

УДК 51-7

Н.С. Косило¹, Д.Б. Ельчанинов²¹Национальный технический университет «ХПИ», Харьков²Национальный институт стратегических исследований, Киев

ИЕРАРХИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ МУЛЬТИМНОЖЕСТВ В ПРОЦЕССЕ КОНСОЛИДАЦИИ ИНФОРМАЦИИ

Методы теории множеств рассматриваются как один из путей решения проблемы формализации процесса консолидации информации. Сопоставлены процесс консолидации информации и процедура иерархической кластеризации. Приведен пример использования процедуры иерархической кластеризации в процессе консолидации информации относительно фирм-конкурентов.

Ключевые слова: консолидация информации, множество, кластеризация, дендрограмма.

Введение

Постановка проблемы. Сегодня уже не нужно доказывать необходимость разведывательной деятельности в экономике. В конкурентной борьбе именно информация может обеспечить решающее преимущество организациям и предприятиям. Практика свидетельствует, что 95% информации можно получить из открытых источников [1]. Причем весомая часть полезной информации является качественной (не количественной). Поэтому важной научной задачей является развитие нового направления в теории принятия решений – вербального анализа решений, методы которого можно использовать при решении практических задач обработки качественной информации [2].

Анализ последних исследований и публикаций. Процесс консолидации информации акцентирует внимание на четырех последовательных этапах: выяснение, накопление и упорядочение исходных требований и данных; выявление необходимой информации (формирование информационных ре-

сурсов); анализ и синтез информации (добывание и приобретение знаний); своевременное обеспечение использования информации в необходимой форме для поддержки принятия решений (рис. 1) [3].

В работе [4] делается упор на необходимости разработки технологического обеспечения этого процесса в трех направлениях, по которым происходит развитие консолидированной информации: в производственной и непроизводственной сфере, а также в системах специального назначения. Несмотря на то, что в теории и практике информационного анализа используется множество формализованных методов, информационно-аналитическая деятельность до сих пор остается творческим процессом. Для упорядочения этого процесса предлагается использовать системологический подход [5], теорию постоянного развития [6], сбалансированную систему показателей [7] и методологию IDEF0 [8]. Особое внимание необходимо также уделять разработке и развитию информационных технологий и средств, которые поддерживают процесс консолидации информации [9].

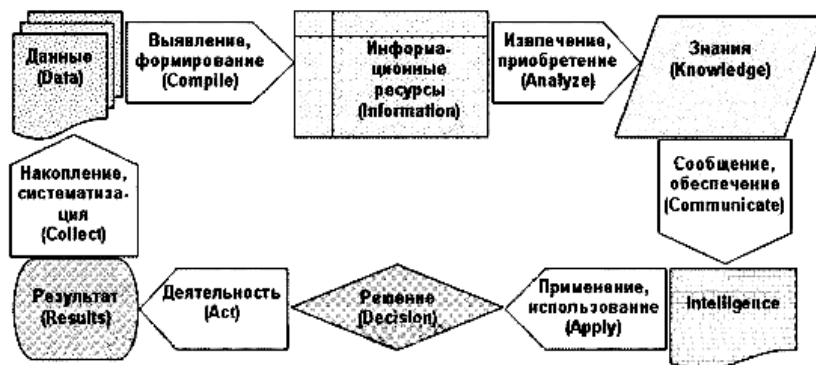


Рис. 1. Процесс консолидации информации

Отличительной особенностью процесса консолидации информации является множественность и повторяемость данных, которые описывают как рассматриваемые объекты, так и их свойства. С точки зрения математики такие многозначные объекты можно представить как мультимножества или мно-

жества с повторяемыми элементами. Например, мультимножество $M = \{a, a, b, b, b, c\}$ содержит два элемента a , три элемента b и один элемент c . Более компактная запись этого мультимножества $M = \{2a, 3b, 1c\}$, где число перед элементом указывает на так называемую «кратность вхождения» элемента в му-

льтимножество. Мультимножество можно рассматривать или как одну из отдельных форм множественного числа, или как самостоятельное понятие, более общее, чем множество [10]. Возможность многократного вхождения элементов в мультимножество создает новое качество, которое отличает мультимножество от обычных ординарных множеств и порождает существенно большее, чем у множеств, разнообразие видов и особенностей мультимножеств [11].

Одной из основных характеристик мультимножества является его мощность - количество экземпляров всех его элементов. Например, мощность мультимножества M (обозначается через $|M|$) равняется $|M| = 2 + 3 + 1 = 6$. Для определения отличия между мультимножествами используется формальная операция симметричной разности. Результатом этой операции есть мультимножество, у которого кратность каждого элемента равна модулю разности кратности соответствующих вычитаемых элементов мультимножеств.

Например, симметричная разность мультимножеств $N = \{3a, 1b, 2c\}$ и M (обозначается $N \Delta M$) равняется $N \Delta M = \{|3-2|a, |1-3|b, |2-1|c\} = \{1a, 2b, 1c\}$.

Понятие мультимножества поддерживается стандартом SQL 2003 [12], поэтому программная поддержка процесса консолидации информации на основе методов теории мультимножеств возможна на базе любой платформы, реализующей этот стандарт.

Нерешенные части общей проблемы. Обработка неполной, противоречивой, вербальной информации, которая поступает из разных источников, значительно усложняет процесс консолидации информации.

Постановка задачи. Адаптировать методы и средства теории мультимножеств к решению указанной проблемы.

Результаты исследований

Сопоставление консолидации и кластеризации. В процессе консолидации информации о конкурентах можно использовать процедуру иерархической кластеризации. В этом случае процесс накопления данных состоит в сборе данных о конкурентах. Затем происходит формирование информационных ресурсов путем представления собранных данных в виде мультимножеств. Дальнейшая кластеризация мультимножеств будет отвечать процессу добывания знаний. Консолидированной информацией в этом случае является представление результатов в виде дендрограмм. После этого принимается решение относительно занятия наиболее свободного рынка и, в конце концов, реализация принятого решения путем организации соответствующего производства (табл. 1). В этом случае процесс накопления данных состоит в сборе данных о конкурентах. Затем происходит формирование информационных ресурсов путем представления собранных данных в виде мультимножеств. Дальнейшая кластеризация мультимножеств будет отвечать процессу добывания знаний.

Таблица 1

Сопоставление консолидации и кластеризации

| Процесс консолидации информации | Процесс иерархической кластеризации |
|---|--|
| Накопление данных | Сбор данных о конкурентах |
| Формирование информационных ресурсов | Представление данных в виде мультимножеств |
| Добывание знаний | Кластеризация мультимножеств |
| Предоставление консолидированной информации | Представление результата в виде дендрограммы |
| Принятие решения | Занять свободный рынок |
| Реализация решения | Организация производства |

Консолидированной информацией в этом случае является представление результатов в виде дендрограмм. После этого принимается решение относительно занятия наиболее свободного рынка и, в конце концов, реализация принятого решения путем организации соответствующего производства.

Пример. Пусть из 6 источников поступает неполная и противоречивая информация относительно 8 фирм A_1 - A_8 , производящих мебель. Предположим, что эти фирмы характеризуются 3-мя признаками, определяющими направления их деятельности: офисная мебель (K_1), шкафы-купе (K_2), спальные (K_3). При этом признак K_1 имеет 2 значения: высококачественная (k_{11}), низкого качества (k_{12}); признак K_2 имеет 3 значения: дорогая (k_{21}), доступная (k_{22}), дешевая (k_{23}); признак K_3 имеет 4 значения: очень большая (k_{31}), большая (k_{32}), средняя (k_{33}), маленькая (k_{34}).

Информацию, поступающую о фирмах-конкурентах, можно представить в виде мультимножеств так, как приведено в табл. 2, где числа указывают, сколько раз из 6 источников поступала информация о характеристиках фирмы с данным значением признака.

Таблица 2

Информация о фирмах-конкурентах

| Фирмы | K_1 | | K_2 | | | K_3 | | | |
|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | k_{11} | k_{12} | k_{21} | k_{22} | k_{23} | k_{31} | k_{32} | k_{33} | k_{34} |
| A_1 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 3 | 3 |
| A_2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 2 |
| A_3 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| A_4 | 3 | 0 | 3 | 3 | 0 | 3 | 3 | 0 | 0 |
| A_5 | 0 | 3 | 0 | 3 | 3 | 0 | 0 | 3 | 3 |
| A_6 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| A_7 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 |
| A_8 | 3 | 0 | 3 | 0 | 0 | 3 | 3 | 0 | 0 |

В один класс группируются мультимножества, расстояния между которыми наименьшие. Для этого необходимо определить симметричную разность между всеми парами мультимножеств и найти мощности получаемых в результате мультимножеств, которые являются расстояниями между соответствующими мультимножествами (табл. 3).

Таблиця 3

Расстояния между мультимножествами

| | A ₁ | A ₂ | A ₃ | A ₄ | A ₅ | A ₆ | A ₇ | A ₈ |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| A ₁ | 0 | 9 | 18 | 27 | 3 | 10 | 17 | 24 |
| A ₂ | 9 | 0 | 9 | 18 | 10 | 1 | 8 | 17 |
| A ₃ | 18 | 9 | 0 | 9 | 17 | 8 | 1 | 10 |
| A ₄ | 27 | 18 | 9 | 0 | 24 | 17 | 10 | 3 |
| A ₅ | 3 | 10 | 17 | 24 | 0 | 9 | 18 | 27 |
| A ₆ | 10 | 1 | 8 | 17 | 9 | 0 | 9 | 18 |
| A ₇ | 17 | 8 | 1 | 10 | 18 | 9 | 0 | 9 |
| A ₈ | 24 | 17 | 10 | 3 | 27 | 18 | 9 | 0 |

В данном случае существует 2 пары мультимножеств с одинаковыми минимальными расстояниями между ними: (A₂, A₆) и (A₃, A₇). Выбираем для группирования пару (A₂, A₆) – оно состоит в добавлении этих мультимножеств. Результат – новое мультимножество B₁ = A₂ + A₆ = (2, 4, 2, 3, 4, 2, 2, 4, 4).

Заменяем старые мультимножества A₂ и A₆ новым мультимножеством B₁. Затем аналогично вычисляем расстояния между мультимножествами, группируем ближайшие и заменяем сгруппированные мультимножества новыми:

$$B_2 = A_3 + A_7 = (4, 2, 4, 3, 2, 4, 4, 2, 2);$$

$$B_3 = A_1 + A_5 = (0, 6, 0, 3, 6, 0, 0, 6, 6);$$

$$B_4 = A_4 + A_8 = (6, 0, 6, 3, 0, 6, 6, 0, 0);$$

$$C_1 = B_1 + B_3 = (2, 10, 2, 6, 10, 2, 2, 10, 10);$$

$$C_2 = B_2 + B_4 = (10, 2, 10, 6, 2, 10, 10, 2, 2);$$

$$D = C_1 + C_2 = (12, 12, 12, 12, 12, 12, 12, 12, 12).$$

Результат кластеризации представлен на рис. 2.

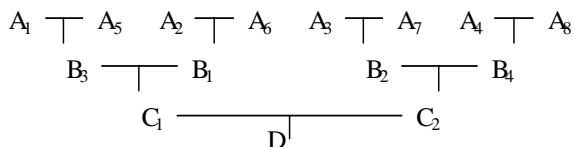


Рис. 2. Результат кластеризации

Таким образом, рынок мебели равномерно разделен между 8 фирмами. В нем нет свободного места, поэтому необходимо организовывать производство новых видов мебели, отличающихся от существующего ассортимента (офисная мебель, шкафы-купе и спальные).

ІЄРАРХІЧНА КЛАСТЕРІЗАЦІЯ МУЛЬТИМНОЖИН В ПРОЦЕСІ КОНСОЛІДАЦІЇ ІНФОРМАЦІЇ

Д.Б. Єльчанінов, М.С. Косіло

Засоби теорії мультимножин розглядаються як один із шляхів вирішення проблеми формалізації процесу консолідації інформації. Зіставлені процес консолідації інформації та процедура ієрархічної кластеризації. Наведено приклад використання процедури кластеризації у процесі консолідації інформації щодо фірм-конкурентів.

Ключові слова: консолідація інформації, мультимножина, кластеризація, дендрограма.

HIERARCHY CLUSTERING PROCEDURE MULTISSETS IN THE COMPETITIVE INTELLIGENCE PROCESS

D.B. Elchaninov, N.S. Kosilo

Means of the multisets theory are considered as a way of the problem formalization problem of competitive intelligence process. Competitive intelligence process and hierarchical clustering procedure are compared. The example of clustering procedure using in the competitive intelligence process relating to firms-competitors is given.

Keywords: competitive intelligence, multiset, clustering procedure, dendrogram.

Выводы

Процесс консолидации информации нуждается в формализации. Особую проблему представляет обработка неполной, противоречивой, вербальной информации, поступающей из разных источников. Для решения этой проблемы можно использовать процедуру иерархической кластеризации мультимножеств. **Перспективы дальнейших исследований.** Целесообразным представляется адаптация других понятий, операций, методов и средств теории мультимножеств к процессу консолидации информации с целью его формализации. Также необходимо исследовать особенности программной реализации указанного подхода.

Список литературы

1. Доронин А.И. Бизнес-разведка / А.И. Доронин. – М.: «Ось-89», 2006. – 496 с.
2. Ларичев О.И. Вербальный анализ решений / О.И. Ларичев. – М.: Наука, 2006. – 181 с.
3. Boyd, J.R. The Essence of Winning and Losing [Электронный ресурс] / Boyd's Last Briefing, 28 June 1995 – Режим доступа: <http://www.chetrichards.com/modern>.
4. Кунанець Н.Е. Консолідована інформація: сучасний фах освітньо-наукового напряму інформаційних наук / Н.Е. Кунанець // Вісник НУ «Львівська політехніка». – 2009. – № 653: Інформаційні системи та мережі. – С. 157-169.
5. Мельников Г.П. Системология и языковые аспекты кибернетики / Г.П. Мельников. – М.: Сов. радио, 1978. – 366 с.
6. Bossel H. Indicators for Sustainable Development: Theory, Method, Applications / H. Bossel // A Report to the Balaton Group. IISD: Winnipeg, Canada, 1999. – 124 p.
7. Каплан Р. Сбалансированная система показателей. От стратегии к действию / Р. Каплан, Д. Нортон. – М.: ЗАО «Олимп-Бизнес», 2008. – 320 с.
8. Маклаков С.В. Моделирование бизнес-процессов с AllFusion Process Modeler / С.В. Маклаков. – Диалог-МИФИ, 2003. – 240 с.
9. Деревянко А.С. Технологии и средства консолидации информации / А.С. Деревянко, М.Н. Солощук. – Х.: НТУ «ХПИ», 2008. – 280 с.
10. Петровский А.Б. Основные понятия теории мультимножеств / А.Б. Петровский. – М.: Ед.УРСС, 2002. – 80 с.
11. Петровский А.Б. Пространства множеств и мультимножеств / А.Б. Петровский. – М.: Ед.УРСС, 2003. – 248 с.
12. Клайн К. SQL. Справочник / К. Клайн. – М.: КУДИЦ-ОБРАЗ, 2006. – 832 с.

Поступила в редколлегию 20.04.2010

Рецензент: д-р техн. наук, проф. Г.В. Певцов, Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков.