетом «банановой кривой» целесообразно составлять прогнозы затрат. Исследования выявили аналогию, которая существует между уравнениями цены оптимального управления и вариационного принципа наименьшего действия. Кривая $C=C(\tau)$ оптимального плана по стоимостным затратам получает интересную механическую интерпретацию.

Она идентична траектории движения материальной точки, оптимальной по энергозатратам при заданных начальных и конечных условиях.

С учетом вышеизложенного, при нарушениях производственного процесса предлагается изменить и оптимизировать ход оставшихся работ с помощью индикаторов скоростей. Управленческие воздействия должны быть такими, чтобы фактическая скорость изменения затрат была бы по возможности близкой к некоторой желаемой (плановой) скорости. Основным критерием будет служить правило: чем ближе кривая к заданной, тем лучше. Можно количественно определить критерий расстояния между двумя кривыми по максимальному расхождению, по методу наименьших квадратов, однако количественный критерий не может заменить визуальную (графическую) оценку соответствия двух кривых. Пусть отклонение скорости $V_f(t)$ от плановой $V(t)$ на отрезке времени $[0, T]$ измеряется выражением

$$I_1 = \int_0^T (F(V_f(t) - V(t))) dt \quad (7)$$

где F - заданная функция, а цена управления

$$I_2 = \int_0^T f^2 dt \quad (8)$$

где f - управляющие воздействия. Для минимизации общей цены управления $I_0 = I_1 + I_2$ рассматривается дискретный вариант процесса оптимального управления и предлагается соответствующий диалоговый алгоритм. С помощью портрета удобно исследовать графические интерпретации принципов устойчивости (стабильности) состояния сложных систем [5, с. 220].

Плотность «упаковки» трехмерного или N -мерного портрета, состоящего из различных видов работ, достигает максимума в состоянии устойчивого равновесия. Получается геометрический аналог принципа плотной упаковки, когда в портретном (трехмерном или N -мерном) образе отсутствуют заштрихованные (пустые) участки.

Обычно говорят об устойчивости того или иного состояния системы.

Можно считать, что система устойчива, если траектория ее модели в пространстве не будет выходить за пределы заданной ограниченной области при некоторых возмущениях достаточно широкого спектра. При любом отклонении от устойчивого равновесия затраты на восстановление равновесия всегда возрастают. Это наглядно прослеживается также на портретных моделях. Анализ устойчивости модели позволяет формулировать различные гипотезы о закономерностях их функционирования, наличие либо отсутствие которых в реальности дает основание судить, в том числе и об адекватности модели. В этой связи выявлены взаимосвязи показателя U с известными показателями оценки уровня организации производства. Каждому его значению соответствуют определенные значения непрерывности, равномерности, совмещения, ритмичности и интенсивности производства работ, которые также получают наглядную интерпретацию на портретной модели.

Исследование различных моделей показало, что наибольшей устойчивостью, характеризуются модели, портреты которых имеют квадратную форму [3, с. 114].

Список литературы

1. Вагнер Г. Основы исследования операций. М.: Мир, 1972. Т. 1-4.
2. Воронин А. А., Мишин С. П. Оптимальные иерархические структуры. М.: ИПУ РАН, 2003. 210 с.
3. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. Лекции по теории графов. М.: Наука, 1990. 384 с.
4. Колосова Е. В., Новиков Д. А., Цветков А. В. Методика освоенного объема в оперативном управлении проектами. М.: Апостроф, 2001. 156 с.
5. Коргин Н. А. Механизмы обмена в активных системах. М.: ИПУ РАН, 2003.