

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЦЕНТРАЛЬНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ
ТЕХНІЧНИЙ УНІВЕРСИТЕТ

УДК 681.518.54

Тези доповідей

IV Міжнародної науково-практичної конференції

"Інформаційна безпека та комп'ютерні
технології"



15– 16 квітня 2021 р.

Кропивницький 2021

УДК 681.518.54

Матеріали IV Міжнародної науково-практичної конференції “Інформаційна безпека та комп’ютерні технології”: тези доповідей, 15 – 16 квітня 2021 р. – Кропивницький: ЦНТУ, 2021. – 81 с.

Наведені тези пленарних та секційних доповідей за теоретичними та практичними результатами наукових досліджень і розробок. Представлені результати теоретичних досліджень в галузях проектування інформаційних систем, технологій захисту інформації, використання сучасних інформаційних технологій в управлінні системами за різними галузями народного господарства.

Матеріали публікуються в авторській редакції.

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор.

© Центральноукраїнський національний
технічний університет, 2021

АНАЛИЗ ПРЕДПОЧТЕНИЙ НА ОСНОВЕ ТОРЦЕВОГО ПРОИЗВЕДЕНИЯ МАТРИЦ

Обработка больших данных является одним из важных направлений применения технологий искусственного интеллекта. В ряде приложений существенная роль при этом отводится поиску статистических закономерностей, например, в интересах предсказания действий на основе анализа предпочтений. В этой связи актуальной задачей является автоматизация процедуры вычисления предпочтительных шаблонов по всему набору предпочтений в интересах раннего распознавания намерений. Целью работы является описание метода решения задач статистического анализа предпочтений на основе предложенного автором в 1996 г. торцевого произведения матриц [1 - 3].

Предположим, что имеется 3 человека, указывающих свой предпочтительный объект в исходном множестве из четырёх объектов. При этом потребуем, чтобы в процессе выбора отдавать предпочтение какому-либо из объектов можно было лишь один раз. Для простоты рассмотрим три последовательных процедуры отдания предпочтений, пример которых представлен в табл. 1.

Таблица 1

Порядковый номер анкетированного	1-е предпочтение	2-е предпочтение	3-е предпочтение
Человек 1	Объект 3	Объект 1	Объект 4
Человек 2	Объект 2	Объект 3	Объект 1
Человек 3	Объект 2	Объект 4	Объект 1

Чтобы сохранить преемственность по отношению к решению задач анализа текста в качестве объектов выбора рассмотрим четыре слова, образующих текстовый фрагмент из трех предложений, рассмотренных в [4]: 1) I like math; 2) You like math; 3) I like you.

Составим для каждого из последовательных предпочтений так называемую матрицу инцидентности [4]. Ее строки будут соответствовать конкретному человеку, а столбцы - выбранному слову. При этом единичные элементы в каждой строке соответствуют слову, которому было отдано предпочтение, тогда как для всех остальных слов будут стоять нули. Количество столбцов должно соответствовать максимальному количеству слов в рассматриваемом текстовом множестве. В указанном выше фрагменте наибольшее количество слов - четыре. Прежде чем непосредственно перейти к матрицам инцидентности, составим с учетом сказанного для наглядности таблицу, соответствующую результатам последовательных предпочтений (табл. 2).

Таблица 2

Порядковый номер анкетированного	X ₁ =I	X ₂ =like	X ₃ =math	X ₄ =you
Первое предпочтение				
1	0	1	0	0
2	1	0	0	0
3	1	0	0	0
Второе предпочтение				
1	1	0	0	0
2	0	1	0	0
3	0	0	0	1
Третье предпочтение				
1	0	0	1	0
2	0	0	0	1
3	0	1	0	0

Отсюда, получим матрицы инцидентности вида:

$$X = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}; Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; Z = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Перейдем к задаче анализа сочетаний объектов в последовательности предпочтений. Как указано в [4], для этого необходимо воспользоваться торцевым произведением матриц. В частности, согласно [4], матрица совместной встречаемости для анализа тройных сочетаний может быть сформирована на основе исходной матрицы инцидентности и ее версии в виде торцевого произведения (символ \square):

$$C = X^T(Y \square Z) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}^T \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Полученная в результате матрица имеет размерность 16×4 и может быть представлена в виде 4 блоков, каждый из которых соответствует одному из слов в парных словосочетаниях (см. табл. 3).

Таблица 3

	Y ₁ =I				Y ₂ =like				Y ₃ =math				Y ₄ =you				
	Z ₁	Z ₂	Z ₃	Z ₄	Z ₁	Z ₂	Z ₃	Z ₄	Z ₁	Z ₂	Z ₃	Z ₄	Z ₁	Z ₂	Z ₃	Z ₄	
X ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X ₂	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X ₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X ₄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Цифры в роли элементов матрицы C характеризуют частоту появления конкретных троек слов в исследуемой последовательности предпочтений с учетом их порядка в последовательности. Например, из первого блока матрицы следует, что набор предпочтений (X₂Y₁Z₃) встречается однажды, равно как (X₁Y₂Z₄) и (X₁Y₄Z₂). Остальные комбинации предпочтений не наблюдались. Фактически табл. 4 отражает диаграммы предпочтений, выбранных пользователями.

Альтернативное решение рассмотренной задачи позволяет получить переход к тройному торцевому произведению матриц инцидентности: $I3 = X \square Y \square Z$. Для анализируемого фрагмента текста такое произведение приводит к 16-блочной матрице, состоящей из 4 кварталов блоков:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Первый блок полученной матрицы соответствует тройкам X₁Y₁Z_m, а последний - X₄Y₄Z_m (m=1,...,4). Совокупная статистика по всем предпочтениям может быть получена путем суммирования строк матрицы X □ Y □ Z, что эквивалентно операции умножения её на вектор-строку единиц:

$$\mathbf{1}^T I3 = \mathbf{1}^T (X \square Y \square Z).$$

Аналогичный результат может быть получен на основе квадратичной формы торцевого произведения: $\mathbf{1}^T (X \square Y \square Z) = \text{diag}[(X \square Y \square Z)^T (X \square Y \square Z)]$, где $\text{diag}(C)$ – вектор-строка из диагональных элементов матрицы C. Указанные соотношения позволяют получить все тот же набор предпочтений, в котором отсутствуют любые комбинации, содержащие X₃ и X₄, тогда как для X₁ и X₂ выбор однократно выпал на сочетания (X₁Y₂Z₄), (X₁Y₄Z₂) и (X₂Y₁Z₃). В общем случае наибольший результат суммирования строк матрицы X □ Y □ Z будет соответствовать комбинации объектов выбора, на которую пришлось максимальное количество предпочтений.

Представленный метод обработки результатов выбора позволяет узнать о “предпочтительных” шаблонах (паттернах) на основе первого предпочтения, например, “если кто-то выберет X в качестве первого предпочтения, то второй выбор, скорее всего, будет...”. Рассмотренная процедура вычисления шаблонов предпочтений по всей совокупности статистик позволяет автоматизировать соответствующий анализ. Это важно при большом количестве элементов выбора и соответствующих ему этапов, поскольку вручную сделать аналогичные подсчёты было бы сложно. В то же время, если предпочтения формулируются в виде коротких ответов из нескольких слов, то применению рассмотренного метода должна предшествовать предварительная токенизация словесного описания выбора.

Список литературы

1. Слюсар В.И. Торцевые произведения матриц в радиолокационных приложениях// Известия высших учебных заведений. Радиоэлектроника.- 1998. - Том 41, № 3.- С. 71 - 75.
2. Слюсар В.И. Семейство торцевых произведений матриц и его свойства// Кибернетика и системный анализ. – 1999.- Том 35; № 3.- С. 379-384.- DOI: 10.1007/BF02733426
3. Основы военно-технических исследований. Теория и приложения. Том. 2. Синтез средств информационного обеспечения вооружения и военной техники. / А.И. Миночкин, В.И. Рудаков, В.И. Слюсар. – Киев: «Гранма», 2012. – С. 7 – 98, 354 – 521.
4. Слюсар В.И. Применение торцевого произведения матриц в задачах обработки естественного языка. //Збірник наукових праць XIX Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТі3-2020». - Краматорськ. -2020. - С. 156 - 162. - DOI: 10.13140/RG.2.2.31568.53762.