На правах рукописи

Вовк Алексей Андреевич

ТЕХНОЛОГИЯ ПОСТРОЕНИЯ ОБОБЩЕННОГО «И/ИЛИ» ДЕРЕВА РЕШЕНИЯ ЗАДАЧ

05.13.17 – теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Красноярск - 2008

Работа выполнена в Сибирском федеральном университете

Научный руководитель:	доктор технических наук, профессор Цибульский Геннадий Михайлович
Официальные оппоненты:	доктор физико-математических наук, профессор Носков Михаил Валерианович, доктор технических наук, профессор Ловчиков Анатолий Николаевич
Ведущая организация:	Институт вычислительной математики и математической геофизики СО РАН

Защита диссертации состоится 30 декабря 2008 года в 14.00 часов на заседании диссертационного совета Д 212.099.11 при Сибирском федеральном университете по адресу: 660074, г. Красноярск, ул. Киренского, 26, корпус Ж, ауд. 1-15.

С диссертацией можно ознакомиться в научной библиотеке Сибирского федерального университета по адресу: ул. Киренского, 26, ауд. Г274.

Автореферат разослан «30» ноября 2008 г.

Ученый секретарь
диссертационного совета
к.т.н., доцент Покидышева Л. И.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Одним из способов описания структуры предметной области, в решающих системах, основанных на знаниях, являются «И/ИЛИ» деревья решения задач. При этом обобщенное «И/ИЛИ» дерево должно быть априори задано в явном виде. Решающая система, взаимодействуя с пользователем, способна в процессе решения задачи строить решающее «И» дерево для конкретной задачи, однако, такие системы не способны к накоплению обобщенного знания о структуре предметной области. Построение же, корректировка и дополнение обобщенного «И/ИЛИ» дерева производится с помощью компонента «приобретение знаний», пользователем которой является инженер по знаниям (когнитолог). В этих условиях актуальной является разработка технологии построения адекватной базы знаний и соответствующего обобщенного «И/ИЛИ» дерева решающей системой в процессе взаимодействия с пользователем-природоведом.

Цель работы. Целью диссертационной работы является разработка технологии построения решающей системой «И/ИЛИ» дерева, изоморфного статической структуре оригинала.

В рамках цели решаются следующие задачи:

1) Анализ применимости операции над графами теории графов к «И/ИЛИ» деревьям решения задач.

2) Определение операции объединения вершин «И/ИЛИ» дерева, представляющих цели в пространстве целей.

3) Определение операции объединения решающих «И» деревьев в «И/ИЛИ» дерево.

4) Разработка алгоритма объединения решающих «И» деревьев в «И/ИЛИ» дерево.

5) Исследование сходимости процесса построения обобщенного «И/ИЛИ» дерева для класса задач.

Научная новизна работы:

1) Введено понятие совпадающих кластеров в пространстве целей и определена операция объединения пары совпадающих кластеров в объединенный кластер. Введено понятие совпадения «И» и «И/ИЛИ» деревьев решения задач через совпадающие вершины, представленные кластерами в пространстве целей.

2) Определена операция объединения «И/ИЛИ» деревьев. Данная операция позволяет путем объединения решающих «И» деревьев, построенных в

результате взаимодействия решающей системы и пользователя-природоведа, сформировать обобщенное «И/ИЛИ» дерево, описывающее статическую структуру оригинала. Разработан новый алгоритм объединения «И/ИЛИ» деревьев решения задач.

3) Сформулировано определение сходимости итерационного процесса построения обобщенного «И/ИЛИ» дерева для класса задач. Данное определение базируется на измерении степени сходства получаемых «И/ИЛИ» деревьев на каждом шаге объединения «И/ИЛИ» дерева с очередным решающим «И» деревом задачи данного класса. Сформулирован и доказан критерий сходимости такого процесса.

4) Выведена теоретическая оценка скорости сходимости процесса построения обобщенного «И/ИЛИ» дерева для класса задач.

Практическая ценность работы:

Разработан программно-аппаратный комплекс, позволяющий осуществлять объединение решающих «И» деревьев задач анализа сложных в обобщенное «И/ИЛИ» дерево. Комплекс внедрен и используется на Институте леса им. В.Н. Сукачёва СО РАН, а также в учебном процессе СФУ.

По материалам исследований опубликовано пять статей, из которых: одна опубликована в сборниках, рекомендованных ВАК, две статьи депонированы.

Диссертационная работа включает: содержание, введение, три главы, и заключение. Работа содержит 104 страницы машинописного текста, 43 рисунка и 21 таблицу. Список литературы содержит 110 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность темы диссертации, формулируется цель и задачи исследования.

В первой главе анализируется понятие «И/ИЛИ» дерева решения задач, как способа представления статической структуры оригинала в пространстве задач. Производится сопоставление понятия «И/ИЛИ» дерева со строгим определением дерева теории графов и применимость операций над графами теории графов к «И/ИЛИ» деревьям решения задач.

В результате установлено, что «И/ИЛИ» дерево лишь частично соответствует определению дерева теории графов и имеет характерные особенности:

1. Наличие циклов в «И/ИЛИ» дереве, что может интерпретироваться следующим образом: одна и та же вершина может являться дочерней для не-

скольких вершин. Это означает, что достижение одной и той же цели может быть необходимо для достижения нескольких целей более высокого уровня.

2. Корневая вершина «И/ИЛИ» дерева соответствует исходной цели, а все остальные вершины – подцелям исходной цели на разном уровне детализации.

3. Между ребрами дерева существуют отношения «И» и «ИЛИ». Наличие отношения «ИЛИ» указывает на, что «И/ИЛИ» дерево содержит несколько путей достижения исходной цели.

4. Подцели в дереве целей иерархически упорядочены. Каждая дочерняя вершина представляет собой подцель цели, соответствующей родительской вершине, находящейся на более высоком уровне дерева. В дереве целей отсутствуют ребра, связывающие вершины одного уровня, т.е. на одном и том же уровне дерева цели взаимно независимы. Уровень вершины в дереве указывает на уровень детализации или уровень абстракции цели, представленной данной вершиной, при редукции исходной цели.

Выяснено, что в чистом виде алгебраические операции теории графов неприменимы к «И/ИЛИ» деревьям в виду того, что операции над графами в теории графов порождены теоретико-множественными операциями над множествами вершин и ребер, и для применения данных операций множество вершин должно быть априори определено и поименовано. В случае с «И/ИЛИ» деревьями имеют место ситуации, когда множество вершин невозможно задать априори, т.е. вершины строятся в процессе построения дерева. Кроме того, вершины не могут быть поименованы так, чтобы к ним можно было применить теоретико-множественные операции. Так же нет определения соответствующих операции над иерархически упорядоченными множествами вершин и множествами ребер, между которыми существуют отношения «И» и «ИЛИ».

Установлено, что сегодня нет систем формирования знаний, которые способны, оперируя решающими «И» деревьями, строить обобщенное «И/ИЛИ» дерево в процессе взаимодействия с пользователем-природоведом. Существующий подход к построению обобщенного «И/ИЛИ» дерева ориентирован на работу с инженером по знаниям посредством использования компонента «приобретение знаний».

Во второй главе определяется операция объединения «И» деревьев в «И/ИЛИ» дерево. Объединение «И» деревьев осуществляется посредством поиска совпадающих вершин объединяемых деревьев и построения объединенных вершин-элементов результирующего дерева. С этой целью вводится понятие совпадающих кластеров и способ построения объединенного кластера. Процесс построения «И/ИЛИ» дерева исследуется на сходимость: формулируется критерий сходимости и оценка скорости сходимости.

Задачу классификации определим следующим образом: В признаковом пространстве X задано m классов множеством своих эталонов $\{E_i\}_{i=1}^{m}$, $\rho(x, y)$ – метрика, определенная в пространстве X. Для каждого класса задан критерий компактности ε_i. Для всех классов выполнено условие ε-непересекаемости. Всякий объект x относится к тому классу, для которого $\rho(E_i, x_i) < \varepsilon_i$, т.е. расстояние между исследуемым объектом и эталоном класса, удовлетворяет критерию компактности данного класса. Если же объект не будет отнесён ни к одному известному классу, он будет классифицирован как «неизвестный» или как фон.

Каждая цель «И/ИЛИ» дерева целей является задачей опознавания, определяемая кластером, заданным в признаковом пространстве $X(x_1,x_2,\ldots,x_n)$ значениями эталона $E(e_1,e_2,\ldots,e_n)$ и критерия компактности $\varepsilon(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)$.

Пусть в некотором признаковом пространстве $X(x_1,x_2,\ldots,x_n)$ заданы два кластера A_1 и A_2 своими эталонами $E_1(e_{1,1},e_{1,2},\ldots,e_{1,n})$ и $E_2(e_{2,1},e_{2,2},\ldots,e_{2,n})$ и критериями компактности $\varepsilon_1(\varepsilon_{1,1},\varepsilon_{1,2},\ldots,\varepsilon_{1,n})$ и $\varepsilon_2(\varepsilon_{2,1},\varepsilon_{2,2},\ldots,\varepsilon_{2,n})$.

Кластеры A_1 и A_2 называются *совпадающими*, если эталон каждого из них принадлежит другому кластеру (рис. 1). Т.е. $|e_{1,i} - e_{2,i}| < \min(\varepsilon_{1,i}, \varepsilon_{2,i})$, $i = 1,\ldots,n$.

Объединенным кластером для пары совпадающих кластеров A_1 и A_2 называется кластер A_3, эталон $E_3(e_{3,1}, e_{3,2},\ldots, e_{3,n})$ и критерий компактности $\varepsilon_3(\varepsilon_{3,1},\varepsilon_{3,2},\ldots,\varepsilon_{3,n})$ которого рассчитываются по формулам:

$$E_3 = \frac{C_{A_1}E_1 + C_{A_2}E_2}{C_{A_1} + C_{A_2}},$$

где C_{A_1} и C_{A_2} – весовые коэффициенты кластеров A_1 и A_2;

$$\varepsilon_{3,i} = \max(\varepsilon_{1,i} + |e_{1,i} - e_{3,i}|, \varepsilon_{2,i} + |e_{2,i} - e_{3,i}|), \mathrm{i} = \overline{1,\mathrm{n}}.$$

Весовой коэффициент (C) кластера A равен единице ($C_A = 1$), если кластер A был найден в процессе достижения цели или был задан априори. Если кластер A был построен, как объединенный для пары совпадающих кластеров A_1 и A_2, то: $C_A = C_{A_1} + C_{A_2}$.

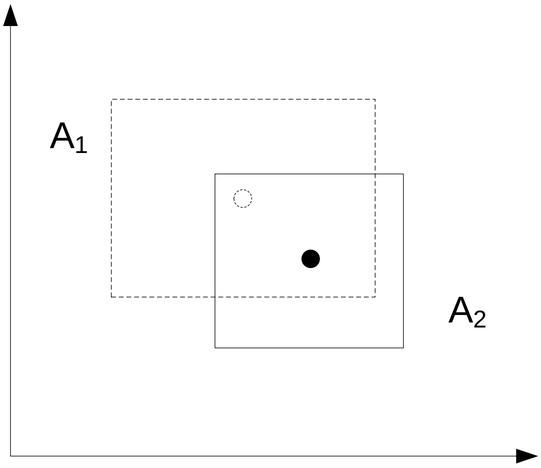

Рисунок 1 – Кластеры A_1 и A_2 совпадают

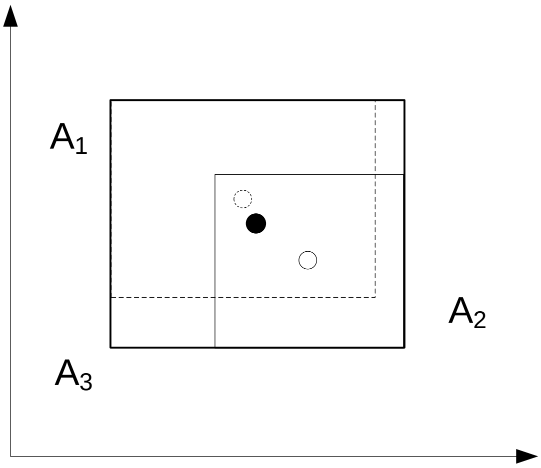

Рисунок 2 – Объединение совпадающих кластеров

Так как каждой вершине «И» дерева соответствует кластер то, вершины называются *совпадающими*, если совпадают соответствующие им кластеры. Вершина A_3 называется *объединенной вершиной* для совпадающих вершин A_1 и A_2, если она соответствует объединенному кластеру для совпадающих кластеров вершин A_1 и A_2. *Весовым коэффициентом вершины* называется весовой коэффициент кластера, соответствующего данной вершине. Для обозначения совпадающих вершин используется знак “$\approx$”.

Используя данные определения совпадающих вершин, введем еще несколько определений.

Два двухуровневых «И» дерева называются *полностью совпадающими,* если на соответствующих уровнях данных деревьев все вершины совпадают.

Два двухуровневых дерева называются *частично совпадающими*, если у них совпадает хотя бы одна вершина.

Два двухуровневых «И/ИЛИ» дерева называются *полностью совпадающими*, если:

1) каждая вершина одного дерева совпадает с вершиной соответствующего уровня другого дерева;

2) для каждого двухуровневого «И» поддерева одного из деревьев найдется полностью совпадающее с ним «И» поддерево другого дерева.

Два «И/ИЛИ» дерева называются *полностью совпадающими*, если для всякого двухуровневого «И/ИЛИ» поддерева в одном из деревьев существует полностью совпадающее с ним двухуровневое «И/ИЛИ» поддерево в другом дереве.

Два «И/ИЛИ» дерева называются *полностью совпадающими*, если для всякого двухуровневого «И/ИЛИ» поддерева в одном из деревьев существует полностью совпадающее с ним двухуровневое «И/ИЛИ» поддерево в другом дереве.

Два «И/ИЛИ» дерева называются *частично совпадающими*, если хотя бы один узел одного дерева совпадает с узлом другого соответствующего уровня другого дерева.

Введем способ задания «И/ИЛИ» дерева. «И/ИЛИ» дерево G описывается тройкой $G = \{G_V, G_E, G_I\}$, где:

1) множество вершин $G_V = \{G_1, G_2, ..., G_n\}$, где

G_i – множество вершин i-го уровня дерева G, n – число уровней дерева G;

2) G_E – множество ребер дерева G;

3) G_I – множество двухуровневых «И» поддеревьев дерева G.

Приведем алгоритм объединения двух «И/ИЛИ» деревьев. Т.к. «И» дерево является частным случаем «И/ИЛИ» дерева, следовательно, данный алгоритм применим и для объединения «И» дерева с «И/ИЛИ» деревом.

Пусть заданы два дерева $G^1 = (\{G_1^1, G_2^1, ..., G_n^1\}, G_E^1, G_I^1)$ и $G^2 = (\{G_1^2, G_2^2, ..., G_n^2\}, G_E^2, G_I^2)$. Результатом работы алгоритма является «И/ИЛИ» дерево G^3, обобщающее деревья G^1 и G^2: $G^3 = (\{G_1^3, G_2^3, ..., G_n^3\}, G_E^3, G_I^3)$.

Суть алгоритма в последовательном просмотре уровней обоих деревьев и поиске совпадающих вершин на соответствующих уровнях. Если совпадающие вершины найдены, то для каждой пары совпадающих вершин строится объединенная вершина. При этом вхождение совпадающих вершин в множества $G_E^1, G_I^1, G_E^2, G_I^2$ заменяются на построенную объединенную вершину. Результирующие дерево строится из объединенных вершин для совпадающих вершин исходных деревьев, вершин, для которых не нашлось совпадающих, и всех ребер исходных деревьев с учетом замены совпадающих вершин на объединенные вершины.

1) Полагаем $i = 1$.

2) Строим новое множество $G_i^3 = \varnothing$.

3) Если множество G_i^2 пустое, то переходим к шагу 11, иначе переходим к шагу 4.

4) Если множество G_i^1 пустое, то переходим к шагу 11, иначе переходим к шагу 5.

5) Первую вершину из множества G_i^2 обозначим *A*. Поочередно сравниваем *A* с каждой вершиной множества G_i^1. Если найдена вершина множества G_i^1, совпадающая с *A*, то обозначаем ее *B* и переходим к шагу 6, иначе переходим к шагу 10.

6) Для пары совпадающих вершин *A* и *B* строим объединенную вершину, которую обозначим *C*. Добавляем *C* к множеству G_i^3.

7) Все вхождения вершины *A* в ребра множества G_E^2 заменяем на вершину *C*. Все вхождения вершины *A* в поддеревья множества G_I^2 заменяем на вершину *C*. Если в G_E^2 образовались одинаковые ребра, то повторения удаляем. Если в G_I^2 образовались одинаковые поддеревья, то повторения удаляем.

8) Все вхождения вершины *B* в ребра множества G_E^1 заменяем на вершину *C*. Все вхождения вершины *B* в поддеревья множества G_I^1 заменяем на вершину *C*. Если в G_E^1 образовались одинаковые ребра, то повторения удаляем. Если в G_I^1 образовались одинаковые поддеревья, то повторения удаляем.

9) Исключаем вершину *B* из множества G_i^1.

10) Исключаем вершину *A* из множества G_i^2. Переходим к шагу 3.

11) Строим множество $G_i^3 = G_i^3 \cup G_i^1 \cup G_i^2$. (Данную операцию можно интерпретировать так: добавляем к множеству G_i^3 все элементы множеств G_i^1 и G_i^2, если таковые остались после выполнения предыдущих шагов алгоритма.

Таким образом, G_i^3 содержит построенные объединенные вершины для совпадающих вершин множеств G_i^1 и G_i^2 и все несовпадающие вершины этих множеств.)

12) Если $i = n$, то алгоритм завершен, иначе полагаем $i = i + 1$ и переходим к шагу 2 (другими словами, повторяем шаги 2-11 для каждого уровня объединяемых деревьев).

Результатом работы алгоритма является «И/ИЛИ» дерево G^3, обобщающее деревья G^1 и G^2: $G^3 = (\{G_1^3, G_2^3, ..., G_n^3\}, G_E^3, G_I^3)$. Схема предложенного алгоритма приведена на рисунке 3.

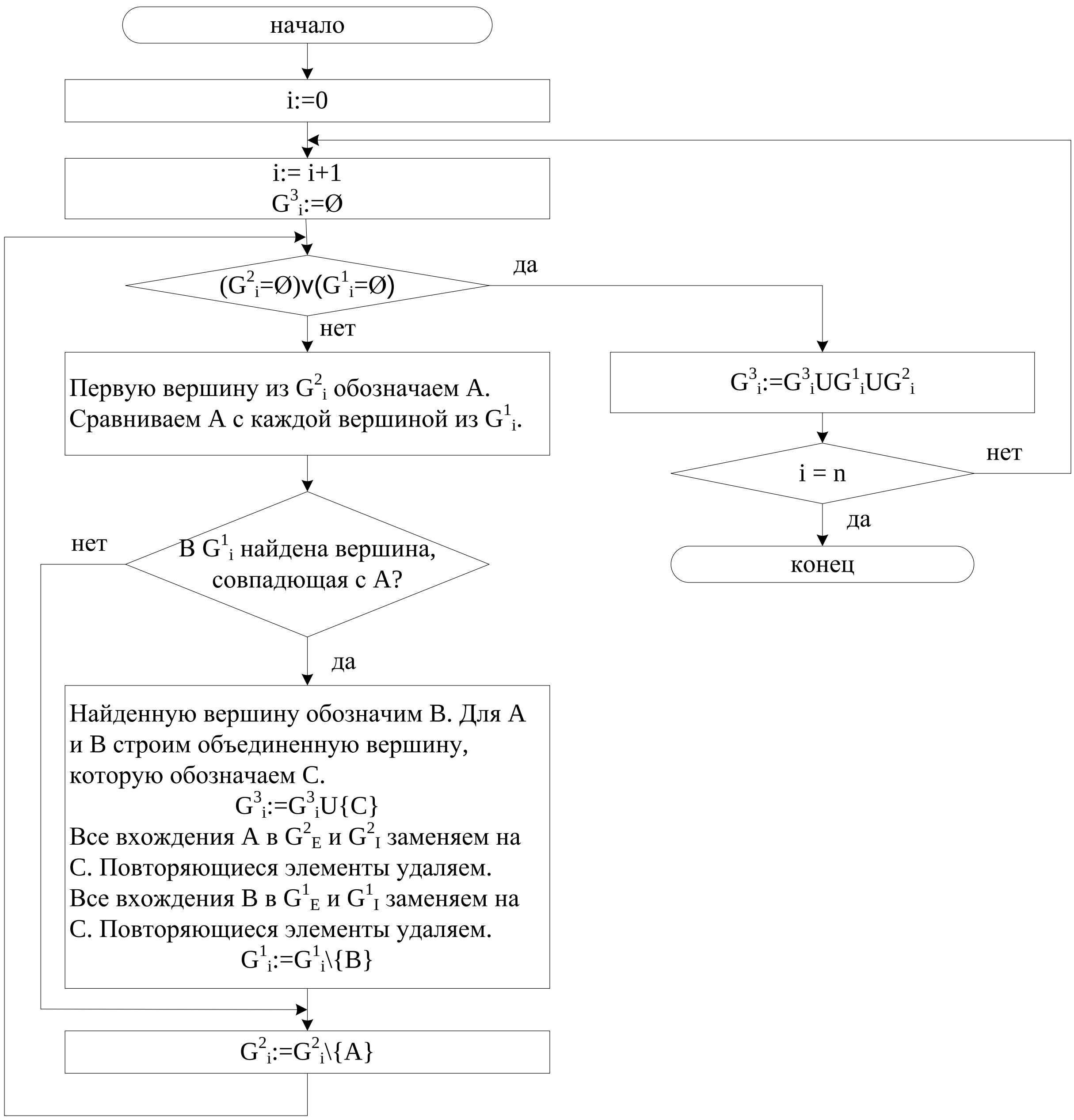

Рисунок 3 – Схема алгоритма объединения «И/ИЛИ» деревьев

Пусть $\boldsymbol{u(G_1, G_2)}$ – операция построения обобщенного «И/ИЛИ» дерева для двух «И/ИЛИ» деревьев G_1 и G_2, определенная в соответствии с разработанным алгоритмом. Операнды могут быть G_1 и G_2 как «И/ИЛИ» деревьями, так и «И» деревьями.

Пусть «И/ИЛИ» дерево $G_3 = u(G_1, G_2)$. От дерева G_1 дерево G_3 может в общем случае отличаться следующими параметрами:

1) структурой «И/ИЛИ» дерева, а именно:

а) в дереве G_3 добавились новые вершины, которых не было в G_1;

б) в дереве G_3 появились отношения «И» или «ИЛИ» между ветвями, которых не было в G_1;

2) изменились вершины (эталоны и критерии компактности соответствующих вершинам кластеров) дерева G_1, входящие теперь в состав дерева G_3.

Нетрудно видеть, что структура «И/ИЛИ» дерева G_3 не изменяется по сравнению с деревом G_1 при объединении его с «И» деревом G_2, если дерево G_2 удовлетворяет условию: *Всякое двухуровневое «И» поддерево дерева G_2 полностью совпадает с некоторым двухуровневым «И» поддеревом дерева G_1.*

Пусть $\boldsymbol{s(G)}$ – функция, вычисляющая сумму весовых коэффициентов всех вершин дерева G.

Из определения весового коэффициента кластера и операции $\boldsymbol{u}$ следует, что:

$$s(u(G_1, G_2)) = s(G_1) + s(G_2). \tag{1}$$

Пусть $\boldsymbol{q(G)}$ – функция, вычисляющая количество вершин дерева G.

Пусть заданы множество $\{T_1, T_2, \dots, T_n\}$ решающих «И» деревьев задач одного класса и последовательность $\{G_i\}_{i=1}^{n}$ обобщенных «И/ИЛИ» деревьев для задач данного класса, определенных следующим образом:

$G_1 = T_1$,

$G_i = u(G_{i-1}, T_i)$, $i = 2,3,\dots,n$.

(2)

Тогда *коэффициентом прироста* дерева G_{k+1} по сравнению с деревом G_k назовем величину:

$$d(G_k, G_{k+1}) = \frac{q(G_{k+1}) - q(G_k)}{s(G_k)}.$$

Очевидно, что $d(G_k, G_{k+1}) \geq 0$.

Равенство $d(G_k, G_{k+1}) = 0$ говорит о том, что в «И/ИЛИ» дереве G_{k+1} не появились новые узлы по сравнению с исходным «И/ИЛИ» деревом G_k при объединении его с решающим «И» деревом T_{k+1}.

Пусть A_1, A_2 – кластеры, определенные в признаковом пространстве X. E_1, ε_1 – эталон и критерий компактности кластера A_1, E_2, ε_2 – эталон и критерий компактности кластера A_2.

$\boldsymbol{U(A_1, A_2)}$ – операция построения объединенного кластера для пары совпадающих кластеров A_1, A_2.

Пусть $A_3 = U(A_1, A_2)$ – объединенный кластер для совпадающих кластеров A_1 и A_2; E_3, ε_3 – эталон и критерий компактности кластера A_3.

Коэффициентом изменения кластера A_3 по сравнению с кластером A_1 для совпадающих вершин A_1 и A_2 будем называть величину:

$$D(A_1, A_3) = \frac{\rho(E_1, E_3)}{|\varepsilon_1|},$$

где $\rho(E_1, E_3)$ – евклидово расстояние между E_1 и E_3 в пространстве X.

Использование ε_1 в знаменателе данной формулы необходимо, т.к. кластеры, соответствующие вершинам дерева, на разных уровнях дерева описываются качественно различными признаками с различным разбросом значений. На пример, в решающем «И» дереве задачи анализа космофотоснимков лесов Саян кластеры, соответствующие вершинам на уровне пикселей описываются текстурными признаками с разбросом значений 0-90. Значения признаков (площадь, периметр и др.) уровня сегментов имеют разброс 100-2500.

Итак, пусть задана последовательность $\{G_i\}_{i=1}^{n}$ обобщенных «И/ИЛИ» деревьев для класса задач, определенная в (2). Процесс построения обобщенного «И/ИЛИ» дерева для данного класса задач называется *сходящимся*, если найдется такой номер n, начиная с которого выполняются условия:

1) коэффициент прироста обобщенного «И/ИЛИ» дерева $d(G_n, G_{n+1})$ не превышает некоторую наперед заданную малую величину;

2) максимальный коэффициент изменения $D(A_i, A_j)$ по всем изменившимся вершинам обобщенного «И/ИЛИ» дерева G_{n+1} по сравнению с обобщенным «И/ИЛИ» деревом G_n не превышает некоторую наперед заданную малую величину.

Сходимость процесса формирования обобщенного «И/ИЛИ» дерева означает, что на некотором шаге объединения текущего «И/ИЛИ» дерева с очередным «И» деревом полученное «И/ИЛИ» дерево будет отличаться от текущего не более, чем на требуемую величину. Т.е. обобщенное «И/ИЛИ» дерево приобретет устойчивость.

Замечание: Следует отличать введенное понятие сходимости процесса построения обобщенного «И/ИЛИ» дерева решения задач от сходимости последовательностей или рядов. Сходящимися в определениях математического анализа здесь являются последовательности коэффициентов прироста дерева и коэффициентов изменения кластера. Поэтому сходимость последовательности «И/ИЛИ» деревьев правильно будет считать внутренней сходимостью или стабилизацией обобщенного «И/ИЛИ» дерева.

Утверждение: Пусть $\{T_1, T_2, \ldots\}$ множество решающих «И» деревьев задач одного класса и последовательность $\{G_i\}_{i=1}^{\infty}$ обобщенных «И/ИЛИ» деревьев для задач данного класса, определенны следующим образом:

$G_1 = T_1$,

$G_i = u(G_{i-1}, T_i)$, $i = 2,3,\ldots$

Если для всех «И» деревьев T_i задач заданного класса существует $M \in N$ такое что, $q(T_i) \le M$, т.е. количество вершин во всех возможных решающих «И» деревьях задач данного класса ограничено, то процесс построения обобщенного «И/ИЛИ» дерева для данного класса задач сходится.

Доказательство:

1. Докажем, что $\lim\limits_{n\to\infty} d(G_n, G_{n+1}) = 0$.

$$\lim_{n\to\infty} d(G_n, G_{n+1}) = \lim_{n\to\infty} \frac{q(G_{n+1}) - q(G_n)}{s(G_n)} = \lim_{n\to\infty} \frac{q(u(G_n, T_{n+1})) - q(G_n)}{s(G_n)} .$$

Величина $q(u(G_n, T_{n+1})) - q(G_n)$ показывает, сколько новых узлов появилось в обобщенном «И/ИЛИ» дереве G_n при объединении его с «И» деревом T_{n+1}. Так как у деревьев G_n и T_{n+1} совпадают, по крайней мере, корневые вершины, то:

$$q(u(G_n, T_{n+1})) - q(G_n) \le q(G_n) + q(T_{n+1}) - 1 - q(G_n) = q(T_{n+1}) - 1 \le M - 1$$

.

Тогда:

$$\lim_{n\to\infty} \frac{q(u(G_n, T_{n+1})) - q(G_n)}{s(G_n)} \le \lim_{n\to\infty} \frac{M-1}{s(G)} .$$

Учитывая (1), $s(G_n) = \sum_{i=1}^{n} s(T_i)$.

Следовательно, $\inf s(G_n) = n$, т.к. все деревья T_i, содержат как минимум одну вершину. Тогда:

$$\lim_{n\to\infty}\frac{M-1}{s(G_n)}\le\lim_{n\to\infty}\frac{M-1}{n}$$

Итак:

$$0\le\lim_{n\to\infty}d(G_n,G_{n+1})\le\lim_{n\to\infty}\frac{M-1}{n}=0,$$

следовательно: $\lim\limits_{n\to\infty}d(G_n,G_{n+1})=0$.

2. Пусть $\{A_1, A_2, \ldots\}$ множество кластеров и последовательность кластеров $\{B_i\}_{i=1}^{\infty}$, определенны в признаковом пространстве X, так что:

$B_1 = A_1$,

$B_i = U(B_{i-1}, A_i)$, $i = 2,3,\ldots$

Пусть $E_n(e_{n,1},e_{n,2},\ldots,e_{n,m})$, $\varepsilon_n(\varepsilon_{n,1},\varepsilon_{n,2},\ldots,\varepsilon_{n,m})$ – эталон и критерий компактности кластера B_n, $E_{n+1}(e_{n+1,1},e_{n+1,2},\ldots,e_{n+1,m})$, $\varepsilon_{n+1}(\varepsilon_{n+1,1},\varepsilon_{n+1,2},\ldots,\varepsilon_{n+1,m})$ – эталон и критерий компактности кластера A_{n+1}. Докажем, что $\lim\limits_{n\to\infty}D(B_n,B_{n+1})=0$.

$$\lim_{n\to\infty}D(B_n,B_{n+1})=\lim_{n\to\infty}D(B_n,U(B_n,A_{n+1}))=\lim_{n\to\infty}\frac{\rho(E_n,\dfrac{C_{B_n}E_n+C_{A_{n+1}}E_{n+1}}{C_{B_n}+C_{A_{n+1}}})}{|\varepsilon_n|}=$$

$$=\lim_{n\to\infty}\frac{\rho(E_n,\dfrac{nE_n+E_{n+1}}{n+1})}{|\varepsilon_n|}=\lim_{n\to\infty}\frac{\sqrt{\sum\limits_{i=1}^{m}(e_{n,i}-\dfrac{ne_{n,i}+e_{n+1,i}}{n+1})^2}}{|\varepsilon_n|}=$$

$$=\lim_{n\to\infty}\frac{\sqrt{\sum\limits_{i=1}^{m}(\dfrac{ne_{n,i}+e_{n,i}-ne_{n,i}-e_{n+1,i}}{n+1})^2}}{|\varepsilon_n|}=\lim_{n\to\infty}\frac{\sqrt{\dfrac{1}{(n+1)}\sum\limits_{i=1}^{m}(e_{n,i}-e_{n+1,i})^2}}{|\varepsilon_n|}\le$$

$$\le\lim_{n\to\infty}\frac{\min(|\varepsilon_n|,|\varepsilon_{n+1}|)\sqrt{\dfrac{1}{(n+1)}}}{|\varepsilon_n|}\le\lim_{n\to\infty}\sqrt{\frac{1}{(n+1)}}=0.$$

Итак:

$$0\le D(B_n,B_{n+1})\le\lim_{n\to\infty}\sqrt{\frac{1}{n+1}}=0,$$

следовательно: $\lim\limits_{n\to\infty}D(B_n,B_{n+1})=0$.

Утверждение доказано.

Следствие 1: Пусть $\{T_1, T_2, \ldots\}$ множество решающих «И» деревьев задач одного класса и последовательность $\{G_i\}_{i=1}^{\infty}$ обобщенных «И/ИЛИ» деревьев для задач данного класса, определенны следующим образом:

$G_1 = T_1$,

$G_i = u(G_{i-1}, T_i)$, $i = 2,3,\ldots$

Если для всех «И» деревьев T_i, i=1,2,… задач заданного класса $\exists M \in N$ такое что, $q(T_i) \le M$, то $\forall \delta > 0$ выполняется $d(G_n, G_{n+1}) < \delta$ для $n > \dfrac{M-1}{\delta}$, $n \in N$.

Доказательство:

$$d(G_n, G_{n+1}) = \frac{q(G_{n+1}) - q(G_n)}{s(G_n)} = \frac{q(u(G_n, T_{n+1})) - q(G_n)}{s(G_n)} =$$

$$= \frac{q(G_n) + q(T_{n+1}) - 1 - q(G_n)}{s(G_n)} = \frac{q(T_{n+1}) - 1}{s(G_n)} \le \frac{q(T_{n+1}) - 1}{n} \le \frac{M-1}{n} .$$

Т.к. $n > \dfrac{M-1}{\delta}$, то: $\dfrac{M-1}{n} < \dfrac{M-1}{\frac{M-1}{\delta}} = \delta$

Итак: $d(G_n, G_{n+1}) < \delta$.

Следствие доказано.

Суть следствия заключается в том, что если количество вершин во всех возможных решающих «И» деревьях задач данного класса ограничено некоторым числом M, то необходимая точность δ будет достигнута не более, чем за $\dfrac{M-1}{\delta}$ шагов.

Следствие 2: Пусть $\{A_1, A_2, \ldots\}$ множество кластеров и последовательность кластеров $\{B_i\}_{i=1}^{\infty}$, определенны в признаковом пространстве X, так что:

$B_1 = A_1$,

$B_i = U(B_{i-1}, A_i)$, $i = 2,3,\ldots$

Тогда $\forall \delta > 0$ выполняется $D(B_n, B_{n+1}) < \delta$ для $n > \left(\dfrac{1}{\delta}\right)^2 -1$, $n \in N$.

Доказательство:

Пусть $E_n(e_{n,1}, e_{n,2}, \ldots, e_{n,m})$, $\varepsilon_n(\varepsilon_{n,1}, \varepsilon_{n,2}, \ldots, \varepsilon_{n,m})$ – эталон и критерий компактности кластера B_n, $E_{n+1}(e_{n+1,1}, e_{n+1,2}, \ldots, e_{n+1,m})$, $\varepsilon_{n+1}(\varepsilon_{n+1,1}, \varepsilon_{n+1,2}, \ldots, \varepsilon_{n+1,m})$ – эталон и критерий компактности кластера A_{n+1}.

$$D(B_n, B_{n+1}) = D(B_n, U(B_n, A_{n+1})) = \frac{\rho(E_n, \dfrac{C_{B_n} E_n + C_{A_{n+1}} E_{n+1}}{C_{B_n} + C_{A_{n+1}}})}{|\varepsilon_n|} =$$

$$= \frac{\rho(E_n, \dfrac{nE_n + E_{n+1}}{n+1})}{|\varepsilon_n|} = \frac{\sqrt{\sum_{i=1}^{m}(e_{n,i} - \dfrac{ne_{n,i} + e_{n+1,i}}{n+1})^2}}{|\varepsilon_n|} =$$

$$= \frac{\sqrt{\sum_{i=1}^{m}(\dfrac{ne_{n,i} + e_{n,i} - ne_{n,i} - e_{n+1,i}}{n+1})^2}}{|\varepsilon_n|} = \frac{\sqrt{\dfrac{1}{(n+1)}\sum_{i=1}^{m}(e_{n,i} - e_{n+1,i})^2}}{|\varepsilon_n|} =$$

$$= \frac{\rho(E_n, E_{n+1})\sqrt{\dfrac{1}{(n+1)}}}{|\varepsilon_n|} < \sqrt{\frac{1}{n+1}} < \sqrt{\frac{1}{\left(\dfrac{1}{\delta}\right)^2 - 1 + 1}} = \delta .$$

Итак: $D(B_n, B_{n+1}) < \delta$.

Следствие доказано.

Суть следствия заключается в том, что объединенный кластер для множества совпадающих кластеров станет устойчивым не более, чем за $\left(\frac{1}{\delta}\right)^2$ -1 шагов.

Теоретическая оценка сходимости основывается на требовании принадлежности решаемых задач одному классу: *две задачи принадлежат одному классу задач тогда и только тогда, когда совпадают корневые вершины решающих «И» деревьев данных задач*. Такая оценка является точной верхней гранью скорости сходимости, т.к. следует из предположения о минимальной схожести решающих «И» деревьев задач одного класса.

В третьей главе представлены экспериментальные результаты.

Целью проводимых экспериментальных исследований является проверка работоспособности разработанной технологии построения обобщенного «И/ИЛИ» дерева для решающих «И» деревьев класса задач.

В эксперименте были использованы космофотоснимки «Landsat™» лесов Саян и четыре выборки, представляющие участки леса с различными характеристиками. Для каждой цели, представленной выборкой, производилось построение «И» дерева следующим образом.

1) Системой выполнялся расчет значений признаков уровня пикселей (яркостные и текстурные) для каждого элемента изображения. По рассчитанным значениям системой определялись информативные признаки, наилучшим образом отделяющие элементы выборки от всего остального (фона). В сформированной системе признаков элементы производилось разбиение элементов выборки на априори не заданное число кластеров. Для построения кластеров использовался метод расширяющихся окружностей. Сформированные кластеры соответствуют вершинам «И» дерева уровня пикселей.

2) На множестве пикселей изображения, попадающих в каждый из построенных кластеров, системой выделялись примитивы. Для каждого примитива рассчитывались средние значение признаков уровня пикселей по всем входящим в него пикселям и размер примитива. Таким образом, на пикселях каждого класса системой были построены кластеры, соответствующие вершинам «И» дерева уровня примитивов.

3) Для множества выделенных примитивов системой осуществлялось построение сегментов. Для каждого сегмента производился расчет признаков уровня сегментов. В полученной системе признаков множестве сегментов разбивалось на кластеры. Разбиение производилась с помощью алгоритма ISODATA на априори не заданное число кластеров. Сформированные таким образом кластеры, соответствуют вершинам «И» дерева уровня сегмент.

4) Системой вычислялись средние значения по всем признакам уровня сегментов для всех построенных сегментов. По полученным значениям формировались кластеры, соответствующие вершинам «И» дерева уровня класс сегментов.

Описание всех построенных вершин приводится в соответствующих таблицах.

Таким образом, было построено четыре решающих «И» дерева для целей пользователя: T_1, T_2, T_3, T_4. Последовательность обобщенных «И/ИЛИ» деревьев имела вид:

$G_2 = u(T_1,T_2)$;

$G_3 = u(\mathrm{G}_2,T_3)$;

$G_4 = u(\mathrm{G}_3,T_4)$.

На каждом шаге объединения «И» деревьев вычислялись коэффициенты прироста дерева и коэффициенты изменения кластеров. На первом шаге коэффициенты равны:

$d(T_1,G_2) = 0{,}429$

$$\max_{G_2} D = 0{,}202 .$$

На третьем шаге соответственно:

$d(G_3,G_4) = 0{,}0467$

$$\max_{G_4} D = 0{,}086 .$$

На основании полученных результатов в целом можно сделать вывод об адекватности предложенных операций построений объединенного кластера и объединения «И/ИЛИ» деревьев. Экспериментально продемонстрировано, что имеет место сходимость процесса построения обобщенного «И/ИЛИ» дерева для класса задач. Установлено также, что скорость сходимости на практике гораздо выше чем, предложенная теоретическая оценка.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1) Введено понятие совпадающих кластеров в пространстве целей и определена операция объединения пары совпадающих кластеров в объединенный кластер. Введено понятие совпадения «И» и «И/ИЛИ» деревьев решения задач через совпадающие вершины, представленные кластерами в пространстве целей.

2) Определена операция объединения «И/ИЛИ» деревьев. Данная операция позволяет путем объединения решающих «И» деревьев, построенных в результате взаимодействия решающей системы и пользователя-природоведа, сформировать обобщенное «И/ИЛИ» дерево, описывающее статическую структуру оригинала. Разработан новый алгоритм объединения «И/ИЛИ» деревьев решения задач.

3) Сформулировано определение сходимости итерационного процесса построения обобщенного «И/ИЛИ» дерева для класса задач. Данное определение базируется на измерении степени сходства получаемых «И/ИЛИ» деревьев на каждом шаге объединения «И/ИЛИ» дерева с очередным решающим «И» деревом задачи данного класса. Сформулирован и доказан критерий сходимости такого процесса.

4) Выведена теоретическая оценка скорости сходимости процесса построения обобщенного «И/ИЛИ» дерева для класса задач.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Вовк, А. А. Технология формирования «И/ИЛИ» дерева задач анализа изображений / А. А. Вовк, Г. М. Цибульский, А. А. Латынцев // Вестник Вестник Сибирской аэрокосмической академии им. академика М. Ф. Решетнева. – 2008. – №2 (19). – С. 33-41.

2. Вовк, А. А. Построение обобщенного «И/ИЛИ» дерева решения задач анализа изображений / А. А. Вовк // Аналитические и численные методы моделирования естественнонаучных и социальных проблем: сборник статей III Международной научно-технической конференции. – Пенза: Приволжский Дом знаний, 2008. – С. 295-297.

3. Вовк, А. А. Сходимость процесса формирования «И/ИЛИ» дерева решения задач анализа изображений / А. А. Вовк; СФУ-Красноярск, 2008. – 12 с. – ил. – Библиогр.: 3 назв. – Рус. – Деп. в ВИНИТИ 19.09.08 № 739 – В 2008.

4. Вовк, А. А. Исследование сходимости процесса формирования «И/ИЛИ» дерева решения задач анализа изображений / А. А. Вовк, А. А. Латынцев; СФУ-Красноярск, 2008. – 13 с. – ил. – Библиогр.: 3 назв. – Рус. – Деп. в ВИНИТИ 19.09.08 № 740 – В 2008.

5. Вовк, А. А. Формирование обобщенного «И/ИЛИ» дерева для класса задач анализа изображений / А. А. Вовк // Решетневские чтения: материалы XII Международной научной конференции, посвящ. памяти генерального конструктора ракетно-космических систем академика М.Ф. Решетнева; под общ. ред. И.В. Ковалева / Сиб. гос. аэрокосмич. ун-т. – Красноярск, 2008. – С. 101-102.

Вовк Алексей Андреевич
Технология построения обобщенного «И/ИЛИ» дерева решения задач.
Автореферат диссертации на соискание ученой степени кандидата технических наук.
Подписано в печать 25.11.2008. Заказ №
Формат 60×90/16. Усл. печ. л. 1. тираж 100 экз.