ЗНАНИЕОРИЕНТИРОВАННЫЕ МОДЕЛИ ПРИНЯТИЯ ОПТИМАЛЬНЫХ РЕШЕНИЙ

Козлова М. Г., ассистент

В статье рассмотрена модель и разработан алгоритм принятия оптимального решения на основе имеющейся начальной информации в виде знаний. В основе подхода лежит каноническое представление задач псевдобулевой оптимизации в форме с дизъюнктивным ограничением.

Развитие информатики, важность проблемы интеллектуализации программного обеспечения, растущий спрос на новые информационные технологии – определяют актуальность разработки моделей принятия решений, практически пригодных для использования в сферах производства и управления. Важнейшим направлением современной теоретической и прикладной информатики является разработка и исследование моделей принятия решений на основе знаний как специальной формы представления информации. Знаниеориентированные (knowledge-based) модели привлекают математиков-прикладников как объект теоретических исследований, направленных на изучение полноты представления, адекватности, точности, и имеют практическое значение как основа принятия решений в экспертных системах и системах поддержки принятия решений. Они востребованы разработчиками экспертных систем и в значительной степени определяют прогресс в области искусственного интеллекта [1].

Теоретическое значение моделей принятия решений при неполной информации состоит в расширении методов прикладной математики, в частности, развитии методов оптимизации, применимых для широкого класса информационных систем с повышенным уровнем интеллектуализации.

Под принятием решений обычно понимается выбор наиболее предпочтительного варианта достижения поставленной цели из некоторого множества допустимых альтернатив. Такой выбор традиционно рассматривается как одна из составляющих кибернетического подхода и практически воплощается в большинстве современных интеллектуализированных информационных систем [2].

В существующих интеллектуальных системах для формализации задачи представления знаний используются фреймовые, реляционные, вычислительные, графовые, продукционные, логические модели и семантические сети. Тем не менее проблема представления знаний и их использования для решения практических задач далека от полного разрешения.

При рассмотрении этой проблемы выделим два аспекта: собственно содержание знаний и форму их представления. Обладая свойством универсальности по отношению к содержанию знаний, указанные модели не являются таковыми относительно формы представления знаний. Это обстоятельство усложняет процесс внесения знаний, поскольку требуется преобразование представления знаний из естественного для той или иной предметной области в допускаемое конкретной моделью.

В своей тьюринговской лекции [3] профессор Е.Фейгенбаум отметил особенную важность представления знаний в форме правил и широкое распространение и полезность знаниеориентированных технологий для использования при проектировании интеллектуализированных информационных систем. Практическое использование продукционных систем, баз знаний (БЗ), определило важность разработки знаниеориентированных оптимизационных моделей выбора решений, существенно более сложных по сравнению с моделями вычисления свойств (вывода предикатов).

Сложность нахождения элементов знаний, практического заполнения БЗ выдвинула проблему извлечения знаний. Наиболее приемлемым оказался подход к извлечению знаний из эмпирики – путем анализа и обобщения прецедентов [4]. Последний подход, реализующий принцип эмпирической индукции, дополняет и обогащает дедуктивный подход на основе правил.

Целью настоящей работы является разработка модели и алгоритма принятия оптимального решения на основе имеющейся начальной информации в виде знаний (и, возможно, дополнительной информации в виде прецедентов). В основе подхода лежит каноническое представление задач псевдобулевой оптимизации в форме с дизъюнктивным ограничением (ПБО с ДНФО) [5]. При использовании начальной информации в виде продукционных знаний используется строгая формализация понятий «логическая продукция» и «логическая продукционная система», введенных в работе [6]. Все необходимые для понимания излагаемого подхода определения можно найти в работах [5-7]. Дополнительный материал можно найти в обзорах [1, 8]. Излагаемые далее результаты частично анонсировались в работах автора [1,2,11].

Определим множество $B^n = \{0,1\}^n$ – единичный n-мерный куб. Функции вида $F: B^n \to \mathbf{R}$ называются псевдобулевыми. Класс таких функций обозначают $PS_2(n)$.

В общем случае однокритериальные псевдобулевые модели имеют вид:

$$extr\, F(\tilde{x}),\quad \tilde{x} \in \Omega \subseteq B^n,\quad F(\tilde{x}) \in PS_2(n), \tag{1}$$

где Ω – область допустимых решений. Если ввести характеристическую функцию

$$f_\Omega(\tilde{x}) = \begin{cases} 1, & \tilde{x} \in \Omega, \\ 0, & \tilde{x} \in B^n \setminus \Omega, \end{cases}$$

то задачу (1) можно представить в виде:

$$extr\, F(\tilde{x}),\quad f_\Omega(\tilde{x}) = 1,\quad f_\Omega(\tilde{x}) \in P_2(n),\quad \tilde{x} \in B^n,\quad F(\tilde{x}) \in PS_2(n). \tag{2}$$

Рассмотрим произвольную дизъюнктивную нормальную форму (ДНФ) для функции $f_\Omega(\tilde{x})$:

$$D_{f_\Omega}(\tilde{x}) = \bigvee_{j=1}^{m} K_j,\ \text{где}\ K_j = x_{i_1}^{\sigma_{i_1}} \& \ldots \& x_{i_{r_j}}^{\sigma_{i_{r_j}}},\quad x_{i_k}, \sigma_{i_k} \in \{0,1\},\ k = \overline{1, r_j},\ j = \overline{1, m}.$$

Тогда задача, эквивалентная задачам (1) и (2), имеет вид:

$$extr\, F(\tilde{x}),\quad D_{f_\Omega}(\tilde{x}) = 1,\quad \tilde{x} \in B^n,\quad F(\tilde{x}) \in PS_2(n). \tag{3}$$

Задача (3) называется задачей псевдобулевой оптимизации с дизъюнктивным ограничением. Задачи псевдобулевой оптимизации с дизъюнктивным ограничением возникли при использовании кибернетических методов решения задач синтеза дискретных решающих алгоритмов.

Синтез ДНФ $D_{f_\Omega}(\tilde{x})$ более сложен, чем решение задачи (3), так как этот вопрос тесно связан с проблемой минимизации дизъюнктивных нормальных форм.

Будем опираться на следующий фундаментальный факт.

Любая задача скалярной псевдобулевой условной оптимизации может быть представлена в канонической форме с дизъюнктивным ограничением [5]:

$$extr\,F(\tilde{x}), \quad \bigvee_{j=1}^{m} x_{j_1}^{\sigma_{j1}} \& \ldots \& x_{j_{r_j}}^{\sigma_{jr_j}} = 1 . \qquad (4)$$

Заметим, что если существует дополнительная информация о значении экстремума целевой функции, то задача (4) сводится к исследованию множества, заданного дизъюнктивным ограничением, которое содержит дизъюнкцию, соответствующую целевой функции.

Для определения модели (4) функция F и ДНФ-ограничение должны быть заданы полностью. В реальных моделях, как правило, дана только часть информации на множестве вариантов или правил и, следовательно, возникает задача с неполной начальной информацией.

Класс задач псевдобулевой оптимизации, представимых в форме с ДНФ-ограничением, включает такие задачи, как задача о ранце, задача коммивояжера и другие известные NP-полные задачи.

Каноническая модель является исчерпывающей в своем классе в силу полноты, устанавливаемой моделью (4). Левая часть единственного ограничения в (4) является дизъюнктивной нормальной формой характеристической функции множества Ω. Существование логической системы продукций (ЛСП), позволяющей выводить целевой факт g_0 ="$\tilde{x}$ – допустимое решение ", становится очевидным при рассмотрении рис. 1.

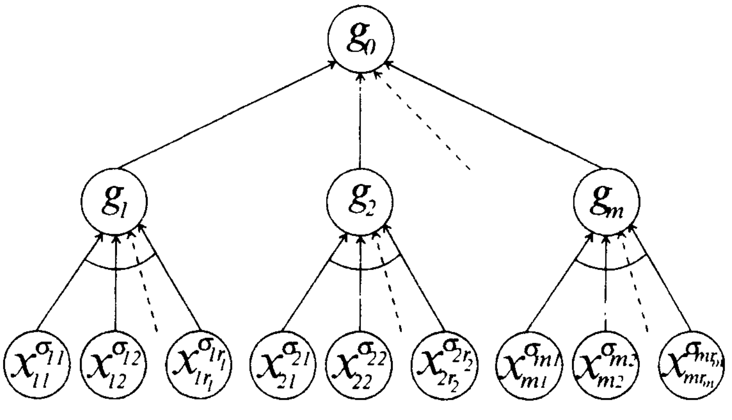

Рис.1. И/ИЛИ граф ДНФ-ограничения канонической модели.

Соответствующая графу на рис.1 ЛСП имеет вид:

$$\begin{cases} x_{j_1}^{\sigma_{j1}}, \dots, x_{j_{r_j}}^{\sigma_{jr_j}} \to g_j; \\ g_j \to g_o;\ j = \overline{1,m}. \end{cases}$$

И/ИЛИ граф ограничения канонической модели является трехъярусным. Легко показать, что любой граф ЛСП, не имеющий циклов, может быть сведен к трехъярусному и представлен в виде ДНФ.

Теперь можно сформулировать требование, предъявляемое к БЗ системы принятия решений. Если решения должны удовлетворять некоторым ограничениям, то ЛСП должна обеспечивать возможность *вывода целевых* предикатов, соответствующих этим ограничениям. Группе ограничений, которые должны выполняться одновременно, соответствует вершина типа «И». Вопрос полноты знаний об ограничениях в БЗ является важным, но он изучается отдельно в теории знаниеориентированных систем. Отметим, что ДНФ ограничение может быть получено путем обучения по эмпирической (прецедентной) информации [8], а также при помощи дуального [7] и других гибридных подходов.

Для синтеза ДНФ по заданным ЛСП были разработаны *D*- и *DS*-алгоритмы [6], реализующие, соответственно, стратегии «сверху вниз» и «снизу вверх». Если знаниеориентированный подход к построению ограничений в моделях принятия решений достаточно легко объясняется в терминах продукций, то при формировании знаний о целевой функции и их извлечении при синтезе модели ситуация сложнее.

Поскольку любую задачу псевдобулевой оптимизации можно представить в эквивалентной форме с ДНФ-ограничением, то это справедливо для задач с линейной целевой функцией. Следовательно, решение любой линейной задачи теоретически можно осуществлять по такой схеме:

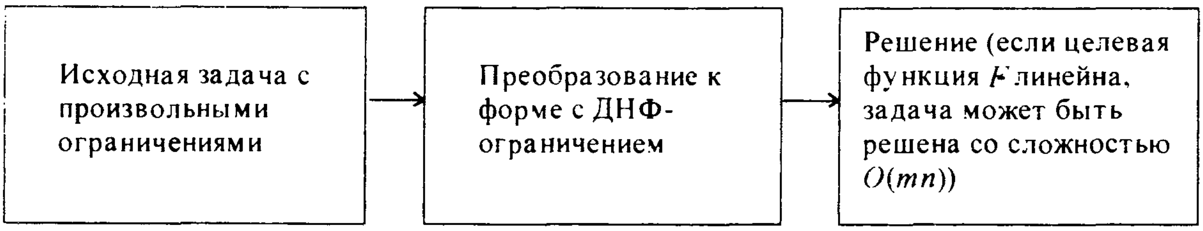

Рис.2. Решение задачи псевдобулевой оптимизации, приведенной к форме с ДНФ-ограничением.

Как правило, задачи псевдобулевой оптимизации не приведены к форме с дизъюнктивным ограничением. Они представляются в такой форме, когда используются кибернетические методы синтеза дискретных моделей принятия решений, основанные на И/ИЛИ правилах.

Схему решения знаниеориентированных задач с неполной начальной информацией можно представить в виде

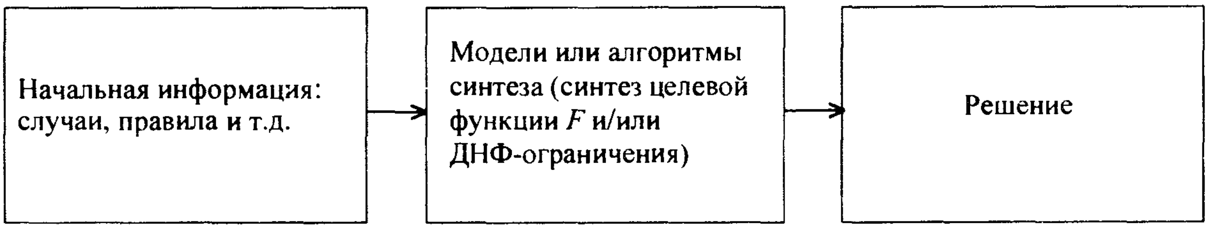

Рис.3. Решение задачи с неполной начальной информацией.

Две схемы, приведенные на рис. 2, 3, имеют следующее общее свойство: второй этап в каждой из трехэтапных схем имеет, как правило, неполиномиальную сложность.

Полученное в работе [5] утверждение: каждая псевдобулева функция $F(\tilde{x})$ может быть представлена в виде полинома над полем вещественных чисел:

$$F(\tilde{x}) = \sum_{j=1}^{k_j} c_j x_{j_1} \cdot \ldots \cdot x_{j_{r_j}} + c_0, \quad c_j \in R, \quad j = \overline{0, k_j}\,.$$

показывает, что любая целевая функция в канонической модели имеет не более чем полиномиальную сложность, и это сразу же подсказывает, какие знания об этой функции можно использовать. Например, полезна такая информация:

1) линейность (нелинейность);

2) знаки коэффициентов целевой функции (положительность, отрицательность или равенство нулю, указывающее на невхождение переменной в целевую функцию);

3) значения функции в некоторых точках (прецеденты).

В случае наличия такой информации ее можно считать знаниями, имеющими вид фактов – изначально истинных предикатов: «F – линейна» или «$F(\tilde{x}) = c_1x_1 + \ldots + c_nx_n$», «$c_i < 0$», «$c_k = 0$» и т.д. Знания о знаках, с точки зрения экспертов, вполне естественны и часто известны: эксперт может оценить ситуацию, если предикат x_i выполнится.

Наиболее продуктивным является использование знаний для канонической модели, то есть модели в виде задачи оптимизации линейной функции в форме с дизъюнктивным ограничением. Такой подход не сужает проблему в силу следующего факта: для любой нелинейной задачи оптимизации функции $F \in PS_2(n)$ найдется эквивалентная ей задача оптимизации линейной функции $G \in PS_2(\mu)$ в форме с дизъюнктивным ограничением, где μ – число слагаемых приведенного полинома для функции F без учета свободного члена [5].

При таком подходе даже небольшое количество знаний о целевой функции канонической модели в некоторых случаях позволяет получить точное решение. В [8] описан подход к синтезу решений при наличии значений линейной целевой функции в некоторых точках, основанный на построении выпуклого многогранного конуса, заведомо содержащего неизвестный вектор коэффициентов. Ниже представлен метод, использующий минимальное число начальных данных.

Рассмотрим алгоритм принятия решения на основе начальной информации в виде знаний.

Будем считать, что информация о целевой функции F в канонической модели (4) представлена следующими *достоверными* знаниями: F – линейная функция, то есть имеет вид $F(\tilde{x}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n + c_0$ и для каждого $i = \overline{1,n}$ является истинным только один из трех предикатов: «$c_i > 0$», «$c_i < 0$», «$c_i = 0$». Не теряя общности, будем считать, что отыскивается максимум функции F.

$$\max \sum_{i=1}^{n} c_i x_i, \quad \bigvee_{j=1}^{m} x_{i_1}^{\sigma_{i_1}} \cdot \ldots \cdot x_{i_r}^{\sigma_{i_r}} = 1, \quad \tilde{x} \in B^n. \tag{5}$$

Обозначим $N_j = \{\tilde{x} : x_{j_1} = \sigma_{j_1}, \ldots, x_{j_r} = \sigma_{j_r}\}$ – интервал, состоящий из всех булевых наборов, на которых j-я конъюнкция ДНФ-ограничения канонической модели обращается в единицу. Область допустимых решений $\Omega = \bigcup_{j=1}^{m} N_j$. Тогда задача (5) может быть представлена в эквивалентной форме

$$\max \sum_{i=1}^{n} c_i x_i, \quad \tilde{x} \in \bigcup_{j=1}^{m} N_j.$$

Экстремальное решение задачи в каждом интервале ($\tilde{x} \in N_j \subset \Omega$) достигается в любой точке $\tilde{\alpha}_j^* = (\alpha_{j1}^*, \ldots, \alpha_{jn}^*) \in N_j$, удовлетворяющей следующему условию для каждого $i = \overline{1,n}$:

$$\alpha_{ji}^* = \begin{cases} \sigma_i, & i \in I_j; \\ 1, & (i \notin I_j) \wedge (c_i > 0); \\ 0, & (i \notin I_j) \wedge (c_i < 0); \\ \delta, & (i \notin I_j) \wedge (c_i = 0). \end{cases} \tag{6}$$

где $\delta \in \{0,1\}$ – произвольное значение, $I_j = \{j_1, \ldots, j_{r_j}\}$.

Определение 1. Точка $\tilde{\beta}$ мажорирует точку $\tilde{\alpha}$ по критерию F (обозначение – $\tilde{\beta} \overset{F}{\succ} \tilde{\alpha}$) относительно начальной информации о знаках коэффициентов линейной функции F, если для любого $i = \overline{1,n}$ выполняется условие

$$[(\beta_i \geq \alpha_i) \& (c_i > 0)] \vee [(\beta_i \leq \alpha_i) \& (c_i < 0)] \vee (c_i = 0)$$

и одновременно найдется такое i, что $[(\beta_i > \alpha_i) \& (c_i > 0)] \vee [(\beta_i < \alpha_i) \& (c_i < 0)]$.

Предположим, что найдены множества экстремальных точек в каждом интервале. Обозначим их $A_1^*,\dots,A_j^*,\dots,A_m^*$. Любая точка $\tilde{\alpha}_j^* \in A_j^*$ удовлетворяет условию (6). В каждом из множеств A_j^* все точки по критерию F равноценны (значения $F(\tilde{\alpha}_j^*)$ одинаковы для всех $\tilde{\alpha}_j^* \in A_j^*$).

Определение 2. Подмножество допустимых решений называется À-паретовским, если оно получено из множества À=$\{A_1^* \bigcup \dots \bigcup A_m^*\}$ путем удаления из него всех точек, которые мажорируются хотя бы одной другой точкой из À.

À -паретовское множество легко вычисляется путем попарных сравнений точек из À с учетом очевидной транзитивности отношения "$\overset{F}{\succ}$". Будем обозначать À-паретовское множество П(À).

Утверждение 1. Если À -паретовское множество полностью содержится в некотором интервале N_j, соответствующем j-ой конъюнкции достоверного ДНФ-ограничения, то входящие в него точки исчерпывают все экстремальные решения задачи (4), даже при неполностью заданной начальной информации.

Действительно, пусть П(À)$\subseteq N_j$. Тогда из условия (6) следует, что значения целевой функции F в каждой паре точек из П(À) одинаковы, так как одинаковы значения функции F в каждой паре точек из N_j. Согласно включению П(À)$\subseteq N_j$, точки из множества A_j^* в совокупности мажорируют все остальные точки экстремальных решений в интервалах множества

$$\Omega = \bigcup_{j=1}^{m} N_j \ .$$

Каждая мажорируемая точка, согласно определению 1, доставляет меньшее значение целевой функции по сравнению с мажорирующей ее точкой, что и доказывает утверждение .

Пример 1. Дана задача:

$$\max(c_1x_1 + c_2x_2 + c_3x_3);\quad c_1 > 0;\quad c_2 < 0;\quad c_3 > 0;\quad x_1\bar{x}_2 \vee x_1\bar{x}_3 = 1.$$

Нулевых коэффициентов c_i нет, поэтому множества A_1^*, A_2^* содержат по одной точке: $\tilde{\alpha}_1^* = (1,0,1)$; $\tilde{\alpha}_2^* = (1,0,0)$; $\tilde{\alpha}_1^* \overset{F}{\succ} \tilde{\alpha}_2^*$. Экстремальное решение достигается в точке $\tilde{\alpha}_1^*$, П(А)= $\{\tilde{\alpha}_1^*\} \subset N_{x_1x_2}$.

Условие утверждения 1 является достаточным для точного нахождения решения задачи (4) с указанной выше начальной информацией (знаниями) о целевой функции. Но оно, вообще говоря, может не выполняться.

Пример 2. Дана задача:

$\max(c_1x_1 + c_2x_2 + c_3x_3)$; $c_1 > 0$; $c_2 < 0$; $c_3 > 0$; $x_2 \vee \bar{x}_1 x_3 = 1$, $\tilde{\alpha}_1^* = (1,1,1)$; $\tilde{\alpha}_2^* = (0,0,1)$;

$\Pi(A) = \{\tilde{\alpha}_1^* \bigcup \tilde{\alpha}_2^*\}$ $\tilde{\alpha}_1^* \in N_{x_2}$; $\tilde{\alpha}_2^* \in N_{x_1 x_3}$.

На основе начальной информации предпочесть одну из двух точек из А-паретовского множества невозможно, потому что они не удовлетворяют отношению "$\overset{F}{\succ}$". Необходимо привлечение дополнительной информации, то есть дополнительных знаний. В рассматриваемом случае, имея дополнительную информацию об истинности предиката "$|c_1| > |c_2|$", получаем:

$$F(1,1,1) = |c_1| - |c_2| + |c_3| > |c_3| = F(0,0,1),$$

при наличии такой достоверной дополнительной информации находится экстремальное решение $\tilde{\alpha}_1^*$.

Алгоритм решения в общем случае состоит в следующем:

1^0. Отыскивается множество П(À).

2^0. Проверяется достаточное условие разрешимости (утверждение 1).

3^0. Если выполняется условие $\Pi(\grave{A}) \subseteq N_j$, при некотором j, то экстремальные решения исчерпываются множеством точек П(А), и решение заканчивается.

4^0. Если условие $\Pi(\grave{A}) \subseteq N_j$ не выполняется при любом $j = \overline{1,N}$, то у экспертов запрашиваются дополнительные знания: запросы формулируются в виде некоторых неравенств (так, в примере 2 запрос имеет вид "$|c_1| > |c_2|$" ?)

5^0. Если дополнительных знаний достаточно для выбора экстремальных решений, то выдается точный ответ, иначе выдается À -паретовское множество как набор альтернатив для окончательного выбора ЛПР.

Литература.

1. Козлова М.Г. Системы поддержки принятия решений в современной информатике // Программы, системы, модели. – Симферополь, 1996. – № 2. – С.40-45.
2. Козлова М.Г. Синтез дискретных моделей выбора решений на основе знаний // Международной науч. конф. «Интеллектуализация обработки информации». Алушта, 3-7 июня 1996 г.: Тез. докл. /КрАН. – Симферополь, 1996. – С.13-14.
3. Feigenbaum E.A. How the «What» becomes the «How» // Communications of the ACM, 1996, 39, 5, p.97-104.

4. Chadha S.R., Mazlack L.J., Rick R.A. Using existing knowledge sources (cases) to build an expert system // Expert Systems, 1991, 8, 1, p.3-12.
5. Донской В.И. Задачи псевдобулевой оптимизации с дизъюнктивным ограничением // Журн.выч.матем. и матем.физики. – 1994. – 34, 3. С.461-472.
6. Донской В.И. Логические продукционные системы: анализ и синтез // Кибернетика и системный анализ. – 1994. – № 4. С.11-22.
7. Донской В.И. Дуальные экспертные системы // Изв.РАН Техническая кибернетика. – 1993. № 5. – С.111-119.
8. Donskoy V.I. Psevdo-Boolean Scalar Optimization Models with Incomplete Information //GMÖOR Newsletter, 1996, 1/2, p.20-26.
9. Донской В.И. Математические модели принятия решений при неполной информации: принципы разработки и синтетический подход // Программы, системы, модели. Симферополь, 1996. – №2. – С.6-17.
10. Донской В.И., Козлова М.Г. Модели принятия оптимальных решений на основе знаний // VI Международная конференция «Знания-диалог-решение». Ялта, 15-20 сентября 1997 г./ К.:АСПИС. 1997. - С.336-343.
11. Лукьяненко В.А., Руденко Л.И., Козлова М.Г. Принятие решений при неполных данных // VI Международная конференция «Знания-диалог-решение». Ялта, 15-20 сентября 1997 г./ К.:АСПИС. 1997. – С.373-380.