УДК 519.713

ПРЕДСТАВЛЕНИЯ МНОГОМЕСТНЫХ ОТНОШЕНИЙ В ВИДЕ КОМПОЗИЦИИ БИНАРНЫХ ОТНОШЕНИЙ

СИНЕЛЬНИКОВА О.И., ВЕЧИРСКАЯ И.Д.

Рассматривается идея представления и обработки многоместных отношений в виде композиции бинарных отношений. Проводится метод разбиения многоместного отношения на бинарные, а также исследуется подход к построению многоместного отношения по бинарным. Рассматривается проблема формализации естественного языка с точки зрения бинарного представления.

Введение

При попытке формально описать какое-либо отношение, явление реального мира сталкиваемся с многоместными отношениями. Арность отношения определяется предметной областью, а также решаемой задачей. В большинстве случаев практически невозможно построить сразу многоместное отношение, т.е. учесть одновременно влияние всех факторов. Для того чтобы построить такое отношение, нужно досконально знать предметную область, причем не в виде законов, действующих на данной предметной области, а в виде существующих в ней наборов значений. Такой подход не эффективен и неприменим к задачам с большим количеством факторов и/или широкой предметной области.

Однако человек, сам того не подозревая, реализует постоянно в реальной жизни свой особый алгоритм. Суть этого алгоритма в том, что он сначала строит отношение между парами, группами факторов, а затем анализирует отношение между полученными группами. Таким образом, мозг человека строит многоуровневую экспертную систему, результатом которой является ответ на вопрос с учетом большого количества факторов.

Так в чем же состоит интерес идеи?

Во-первых, в том, что многоместное отношение можно представить, как комбинацию бинарных отношений. Это в значительной мере облегчает задачу формализации отношений. Работа с бинарными отношениями довольно простая с точки зрения представления их математическими моделями.

Во-вторых, при таком подходе возможна параллельная обработка информации, что существенно повышает эффективность процесса решения.

В-третьих, можно решать большое количество задач, заданных на одной предметной области.

Однако возникает вопрос о правомерности такого подхода. В данной работе проведен анализ процесса бинаризации, осуществлена попытка выявить условия, при которых это возможно, а также найти методы решения проблемы, когда эти условия не выполняются. Кроме того, осуществлена попытка применить концепцию бинаризации к конкретной задаче, а именно, к проблеме формализации отношений естественного языка, в частности, процесса формирования окончаний непритяжательных, полных имен прилагательных.

Бинаризация

На практике, при обработке информации или построении модели какого-нибудь процесса, приходится обрабатывать многомерные отношения. С помощью какой бы модели данных мы не попытались это реализовать, получаются довольно сложные конструкции. Однако можно попытаться представить многоместное отношение в виде комбинации более простых отношений, т.е. произвести декомпозицию, что значительно упростит работу с ними, а также позволит осуществлять параллельную обработку данных.

Рассмотрим процесс декомпозиции с точки зрения такой модели данных, как алгебра конечных предикатов. В терминах данной модели многоместное отношение можно описать с помощью *n*-арного предиката $\boldsymbol{P}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n})$, определенного на декартовом произведении множеств:

$$\mathrm{A} = \mathrm{A}_1 \times \mathrm{A}_2 \times ... \times \mathrm{A}_\mathrm{n},$$

где $\mathrm{x}_1 \in \mathrm{A}_1, \mathrm{x}_2 \in \mathrm{A}_2, ..., \mathrm{x}_\mathrm{n} \in \mathrm{A}_\mathrm{n}$. Идея декартовой декомпозиции состоит в том, что *n*-арный предикат представляется в виде конъюнкции бинарных предикатов, отсюда название *бинаризация* – представление *n*-арного предиката в виде системы бинарных предикатов. Восстановление предиката $\boldsymbol{P}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n})$ из системы бинарных предикатов называется его декартовой композицией – *ребинаризацией.*

Процесс бинаризации состоит в следующем. Каждому набору $(\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_\mathrm{n}) \in \mathbf{P}$, где $\mathbf{P}$ есть *n*-местное отношение, для которого задан предикат $\boldsymbol{P}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n})$, ставим в соответствие свое единственное имя. Из всех имен образуем множество *B*. Вводим переменное имя $\mathrm{y} \in B$ и функцию $\mathrm{y} = \mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n})$ ($S: \mathrm{P} \to B$), определяющую имя любого набора из $\mathbf{P}$.

Функции $\mathrm{y} = \mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n})$ соответствует некоторый предикат $\mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n}, \mathrm{y})$, задающий пространство $\boldsymbol{S}$ в координатной системе *A*. Предикат $\mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n}, \mathrm{y})$ определен на $A \times B$, имена y рассматриваем как векторы пространства *A*, а соответствующие им наборы $(\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_\mathrm{n})$ – как их координатное представление.

Предикат $\mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n}, \mathrm{y})$ однозначен (каждому набору отвечает не более одного имени), инъективен (каждому имени соответствует не более одного набора), сюръективен (каждому имени отвечает хотя бы один набор), поэтому пространство $\boldsymbol{S}$ – квазидекартово.

Любой инъективный предикат $\mathrm{S}(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_\mathrm{n}, \mathrm{y})$ раскладывается в конъюнкцию его проекционных предикатов $\boldsymbol{G}_1, \boldsymbol{G}_2, ..., \boldsymbol{G}_\mathrm{n}$:

$$\mathrm{S}(x_1, x_2, ..., x_n, y) = \mathrm{G}_1(y, x_1) \wedge \wedge \mathrm{G}_2(y, x_2) \wedge ... \wedge \mathrm{G}_\mathrm{n}(y, x_n) \quad . \tag{1}$$

Предикаты $G_1, G_2, ..., G_n$ определены на

$$B \times A_1, B \times A_2, ..., B \times A_n .$$

В результате декартовой декомпозиции n+1-арный предикат S заменяется равносильной ему системой бинарных предикатов $G_i (i = \overline{1,n})$. Следует отметить тот факт, что бинаризацию предиката $P(x_1, x_2, ..., x_n)$ мы провели через его усложнение, так как вводили дополнительную переменную y и n+1-арный предикат $S(x_1, x_2, ..., x_n, y)$, а только потом осуществляли его декомпозицию.

Однако возникает вопрос: на самом ли деле достигнута таким способом декомпозиция предиката $P(x_1, x_2, ..., x_n)$? Ведь мы производили декомпозицию не предиката P, а какого-то совсем другого предиката S. Да, достигнута, поскольку от предикатов $G_1, G_2, ..., G_n$ всегда можно возвратиться к предикату P.

Восстановление предиката $P(x_1, x_2, ..., x_n)$ по предикатам $G_1, G_2, ..., G_n$, т.е. декартова композиция, осуществляется по формуле:

$$P(x_1, x_2, ..., x_n) = \exists y \in B\, S(x_1, x_2, ..., x_n, y) , \quad (2)$$

где предикат $S(x_1, x_2, ..., x_n, y)$ предварительно восстановлен по формуле (1).

Для перехода к предикату $P(x_1, x_2, ..., x_n)$ в предикате $S(x_1, x_2, ..., x_n, y)$ элиминируем переменную y, исключаем ее с помощью квантора существования.

Проекционные предикаты могут быть найдены по формуле:

$$\begin{gathered} G_i(y, x_i) = \exists x_1 \in A_1\, \exists x_2 \in A_2 \,..., \\ \exists x_{i-1} \in A_{i-1}\, \exists x_{i+1} \in A_{i+1} ... \exists x_n \in A_n, \\ S(x_1, x_2, ..., x_i, ..., x_n, y), \end{gathered} \quad (3)$$

$i = \overline{1,n}$.

Таким образом, осуществляется процесс бинаризации n-местного отношения.

Ре-бинаризация

Однако кроме проблемы упрощения многоместного отношения, существует задача создания формального описания какой-либо предметной области, т.е. именно построение предиката $P(x_1, x_2, ..., x_n)$. Иногда просто невозможно сразу учесть взаимосвязь всех факторов в совокупности и построить адекватную модель реального процесса. Для решения данной проблемы предложен следующий подход, который также сводится к бинарным предикатам, т.е. рассмотрим процесс ре-бинаризации.

Пусть задано пространство $A = A_1 \times A_2 \times ... \times A_n$, а также существует набор предметных переменных $x_1, x_2, ..., x_n$, такой что

$$x_1 \in A_1, x_2 \in A_2, ..., x_n \in A_n .$$

В виде формул алгебры предикатов области определения x_i есть множества, описываемые формулой вида:

$$M(x_i) = \underset{a \in A_i}{\vee} x_i^a, i = 1,2,...n . \quad (4)$$

Пусть также на рассматриваемом пространстве существует отношение P, которое является подмножеством декартового произведения $A_1 \times A_2 \times ... \times A_n$, т.е. если $(a_1, a_2, ..., a_n) \in P$, то можно говорить, что $x_1, x_2, ..., x_n$ связаны отношением P, где $x_1 = a_1, x_2 = a_2, ..., x_n = a_n$, $(a_1, a_2, ..., a_n)$ – набор возможных значений переменных $x_1, x_2, ..., x_n$, который принадлежит множеству $A = A_1 \times A_2 \times .. \times A_n$ [1].

Кроме того, предметная область такова, что можно формально, в виде бинарных предикатов, описать отношения между любыми двумя переменными из набора $x_1, x_2, ..., x_n$. Тогда на пространстве A можно построить n-арный предикат $P'(x_1, x_2, ..., x_n)$, который будет минимальной совершенной дизъюнктивной нормальной формой предиката, описывающего n-местное отношение, существующее на данной предметной области. Предикат $P'(x_1, x_2, ..., x_n)$ строится по следующей формуле:

$$P'(x_1, x_2, ..., x_n) = \underset{i,j}{\wedge} F_{ij}(x_i, x_j), \;\; i,j = \overline{1,n} , \quad (5)$$

где $F_{ij}(x_i, x_j)$ $i,j = \overline{1,n}$ – предикаты, описывающие все возможные бинарные отношения между переменными x_i и x_j.

Итак, обсудим процесс построения предиката $P'(x_1, x_2, ..., x_n)$.

На введенном пространстве существует система отношений между x_i и x_j, I,j=1,...,n предметной переменной. Обозначим эти отношения F_{ij}, что означает: I-я переменная находится в некотором отношении с переменной x_j, или можно записать $x_i F_{ij} x_j$. В общем случае F_{ij} $I,j=1,...,n$ различные отношения, каждое из них определено на декартовом произведении $A_i \times A_j$.

Рассмотрим теперь введенные отношения $x_i F_{ij} x_j$ на $A_i \times A_j$, как преобразование $x_i = F_{ij}(x_j)$ предмета $x_i \in A_i$ в предмет $x_j \in A_j$, такое преобразование называется отображением [2]. Между отношениями $x_i F_{ij} x_j$ и отображениями $x_j = F_{ij}(x_i)$ существует взаимно-однозначное соответствие, выражаемое условием: для любых $x_i \in A_i$, $x_j \in \underline{A}_j$, если $x_i F_{ij} x_j$, то $x_j = F_{ij}(x_i)$; если же $x_i \overline{F_{ij}} x_j$, то $x_j \neq F_{ij}(x_i)$. Потребуем выполнения некоторых свойств для полученных отображений. Пусть рассматриваемое пространство таково, что отображения $x_i F_{ij} x_j$ являются всюду определенными:

$$\forall x_i \in A_i, \exists x_j \in A_j : x_j = Fij(x_i) ,$$

а также сюрьективными:

$$\forall x_j \in A_j, \exists x_i \in A_i : x_j = Fij(x_i) .$$

Выполнение сюрьективности эквивалентно всюду определенности только для обратного отображения.

Следующим шагом в построении структуры бинарных отношений является переход к предикатам, соответствующим введенным отношениям [1]. Переход осуществляется путем введения предиката, как характеристической функции $P(x_i, x_j) = \xi$, отображающей множество $A_i \times A_j$ в множество $\Sigma = \{0,1\}$, т.е. если x_i и x_j находятся в отношении F_{ij}, то $P(x_i, x_j) = 1$, иначе $P(x_i, x_j) = 0$. Обозначим эти предикаты $P_{ij}(x_i, x_j)$. Введенные таким образом предикаты будут иметь вид, например:

$$P_{ij}(x_i, x_j) = x_i^{a1} x_j^{b1} \vee x_i^{a3} x_j^{b2} \vee \ldots \vee x_i^{an} x_j^{bm},$$

где a_j – элементы множества A_i, а b_j – элементы множества A_j, т.е. предикат $P_{ij}(x_i, x_j)$ представляет собой дизъюнкцию возможных на данной предметной области пар значений переменных x_i и x_j, соответственно из множеств A_i и A_j.

Введем понятие тождественно истинного предиката. Значение такого предиката $P_{ij}(x_i, x_j)$ при любых наборах значений переменных x_i и x_j будет равно 1. На рассматриваемом пространстве такой предикат будет конъюнкцией предикатов, соответствующих областям определения переменных x_i и x_j:

$$P_{ij}(x_i, x_j) = M(x_i) \wedge M(x_j) \equiv 1, \qquad (6)$$

для любых наборов значений переменных x_i и x_j.

Рассмотрим теперь следующую конструкцию, образованную коньюнкцией всех возможных на данном пространстве бинарных отношений, выраженных соответствующими предикатами. При размерности пространства *n*, образованного системой переменных $x_1, x_2, \ldots, x_n$, определенных на $A = A_1 \times A_2 \times \ldots \times A_n$, соответственно, таких предикатов будет n^2. В результате такой конъюнкции получили *n*-арный предикат, которому соответствует некоторое отношение. Смысл этого предиката в том, что он определяет существование в данном *n*-мерном пространстве набора из *n* значений переменных $x_1, x_2, \ldots, x_n$. Таким образом, мы получили *n*-местное отношение в виде конъюнкции бинарных отношений.

Однако при размерности пространства больше 3 такая конструкция очень громоздка и избыточна. Приведем некоторый алгоритм, позволяющий минимизировать такую формулу:

$$\boldsymbol{P'}(x_1, x_2, \ldots, x_n) = \bigwedge_{i,j} F_{ij}(x_i, x_j), \quad i,j = \overline{1,n}. \qquad (7)$$

Строим таблицу, которая будет устанавливать характер отношения между переменными x_i и x_j.

– По вертикали и горизонтали расположены переменные $x_1, x_2, \ldots, x_n$.

– На пересечении i-й строки j-го столбца стоит идентификатор соответствующего бинарного отношения F_{ij}.

В таблице наглядно виден процесс упрощения. Далее проведем ее преобразование, в результате которого полученный по таблице предикат будет значительно проще построенного по формуле (7) $\boldsymbol{P'}(x_1, x_2, \ldots, x_n)$:

	x_1	x_2	$\cdots$	x_n
x_1	F_{11}	F_{12}	$\cdots$	F_{1n}
x_2	F_{21}	F_{22}	$\cdots$	F_{2n}
…	…	…	…	…
x_n	F_{n1}	F_{n2}	$\cdots$	F_{nn}

1. Заменим все элементы таблицы, которые соответствуют отношению, описываемому предикатом, тождественно равным 1, прочерками.

Как минимум, таких отношений будет *n,* т.е. те отношения $x_j F_{ji} x_i$, где $i = j$, в силу тождественной истинности предиката:

$$P_{ii}(x_i, x_i) = M(x_i) \wedge M(x_i) \equiv 1. \qquad (8)$$

В результате таких преобразований получим конъюнкцию не более чем $n^2 - n$ предикатов.

2. Ввиду всюду определенности и сюрьективности можно рассматривать только половину таблицы, либо выше, либо ниже главной диагонали, т.е. отношение $x_i F_{ij} x_j$ равнозначно отношению $x_j F_{ji} x_i$.

Таким образом, количество предикатов, входящих в конъюнкцию, будет не более, чем $\dfrac{n^2 - n}{2}$.

После описанных преобразований таблицы построим по ней предикат $\boldsymbol{P'}(x_1, x_2, \ldots, x_n)$ путем образования конъюнкции из всех оставшихся в таблице предикатов. В итоге мы получили минимальную дизъюнктивную нормальную форму *n*-местного отношения, описанного *n*-арным предикатом.

Описанный метод позволяет построить *n*-местное отношение при заданных областях определения каждой предметной переменной и знании законов, существующих между любыми двумя предметными переменными. На практике существует множество задач, исходными данными для которых будут перечисленные выше условия.

Следует отметить, что полученный таким путем предикат $\boldsymbol{P'}(x_1, x_2, \ldots, x_n)$ описывает все возможные сочетания значений предметных переменных. Однако реальные отношения, которые существуют на этой предметной области, могут задавать меньшее количество сочетаний:

$$\boldsymbol{P}(x_1, x_2, \ldots, x_n) \subseteq \boldsymbol{P'}(x_1, x_2, \ldots, x_n), \qquad (9)$$

что означает – предикат – $\boldsymbol{P}$ содержится в предикате $\boldsymbol{P'}$.

Таким образом, для пространств, где нет так называемых “условных” сочетаний, процесс создания формального описания многопараметрического явления, т.е. построение *n*-местного отношения можно осуществлять путем ре-бинаризации.

Рассмотрим, для примера, процесс построения описания окончаний имен прилагательных, т.е. задачу формализации естественного языка.

Формальное описание естественного языка

Рассмотрим в качестве предметной области естественный язык. Попытаемся формально представить процесс формирования окончаний непритяжательных полных имен прилагательных с помощью бинарных отношений.

Для начала опишем математическую модель склонения полных непритяжательных имен прилагательных.

В русском языке наблюдается большая степень согласованности между смыслом, формой и структурой слова, т.е. существует своего рода морфологическое отношение, которое связывает между собой грамматические признаки, форму слова и другие признаки. Таким образом, проведя анализ русского языка, можно выявить те признаки, которые влияют на вид окончания. Поэтому данный процесс можно описать в виде предиката, который отображает множество сочетаний признаков на множество $\sum = \{0,1\}$. Если в отношении, описываемом этим предикатом, существует данный набор, то предикат равен 1, иначе нулю.

В результате анализа окончаний непритяжательных полных имен прилагательных выявлено одиннадцать основных признаков, определяющих вид окончания [3], т.е. необходимо получить предикат $\boldsymbol{P}'(x_1, x_2, ..., x_{11})$.

Рассмотрим эти признаки и введем следующие предметные переменные, обозначающие их: x_1 – первая буква окончания; x_2 – вторая буква окончания; x_3 – третья буква окончания; x_4 – последняя буква основы; x_5 – род; x_6 – число; x_7 – падеж; x_8 – ударность, при чем имеется в виду ударность самого окончания; x_9 – мягкость, определяется по буквам, входящим в окончание; x_{10} – современность; x_{11} – одушевленность.

Таким образом, решение задачи нахождения формы окончания в рамках определенного текста сводится к нахождению таких значений переменных $\mathrm{x}_i, i = 1,2,...,11$, при которых предикат $\boldsymbol{P}'(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_{11}) = 1$.

Предикат $\boldsymbol{P}'(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_{11})$ является формальным описанием сочетаний одиннадцати признаков, всех, которые возможны в русском языке.

Предикат $\boldsymbol{P}'(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_{11}) = \boldsymbol{P}'(\mathrm{x})$ определен на декартовом произведении множеств $\mathrm{M}_1, \mathrm{M}_2, ..., \mathrm{M}_{11}$, которые являются областями определения соответствующих переменных. Представим формальное описание некоторых из этих областей.

Согласно алгоритму ре-бинаризации необходимо построить бинарные отношения, существующие на данной предметной области, между всеми признаками.

Проведенные исследования показали, что многие предикаты, описывающие бинарные отношения между некоторыми признаками, являются тождественно-истинными.

Далее была построена таблица, проведена ее обработка, в результате чего выявлено следующее: чтобы описать все возможные сочетания этих признаков, достаточно 25 бинарных таблиц. Однако это будут все возможные сочетания, но с точки зрения русского языка некоторые наборы будут лишними, т.е. необходимо учитывать “условные” сочетания для построения именно предиката $\boldsymbol{P}(x_1, x_2, ..., x_{11})$, соответствующего правилам русского языка. Однако для некоторых признаков все же удается построить тринарные предикаты $\boldsymbol{P}'(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k)$, которые совпадают с $\boldsymbol{P}(\mathrm{x}_i, \mathrm{x}_j, \mathrm{x}_k)$. Например, предикат $\mathrm{P}(x_1, x_2, x_3)$, устанавливающий зависимость таких признаков, как первая, вторая и третья буквы окончания, найден в следующем виде:

$$\mathrm{P}(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3) = \mathrm{P}_{12}(\mathrm{x}_1, \mathrm{x}_2)\mathrm{P}_{13}(\mathrm{x}_1, \mathrm{x}_3)\mathrm{P}_{23}(\mathrm{x}_2, \mathrm{x}_3).$$

Он задает такие наборы значений признаков x_1, x_2 и x_3, которые соответствует реальным закономерностям русского языка.

Итак, подведем некоторые итоги. В результате анализа таблицы можно сделать вывод, что десятиместное отношение, описывающее окончания непритяжательных полных имен прилагательных, можно получить в виде конъюнкции 25 бинарных предикатов. Такой результат свидетельствует о том, что нахождение возможных сочетаний значений одиннадцати признаков – довольно трудоемкая задача, но реализуемая.

Заключение

Идея обработки информации в виде бинарных отношений находит свое отражение во многих моделях представления данных: бинарные деревья, бинарные отношения в многокритериальной оптимизации, реляционные модели и др. Эти концепции могут быть реализованы с помощью различных программных, технических и других средств. Так, идеи бинаризации и ре-бинаризации можно с успехом применять при организации реляционных баз данных, при схемном моделировании интеллектуальных процессов, при обработке статистических данных, а также при построении информационных систем обработки естественного языка.

Литература: 1.*Дударь З.В., Мельникова Р.В., Шабанов-Кушнаренко Ю.П.* Отношения как объекты формульного описания // Радиоэлектроника и информатика. 1998, № 1. С.115-119. 2.*Дударь З.В., Самуйлик И.Г.,Шабанов-Кушнаренко Ю.П.* Отображения как объекты формульного описания // Радиоэлектроника и информатика. 1998, №1. С. 56-60. 3.*Шабанов-Кушнаренко Ю.П., Бондаренко М.Ф.* Математическая модель склонения полных непритяжательных имен прилагательных // Информационный анализ. 1979, № 6. С. 10-12.

Поступила в редколлегию 11.04.2001

Рецензент: д-р техн. наук, проф. Шабанов-Кушнаренко Ю.П.

Синельникова Ольга Игоревна, студентка ХНУРЭ. Научные интересы: искусственный интеллект, методы оптимального управления. Увлечения: туризм. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 13-17-13.

Вечирская Ирина Дмитриевна, студентка ХНУРЭ. Научные интересы: искусственный интеллект, логическая математика, естественный язык. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 93-56-98.