



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

Лабораторные работы №4-6

По курсу

«Методы машинного обучения в АСОИУ»

Выполнил:

ИУ5-22М Бабин А.С.

14.02.2024

Проверил:

Балашов А.М.

2024 г.

АФП на основе трипонятий

В классическом анализе формальных понятий есть две ключевые сущности: объекты и признаки. Одним из вариантов представления данных такой структуры являются таблицы или объектно-признаковые матрицы, где объектам соответствуют строки, а признакам – столбцы.

Однако такой подход несколько ограничивает возможности анализа, так как далеко не все реальные данные можно свести к такой структуре без потери какой-либо значимой и полезной для дальнейшего анализа информации с присутствующими зависимостями.

В связи с этим был разработан усовершенствованный метод АФП, в котором к уже упомянутым выше объектам и признакам добавляются ещё и условия.

Триадический формальный контекст $K = (G, M, B, Y)$ состоит из множеств G (объекты), M (признаки), B (условий) и тернарного отношения $Y \subseteq G \times M \times B$. Запись $(g, m, b) \in Y$ означает, что g имеет признак m при условии b .

Точно так же, как двоичные контексты часто описываются двумерными перекрестными таблицами, триадические контексты могут быть представлены трехмерными перекрестными таблицами [7].

Трипонятием контекста K называется такая тройка вида (X, Y, Z) , которая максимальна по вложению своих компонент, такая что $X \subseteq G$, $Y \subseteq M$, $Z \subseteq B$ и $X \times Y \times Z \subseteq Y$.

Чтобы теперь определить трикластер, необходимо ввести несколько определений операторов. Во-первых, в случае трикластеризации необходимы производные операторы, но только первого вида, сопоставляющие множеству элементов (например, объема) множества пар других элементов:

$$m' = \{ (g, b) \mid (g, m, b) \in Y \},$$

$$g' = \{ (m, b) \mid (g, m, b) \in Y \},$$

$$b' = \{ (g, m) \mid (g, m, b) \in Y \}.$$

Во-вторых, для тройки $(g, m, b) \in I$ дополнительно вводятся бокс-операторы [8]:

$$g^\square = \{g_i \mid (g_i, b_i) \in m' \text{ or } (g_i, m_i) \in b'\},$$

$$m^\square = \{m_i \mid (m_i, b_i) \in g' \text{ or } (g_i, m_i) \in b'\},$$

$$b^\square = \{b_i \mid (m_i, b_i) \in g' \text{ or } (g_i, b_i) \in m'\}.$$

Тогда трикластером называется тройка вида $T = (g, m, b)$, где $(g, m, b) \in Y$.

Предложенный метод трикластеризации является эффективным средством сокращения количества формальных трипонятий (время работы алгоритма и размер выхода полиномиальны), порождаемых для входных контекстов и может быть использован для решения различных задач анализа данных.

ДСМ-метод в терминах АФП

ДСМ-метод — метод автоматического порождения гипотез. ДСМ-метод оперирует сущностями трёх видов:

- объекты предметной области;
- свойства этих объектов;
- возможные причины свойств.

Предполагается, что объекты имеют структуру и причинами свойств объектов являются фрагменты этой структуры.

На вход ДСМ-метод получает некоторое множество изучаемых объектов и сведения об их структуре, о наличии или отсутствии у них определённых свойств, а результатом применения ДСМ-метода являются гипотезы двух возможных типов:

- гипотезы о связи определённых структурных фрагментов изучаемых объектов со свойствами, которыми они обладают;
- гипотезы о наличии или отсутствии целевых признаков у объектов, для которых изначально это было неизвестно, формируемые на основании установленной взаимосвязи между свойствами объектов и их структурными компонентами.

Теорию ДСМ метода удалось применить и в контексте анализа формальных понятий, несколько модифицировав первоначальный алгоритм. Основное изменение касается введения специального целевого признака объекта.

Пусть $K = (G, M, I)$ – формальный контекст, а $w \notin M$ представляет некий целевой признак объектов, отличный от признаков из множества M . Объекты делятся на три категории:

- положительные примеры, обладающие свойством w ;
- отрицательные примеры, не обладающие свойством w ;
- недоопределённые примеры, которые требуется классифицировать на наличие свойства w .

Для каждой категории строится свой контекст – положительный, отрицательный и недоопределённый. И далее, объединяя аппарат АФП и теорию ДСМ-метода, производится определение гипотез.

Благодаря установке связи между ДСМ-гипотезами и формальными понятиями появилась возможность использовать в АФП обучение по положительным и отрицательным примерам, а также рассматривать ДСМ-гипотезы с точки зрения теории решёток [9].

Приложения АФП

На сегодняшний день с помощью анализа формальных понятий возможно решение задач очень широкого круга задач из разных направлений.

Примеры расширений и приложений анализа формальных понятий:

- Бикластеризация;
- АФП в онтологическом моделировании;
- Построение, слияние, морфизмы и управление качеством онтологий;
- Ассоциативные правила для рекомендательных систем;
- Поиск дубликатов Интернет-документов;
- Таксономии групп пользователей веб-сайтов;
- Трикластеризация для фолксономий;
- АФП в социологии и образовании;
- АФП в машинном обучении;
- АФП в программной инженерии;
- АФП в информационном поиске;
- OLAP приложения АФП;
- Социальные интернет-сервисы и поиск сообществ;
- АФП в моделях контроля доступа;
- АФП в анализе текста.

Таким образом, анализ формальных понятий и его модификации представляет собой удобные модели для:

- построения и визуализации таксономий предметных областей;
- мультимодальной кластеризации объектов из предметных областей;
- компактного представления зависимостей в предметных областях через базисы импликаций и ассоциативных правил;
- представление ряда моделей машинного обучения;
- построения, пополнения, слияния и управлением качества онтологий;
- для приложений, требующих анализа объектно-признаковых данных.

Ниже подробнее описаны основные из направлений.

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.

Сначала формируется формальный контекст, в котором множество объектов – это разнообразие текстов «мешок слов» и признаков, это набор ключевых слов. Таким образом, построенная на основе этих данных решётка будет включать в себя концепты из некоторых наборов текстов, которым свойственен один и тот же список ключевых слов и фраз.

При анализе решётки от вершины к вершине сверху вниз возможно определять источники информации, подходящие под заданную тему или определённый поисковый запрос. Навигация по решётке и учёт иерархических зависимостей приводит к гибкому поиску. Кроме того, АФП позволяет преодолеть проблемы с отсутствием результатов поиска и избытком информации.

Классификация.

Методы классификации на основе анализа формальных понятий используют базовые свойства решетки формальных понятий для выявления объектов с одинаковыми характеристиками. Путем разметки объектов по их принадлежности к определенным классам можно определить характеристики, которые являются определяющими для каждого класса, что в свою очередь позволяет обнаружить новые знания. Однако этот подход имеет недостатки, такие как сложность интерпретации смешанных понятий, в которых содержатся объекты, отнесенные к разным классам в обучающей выборке [6].

Построение онтологических моделей

Знания из какой-либо предметной области можно формализовать с помощью онтологий – наборах понятий и отношений, которые описывают сущности и их связи в предметной области. Онтология определяет основные понятия и их свойства, а также правила и ограничения, которые применяются к этим понятиям.

Онтологические модели относятся к классу познавательных моделей [7]. Онтологические модели – это формальные структуры, которые используются для представления знаний о мире. Они позволяют описывать понятия, отношения и свойства объектов в предметной области. Онтологические модели могут быть представлены в виде графов, деревьев или сетей.

Анализ формальных понятий способен извлекать такие онтологические понятия, строить и уточнять онтологии, а также исследовать атрибуты для выявления экспертных знаний.

Кроме того, существует подход, позволяющий сливать онтологии. Это означает, что при нахождении объектов со схожими атрибутами, они будут объединяться и тем самым будет улучшаться структуризация онтологий предметной области.

Благодаря методу АФП возможно построение онтологических моделей не только для исследования существующих, но и создания новых онтологий на основе извлечённых знаний о рассматриваемой предметной области

Отбор значимых признаков

Благодаря отсечению избыточных признаков можно повысить производительность и эффективность работы моделей и выполнения запросов.

Анализ формальный понятий может применяться для определения наиболее значимых признаков в рассматриваемом наборе данных. АФП способен выделять ключевые признаки, характеризующие какое-либо множество объектов предметной области за счёт оценки объёма формальных понятий как показателя значимости вершин диаграммы решётки понятий.

Также к отбору значимых признаков можно отнести процесс аппроксимации, применяемый для неточных множеств. АФП преобразует многозначную информационную систему со множеством свойств объектов в формальный контекст с единственным свойством. А затем с помощью решётки понятий вычисляется наименьшая верхняя и наибольшая нижняя грани неточного множества.

Формирование рекомендаций

Анализ формальных понятий часто используется в рекомендательных системах. Задача формирования рекомендаций решается за счёт использования отношения частичного порядка на множестве концептов решётки. Благодаря импликации между объектами и содержанием концептов происходит выявление объектов со схожими характеристиками в заданном пространстве поиска.