

УДК 004.652.6

## ОБ ОДНОМ МЕТОДЕ АНАЛИЗА ДАННЫХ В ЗАДАЧЕ ПСИХОЛОГИЧЕСКОЙ ДИАГНОСТИКИ

**А. А. Ключа,**

доктор психол. наук

Главное управление кадров Министерства обороны РФ

**Т. Ю. Морозова,**

канд. физ.-мат. наук, доцент

Московский государственный университет приборостроения и информатики

Приводится метод построения модели предметной области на основе интеллектуального анализа данных. Метод базируется на теории решеток Биркгофа и представляет сформировавшийся в последнее время логико-алгебраический подход, известный как формальный концептуальный анализ. Метод применен к структурированию и формированию логических правил для установления диагноза при клинико-психологическом обследовании.

*This paper presents a way to build the initial model of a domain using a knowledge based system. The method is based on the Birkhoff's lattice theory and represents the recently formed logic-algebra approach known as Formal Conceptual Analysis. We will apply this technique to structure and formulate the logical rules used to state the diagnosis during clinical and psychological inspections.*

Методы интеллектуального анализа данных (Data Mining) [1] применяются для автоматического обнаружения эмпирических закономерностей и использования их при решении задач классификации, распознавания образов и прогнозирования. Особенность этих методов состоит в их ориентации на задачи, для которых использование традиционных статистических методов вызывает большие затруднения. Имеются в виду задачи анализа данных очень большого объема; пораженных шумами; с признаками, измеренными в разнотипных шкалах; при отсутствии оснований для выдвижения гипотез о законах распределения плохо обусловленных таблиц (количество признаков сравнимо с количеством объектов) и т. д.

Целью технологии анализа данных является производство нового знания, выявление отношения в данных. К методам анализа данных следует отнести так называемый формальный концептуальный анализ (ФКА), недостаточно освещенный в отечественной научной литературе и, возможно, поэтому не получивший широкого применения в задачах структурирования данных и формирования баз данных. Формальный концептуальный анализ, введенный Рудольфом Вилле [2], является математическим подходом к анализу данных, базирующимся на теории решеток Биркгофа [3]. Он позволяет получить из неструктурированной информации структурированную. Может широко использоваться в прикладных областях, например в психологии.

Для введения ФКА прежде всего необходимо определить термин *контекст* или *формальный контекст*. Формальный контекст — это тройка  $(G, M, I)$ , где  $G$  — множество объектов;  $M$  — множество атрибутов и  $I$  — бинарные отношения между объектами и атрибутами:  $I \subseteq G \times M$ .

Следующие определения будут полезными для дальнейшего изложения:

1. На прямом произведении  $G \times M$  двух множеств существует частичный порядок, если  $(x_1, y_1) \leq (x_2, y_2)$  тогда и только тогда, когда  $x_1 \leq x_2$  в  $G$  и  $y_1 \leq y_2$  в  $M$ .

2. Решеткой называется множество  $L$ , в котором любые два его элемента имеют точную верхнюю грань, т. е. «объединение»  $x \vee y$ , и точную нижнюю грань, или «пересечение»  $x \wedge y$ .

Представим контекст психических расстройств в виде таблицы, в которой цифрами обозначены различные формы шизофрении: 1 — параноидальная, 2 — кататоническая, 3 — гебефреническая, 4 — простая, 5 — приступообразная, 6 — фебрильная; а буквами русского алфавита — признаки заболеваний в виде множества атрибутов  $M$ .

Таблица может быть интерпретирована следующим образом. Каждый символ «+» помечает пару, являющуюся элементом инцидентного отношения  $I$ .

Например, бинарное отношение (параноидальная шизофрения, 1) означает, что заболевание параноидальная шизофрения (объект) обладает

Атрибут		Объект					
		1	2	3	4	5	6
Возрастные особенности (А)		+	+	+			
Развитие	Стремительное (Б)					+	+
	Другое (В)	+	+	+			
Галлюцинации	Стойкие (Г)	+			+		
	Эпизодические (Д)			+			
Особенности речи (Е)			+		+		
Сознание	Онейроидное помрачение сознания (Ж)		+				
	Стойкий бред (З)	+	+				
	Другое (И)	+	+	+	+	+	+
Эмоции	Специфические (К)		+			+	
	Неспецифические (Л)	+	+	+	+	+	+
Моторика	Специфическая (М)		+				
	Неспецифическая (Н)	+	+	+	+	+	+
Поведение	Специфическое (О)			+			+
	Неспецифическое (П)	+	+	+	+	+	+
Потеря контакта с окружающим миром (Р)							+
Соматические проявления (С)							+

симптомом «наличие у больного стойких галлюцинаций различных типов» (атрибут). Таким образом,  $(g, m) \in I$  означает, что «объект  $g$  обладает свойством  $m$ ».

Главным понятием в ФКА является *формальный концепт*. Концепт  $(A, N)$  определяет пару — объект  $A \subseteq G$  и атрибут  $N \subseteq M$ , которые удовлетворяют некоторым условиям.  $A$  называют экстен- том,  $N$  — интен- том концепта, а множество всех свойств, которыми они обладают, — содержанием (интенционалом). Чтобы определить необходи- мость и достаточность условий для формального концепта, представим два оператора, допустив  $A \subseteq G$ :

$$A' = \{m \in M \mid \forall g \in A : (g, m) \in I\},$$

и соответственно  $N \subseteq M$ :

$$N' = \{g \in G \mid \forall m \in N : (g, m) \in I\}.$$

Приведенные определения означают, что мно- жество  $A'$  содержит все атрибуты, которые явля- ются общими для всех объектов  $A$ , а множество  $N'$  есть множество всех объектов, которые обладают всеми свойствами множества  $N$ .

Тогда пара  $(A, N)$  есть формальный концепт, если и только если

$$A' = N \text{ и } A = N'.$$

Это свойство говорит о том, что все объекты концепта содержат все его атрибуты. Это свиде- тельствует о том, что такое заболевание как пара- ноидальная шизофрения характеризуется в пер- вую очередь нарушениями сознания и сферы вос- приятия, поражает эмоциональную сферу, двига- тельную сферу, воздействует на мотивацию и волю и как следствие вызывает изменения в поведении.

Для формальных концептов природа отноше- ния подконцепт/надконцепт может быть опреде- лена следующим образом:

$$(A_1, N_1) \leq (A_2, N_2) \Leftrightarrow A_1 \subseteq A_2, N_1 \subseteq N_2.$$

Это отношение выявляет дуализм между атри- бутами и объектами концептов. Концепт  $C_1 = (A_1, N_1)$  является подконцептом концепта  $C_2 = (A_2, N_2)$ , если множество его объектов являются подмноже- ством объектов  $C_2$ . Таким образом, множество всех формальных концептов образуют так называемую концептуальную решетку.

Если контекст задан тройкой  $(G, M, I)$ , то инфинум такой решетки образуется множеством  $\{\emptyset, M\}$ , супремум формируется множеством  $(G, \emptyset)$ .

Линейная диаграмма является графическим представлением концептуальной решетки. Она позволяет исследовать и интерпретировать отношения между концептами, объектами и признаками, является эквивалентным представлением контекста, т. е. она содержит точно такую же информацию, как таблица отношений, в которых каждому узлу соответствует концепт из данного контекста.

На диаграмме каждый объект обладает свойствами, приписанными узлу, и свойствами узлов, с которыми этот узел связан дугами снизу вверх. С другой стороны, учитывая дуализм между объектами и свойствами (атрибутами), относительно свойств можно утверждать, что каждое свойство соответствует объектам, приписанным данному узлу, и тем объектам, с узлами которых данный узел связан дугами сверху вниз.

Приведем пример построения решетки концептов для анализа состояния пациента. В таблице задан формальный контекст  $K = (G, M, I)$ , где  $G$  — множество состояний,  $M$  — их свойства,  $I$  — бинарное отношение между состояниями и свойствами. При построении решетки совпадающие столбцы таблицы можно интерпретировать как наличие одного или/и другого признака. Поэтому на линейной диаграмме признаки заключены в скобки. Эта ситуация также может свидетельствовать о линейной зависимости между столбцами.

На рисунке изображена концептуальная решетка контекста «наличие психического расстройства у пациента».

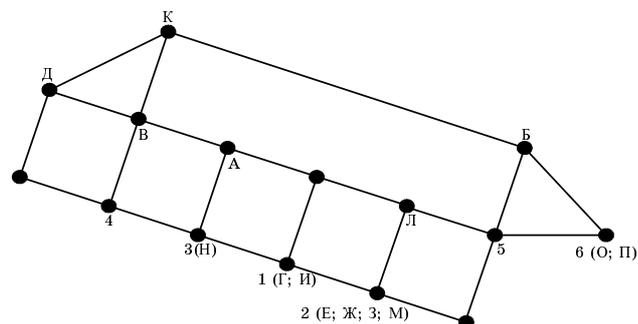
Граф состоит из узлов, которые представляют собой концепты, и ребер, соединяющих эти узлы. Два узла  $C_1$  и  $C_2$  соединены тогда и только тогда, когда  $C_1 \leq C_2$  и нет такого концепта  $C_3$ , что  $C_1 \leq C_3 \leq C_2$ .

Каждый объект и атрибут введен в граф только один раз. Атрибуты и объекты распространяются вдоль граней графа, как своего рода наследование. Атрибуты распространяются вдоль граней к основанию графа. Таким образом, высший элемент графа (верхняя грань контекста) соответствует  $\{G, \emptyset\}$ , где  $G$  — множество объектов. Элемент основания графа (нижняя грань контекста) соответствует  $\{\emptyset, M\}$ , где  $M$  — множество атрибутов.

Имена атрибута отмечаются буквами, а имена объекта отмечены цифрами около узла графа.

Таким образом, граф показывает связи между объектами и атрибутами.

По решетке можно проследить все свойства, которыми обладает то или иное состояние, — это множество всех свойств, лежащих выше узла, по-



■ Наличие психического расстройства у пациента

меченного названием состояние. Каждый узел решетки соответствует концепту.

Заметим, что если для всех объектов контекста, для которых справедливо некоторое свойство  $X$ , справедливо также некоторое свойство  $Y$ , то является истинной и импликация. Иными словами, если импликация  $X \rightarrow Y$  истинна для контекста  $K = (G, M, I)$  и любому объекту  $g \subseteq G$  применим каждый признак из посылки  $X$ , то к нему применим также признак из заключения импликации  $Y$ , где  $X \subseteq M$  и  $Y \subseteq M$ .

Проблема данного подхода состоит в том, что большое количество признаков влечет за собой большой размер таблицы. Другая проблема в том, что результирующая таблица не может содержать полной информации о каждом объекте и, кроме того, информация может быть противоречивой. Противоречия обнаруживаются при непосредственном рассмотрении формального контекста. Эти противоречия решаются при консультации со специалистом. Данная проблема не может быть решена, например, при наличии в модели некоторой ошибки.

Несмотря на это ФКА делает связи между понятиями (концепциями) явными и тем самым помогает из неструктурированной информации получить структурированную, что позволяет делать выводы и принимать решения.

## Литература

1. Дюк В., Самойленко А. Data mining. СПб.: Питер, 2001. 505 с.
2. Биркгоф Г. Теория решеток. М.: Наука, 1984. 337 с.
3. Ganter B., Wille R. Formale concept analysis: mathematical foundation. New York: Springer—Verlag, 1997. 93 с.