



Министерство науки и высшего образования Российской
Федерации Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Московский государственный технический университет имени
Н.Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

Отчет по лабораторным работам №4-6

По курсу

«Методы машинного обучения в АСОИУ»

Выполнил:

ИУ5-22М Барышников М.И.

14.02.2024

Проверил:

Балашов А.М.

2024 г.

В современных реалиях, одновременно с быстрым развитием информационной инфраструктуры становится актуальным направление хранения и обработки больших данных.

Необходимость оперативного доступа к желаемой информации увеличивает нагрузку на вычислительные машины и увеличивает время ожидания обработки пользовательских запросов. По этой причине ученые разрабатывают новые подходы для структурирования и анализа данных больших объемов и разных форматов. Повышение эффективности поддержки принятия решений путем оперативной аналитической обработки и предоставления релевантной информации из больших массивов разнородных данных на текущее время является весьма актуальной задачей развития информационных технологий [6].

Технологии искусственного интеллекта широко применяются во многих областях деятельности человека. Среди множества математических подходов области искусственного интеллекта активно набирает популярность метод анализа формальных понятий (АФП) [3], который основывается на классической теории алгебраических решеток. Анализ формальных понятий совмещает в себе мощный математический аппарат и философию, которые нашли широкое применение в решении прикладных задач. Стремительное развитие подхода обусловлено изящностью теории и простотой реализации алгоритмов. Данный метод получил широкое применение в задачах классификации, информационного поиска, построения онтологий, формирования рекомендаций.

Анализ формальных понятий (АФП) [3] — это метод анализа данных, который является прикладной ветвью алгебраической теории решеток. АФП позволяет пользователю анализировать зависимости между подмножествами объектов и признаков. В качестве отношения «объект - признак» могут выступать данные, которые описывают любую область человеческой деятельности. Пары взаимосвязанных множеств объектов и признаков образуют решетку

понятий, частично упорядоченную отношением «быть более общим понятием чем».

Фактически анализ формальных понятий имеет дело с данными в объектно-признаковой форме, а формальные понятия, определенные с помощью соответствия Галуа, представляют собой пары множеств вида (объем, содержание), им в точности до перестановки строк и столбцов соответствуют максимальные прямоугольники в таблице объект-признак. Основными достоинствами такого определения понятия являются соответствие традиционным представлениям о понятиях используемым в философии: 1) понятие — это пара вида (объем, содержание), 2) при уменьшении объема понятия увеличивается его содержание и наоборот, 3) понятия иерархически упорядочены по отношению, от общего к частому.

Формальные понятия – это понятия, которые определяются исключительно по своим признакам и характеристикам. Они не зависят от конкретных объектов или явлений, а являются абстрактными. Анализ формальных понятий представляет собой процесс разбора и изучения этих понятий.

Основным инструментом анализа формальных понятий является таблица понятий. Она состоит из двух частей: верхней и нижней. Верхняя часть таблицы содержит перечень всех признаков, которые могут быть присущи объекту, относящемуся к данному понятию. Нижняя часть таблицы содержит список объектов, относящихся к данному понятию.

Для того чтобы заполнить таблицу понятий, необходимо провести анализ всех возможных признаков объекта. При этом следует учитывать, что не все признаки могут быть присущи всем объектам данного понятия. Некоторые признаки могут быть общими только для подмножества объектов.

После заполнения таблицы понятий можно провести анализ отношений между объектами. Для этого используются операции формальной

концептуальной анализа: объединение, пересечение и дополнение. Они позволяют выделить подмножества объектов, которые имеют общие признаки.

Важным элементом анализа формальных понятий является определение семантических свойств понятий. Семантические свойства понятий могут быть выражены в виде отношений между ними. Например, одно понятие может быть подмножеством другого или иметь с ним общие элементы.

Таким образом, анализ формальных понятий представляет собой сложный процесс, который включает в себя разбор всех признаков объектов, определение отношений между ними и выделение семантических свойств понятий.

Концепция метода АФП

С точки зрения философии науки, формальное понятие является естественным способом описания группы объектов через их характерные свойства, которое определяется объемом – величиной всех объектов, принадлежащих понятию и содержанием – совокупностью всех признаков, общих для объектов [1-4].

Важным преимуществом метода анализа формальных понятий является его способность выявлять скрытые связи и тенденции в данных, которые могут быть упущены при традиционном анализе. Кроме того, он позволяет автоматизировать процесс анализа данных и сократить время, необходимое для принятия решений.

Первая задача АФП заключается в формировании формального контекста в виде тройки из множеств объектов, признаков и отношений их инцидентности. Обычно формальный контекст представляют в виде объектно-признаковой таблицы, в которой столбцам соответствуют признаки, а строками – объекты. Наличие крестика в таблице на пересечении столбца и строки означает вхождение соответствующей пары «объект, признак» в отношение инцидентности. Воздействуя на контекст оператором Галуа, представленным парой отображений объектов на признаки, и признаков на объекты, получаем (формальные) понятия. Для упорядочивания понятий строится решетка понятий. Вершинами решетки соответствуют (формальные) понятия, с соответствующими им объемами и содержаниями, а ребрами – отношения соседства между самими понятиями.

Отношения между формальными понятиями определяются на основе их свойств. Например, два понятия могут быть связаны отношением включения, если одно понятие является подмножеством другого. Также могут

существовать отношения эквивалентности, когда два понятия имеют одинаковые свойства.

Теория АФП успешно применяется для классификации, информационного поиска, построения онтологий, формирования рекомендаций и прочие [5-7].

Математический базис, используемый в анализе формальных понятий представлен в следующих ключевых понятиях:

- замыкания
- формальный контекст;
- соответствия Галуа;
- формальное понятие;
- решётка понятий;

Формальный контекст

Формальный контекст — это математическая модель, которая представляет собой таблицу, состоящую из трех частей: множества объектов, множества признаков и отношения между ними. Каждый объект может иметь некоторые признаки, и каждый признак может быть присущ только определенным объектам.

Отношение между объектами и признаками может быть двух видов: присутствие и отсутствие.

Контекстом в АФП называют тройку $K = (G, M, I)$, где G — множество объектов, M — множество признаков, а отношение $I \subseteq G \times M$ говорит о том, какие объекты какими признаками обладают.

Для произвольных $A \subseteq G$ и $B \subseteq M$ определены соответствия Галуа:

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

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

Которые задают соответствия между частично упорядоченными множествами объектов и признаков.

Оператор $"$ (двукратное применение оператора $'$) является оператором замыкания и отвечает следующим свойствам:

- идемпотентен ($A''' = A''$),
- монотонен ($A \subseteq B$ влечет $A'' \subseteq B''$),
- экстенсивен ($A \subseteq A''$).

Множество объектов $A \subseteq G$, такое, что $A'' = A$, называется замкнутым. Аналогично для замкнутых множеств признаков — подмножеств множества M . Пара множеств (A, B) , таких, что:

- $A \subseteq G$,
- $B \subseteq M$,
- $A' = B$,
- $B' = A$,

называется формальным понятием контекста K . Множества A и B замкнуты и называются объемом и содержанием формального понятия (A, B) соответственно. Для множества объектов A множество их общих признаков A' служит описанием сходства объектов из множества A , а замкнутое множество A'' является кластером сходных объектов (с множеством общих признаков A').

Отношение “быть более общим понятием” задается следующим образом: $(A, B) \geq (C, D)$ тогда и только тогда, когда $A \supseteq C$. Понятия формального контекста $K = (G, M, I)$, упорядоченные по вложению объемов образуют решетку $B(G, M, I)$, называемую решеткой понятий.

Отношения между объектами и признаками определяются на основе наличия или отсутствия признака у объекта. Если объект обладает определенным признаком, то мы говорим, что между ними существует отношение принадлежности. Если объект не обладает признаком, то между ними существует отношение непринадлежности.

Формальный контекст может быть представлен в виде таблицы (рис. 1), где строки соответствуют объектам, а столбцы - признакам. В каждой ячейке таблицы может быть значение 1 или 0, которое указывает наличие или отсутствие признака у объекта.

	bookable	rentable	driveable	rideable	joinable
hotel	x				
apartment	x	x			
car	x	x	x		
bike	x	x	x	x	
excursion	x				x
trip	x				x

Рисунок 1 – формальный контекст

Формальный контекст может быть использован для построения решетки понятий, которая позволяет визуально представить отношения между понятиями. Также он может быть использован для классификации объектов и понятий, а также для построения систем рекомендаций на основе анализа данных.

Поиск закономерностей

Импликацией формального контекста $K = (G, M, I)$ в анализе формальных понятий называется признаковая зависимость вида $A \rightarrow B$, где $A, B \subseteq M$, при условии, что все объекты, обладающие A , также обладают всеми признаками из B , т.е. $A' \subseteq B'$. Импликация в АФП является частным случаем такой признаковой зависимости как ассоциативное правило в разработке данных, это в точности ассоциативное правило с достоверностью (confidence) равной 1. В свою очередь, ассоциативные правила изучались в АФП задолго до их появления в сообществе разработки данных под названием частичные импликации [1].

Связь импликаций и функциональных зависимостей позволила использовать т. н. базис импликаций Дюкена-Гига для компактного представления функциональных зависимостей (см. теорию баз данных) виде их ограниченного множества, из которого все оставшиеся функциональные зависимости данного многозначного контекста (таблицы данных) выводимы по правилам Армстронга [2].

Решетка понятий

Решетка понятий – это графическое представление формального контекста, которое позволяет визуально выявить отношения между понятиями. В решетке понятий каждое понятие представлено вершиной, а отношения между ними – ребрами.

Решетка понятий состоит из упорядоченного множества понятий, которые связаны между собой отношением включения. Каждое понятие представляет собой набор объектов, которые имеют общие признаки или свойства. Например, если мы анализируем данные о продажах в магазине, то каждое понятие может представлять собой группу товаров, которые покупаются вместе.

Решетка понятий имеет две основные операции - расширение и интенсификация. Расширение понятия происходит при добавлении новых объектов, которые имеют общие признаки с уже существующими объектами в данном понятии. Интенсификация понятия происходит при добавлении новых признаков, которые связаны с уже существующими объектами в данном понятии.

Самый верхний и самый нижний узлы в решетке понятий являются особыми. Верхний узел содержит все формальные объекты (объем формальных понятий). Нижняя концепция содержит все формальные атрибуты (содержание формальных понятий).

Чтобы найти объем формальной концепции, нужно проследить все пути, которые ведут вниз от узла и выделить все попавшиеся на этом пути формальные объекты.

Чтобы восстановить содержание формального понятия, нужно проследить все пути, которые ведут снизу вверх от узла и выделить все встреченные формальные атрибуты.

Решётка понятий основана по понятии иерархии. Есть более общие формальные понятия и более частные, то есть вложенные формальные понятия. Отношение “подпонятие – понятие” является транзитивным, что означает, что подпонятие является частным по отношению к любому понятию, которого можно достичь, двигаясь от него вверх.

Решётка понятий — частично упорядоченное множество, в котором каждое двухэлементное подмножество имеет как точную верхнюю (sup), так и точную нижнюю (inf) грани. Отсюда вытекает существование этих граней для любых непустых конечных подмножеств. Пример решётки понятий представлен на рисунке 2.

$$\sup\{\{x\}, \{x, y\}\} = \{x, y\}$$

$$\inf\{\{x\}, \{x, y\}\} = \{x\}$$

$$\sup\{\{x\}, \{y, z\}\} = \{x, y, z\}$$

$$\inf\{\{x\}, \{y, z\}\} = \{\}$$

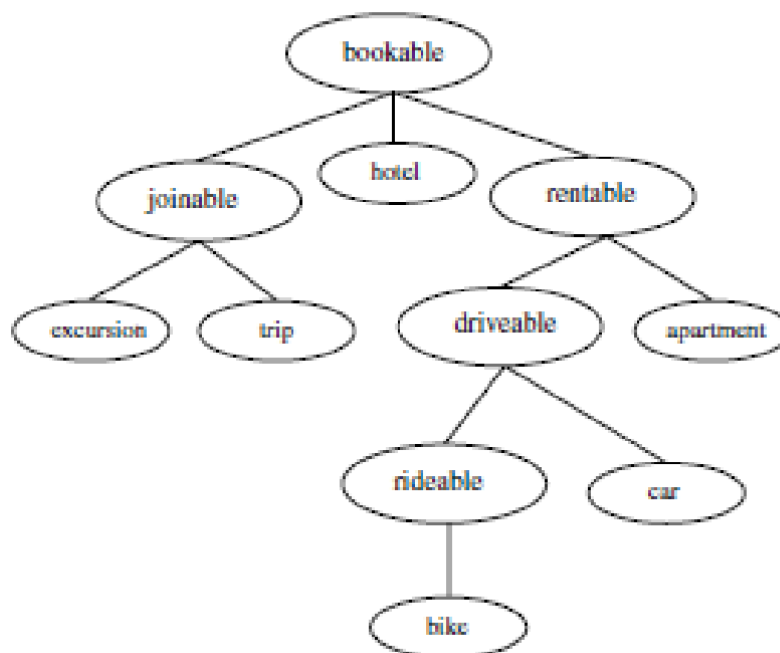


Рисунок 2 – решётка понятий

Довольно часто размер решетки велик, а ее структура сложна. Было предложено много методов: разделять набор элементов и получать вложенные линейные диаграммы или раскладывать решетку, используя соответствия или допуски, чтобы получить представления факторов или атласа. Методы визуализации являются ценными инструментами для интерпретации извлеченных знаний и становятся все более важными, особенно в контексте больших и сложных данных.

Применение метода АФП в решении задач

Одним из основных применений формального контекста решетки понятий является классификация объектов и понятий. Классификация позволяет выделить группы объектов и понятий, которые имеют общие признаки. Это может быть полезно, например, при анализе данных или при поиске решений в различных областях науки и техники.

Кроме того, формальный контекст решетки понятий может использоваться для построения систем рекомендаций. Например, на основе анализа покупок клиентов в интернет-магазине можно выделить группы товаров, которые часто покупают вместе, и предложить клиентам рекомендации на основе этих групп.

Информационный поиск.

Задача информационного поиска заключается в выделении из исходного неструктурированного текста информации, которая соответствовала бы определённому запросу или критерию. С помощью анализа формальных понятий данная задача может быть решена следующим образом.

Задача информационного поиска в случае АФП решается следующим образом: формируется формальный контекст, в котором множество объектов – это разнообразие текстов «мешок слов» и признаков, это набор ключевых слов. Тогда решетка будет состоять из упорядоченных концептов, включающих некоторый набор текстов, которым свойственен один набор ключевых слов. Соответственно, анализируя решетку сверху вниз, а именно переходя от вершины к вершине, мы можем определить источники информации, удовлетворяющие поисковому запросу.

Классификация.

Классификация является одним из подходов в теории машинного обучения и распознавания образов. Суть задачи заключается в разделении объектов по схожим признакам на заранее известное количество классов. На данный момент предложен ряд алгоритмов классификации, основанный на теории АФП.

Методы классификации на основе АФП опираются на базовые свойства решетки формальных понятий, а именно—выявление объектов с одинаковыми свойствами. Разметка объектов относительно принадлежности к определенному классу позволяет определить свойства, характерные (определяющие) для класса, и тем самым обнаружить новые знания. К недостаткам подхода можно отнести сложность интерпретации смешанных понятий, которые содержат объекты, отнесенные к разным классам в обучающей выборке.

Построение онтологических моделей

Формализация некоторой области знаний в виде концептуальных схем, состоящих из структурированных данных, осуществляется с помощью онтологий. Онтология включает в себя множество понятий — сущностей и отношений между ними. В свою очередь решетка понятий, обладающая таксономическими свойствами, пересекается с онтологической концепцией. В настоящее время популярными задачами, которые может решать анализ формальных понятий являются: извлечение онтологических понятий и выявление иерархий или сетей, построение и уточнение онтологий, исследование атрибутов в целях получения экспертных знаний [8], а также слияние онтологий.

Отбор значимых признаков

Классическим примером применения АФП для определения значимых признаков является Теория неточных множеств. Суть данной теории заключается в представлении множества объектов в виде множества подмножеств, определенных ключевыми свойствами объектов, а не конкретными объектами. Ученый дает понятие неточного множества как пары двух множеств - нижнего

и верхнего приближений, построенных из элементарных множеств объектов. Эта идея является ключевой для решения многих других задач: классификации, оценки зависимостей между признаками и классификацией объектов, определения степени такой зависимости, вычисления важности признаков, сокращения количества признаков и порождения решающих правил по исходным данным.

Формирование рекомендаций

Анализ формальных понятий применяется в рекомендательных системах. Множество задач интеллектуальной поддержки принятия решений опирается на рекомендательный подход. Действительно, реализация умных ассистентов помогает лицу принимающего решение определиться с выбором фильма, товара, услуги или мероприятия в соответствии с его интересами.

АФП широко применяется для решения задачи рыночной корзины. В данном подходе в формальном контексте признаками являются товары (предметы), а объектами – чеки (транзакции). По сути, импликация определяет отношения между набором предметов. Пространство поиска решетки позволяет анализировать смежные группы товаров путем просмотра объема и содержания вершин диаграммы решетки – формального понятия.