УДК 65.011.56:519.23

КОГНИТИВНЫЙ ПОДХОД К АНАЛИЗУ ФОРМАЛЬНЫХ ПОНЯТИЙ И БИКЛАСТЕРИЗАЦИИ

Прокопчук Ю.А.,
Институт технической механики НАНУ и ГКАУ, г. Днепропетровск

В работе рассматривается специфика когнитивного подхода к анализу формальных понятий и бикластеризации. Когнитивный подход показывает, как понятия и категории связаны с доконцептуальной структурой опытного знания. Рассмотрены модели на основе замкнутых множеств разного уровня общности, на основе синдромного подхода и на основе орграфов набросков.

Ключевые слова: анализ формальных понятий, бикластреризация, когнитивный подход.

Введение и постановка задачи. Трудно найти что-то более важное для нормального функционирования мышления, восприятия, деятельности и речи, чем процессы категоризации, в результате которых формируются понятия, концепты и категории (по большей части, неосознаваемые). Изучение процессов категоризации особенно важно для разработки формальных моделей различных аспектов интеллектуальной деятельности [1-6]. В работе рассматривается возможность построения некоторых моделей, опираясь на методологию бикластеризации, анализа формальных понятий и принцип предельных обобщений [5, 6].

Анализ Формальных Понятий (АФП) предоставляет решеточные модели особого вида, позволяющие сохранять объектно-признаковое описание сходства группы объектов внутри кластера и, кроме того, строить иерархии таких кластеров по отношению «быть более общим, чем» [1-3]. В рамках этой области сформулировано математическое определение формального понятия (ФП) и описано построение иерархий ФП. Исходно ФП является парой вида (объем, содержание), где под объемом понимается некоторое множество объектов, а под содержанием – множество их общих признаков. Это определение соответствует описанию бикластера. Исходные данные в АФП представляются в виде объектно-признаковой матрицы, состоящей из нулей и единиц, а ФП является максимальный прямоугольник (с точностью до перестановок столбцов и строк) такой матрицы, заполненный единицами. Множество всех ФП упорядоченно и образует полную решетку, называемую «решеткой формальных понятий».

В ряде работ исследуются возможности «ослабления» требований к определению ФП, например, рассматриваются шумоустойчивые понятия и нечеткие решетки понятий [1]. Сообщество FIMI (Frequent Itemset Mining Implementation), изучает проблемы поиска частых (замкнутых) множеств признаков в больших базах данных. Задача поиска частых множеств признаков является одной из центральных тем в Data Mining. Замкнутые множества признаков являются в точности содержаниями ФП [2].

Несмотря на большое разнообразие подходов к бикластеризации, многие вопросы остаются открытыми, в частности, данные подходы не учитывают современных представлений о понятии (категории), как о сложно устроенном когнитивном феномене, не имеющем однозначной семантической репрезентации. Стало ясно, что процесс категоризации (выработки понятия) устроен куда сложнее, чем определение сущности, объединяющей элементы с общими признаками [4]. Когнитивный подход допускает ситуации, когда одни из объектов (прецедентов) в большей степени соответствуют представлению о понятии (категории), чем другие (понятиям, категориям свойственно иметь наилучших представителей). В отличие от обычной таксономии, категории, занимающие «серединное» положение в когнитивной иерархии, являются базовыми. Основная часть знания структурируется именно на этом уровне. Понятия, категории зависят от специфически человеческих способностей к восприятию, к созданию ментальных образов, к обучению и запоминанию, к организации усвоенных фактов и к эффективной коммуникации. Классические теории АФП и бикластеризации в этом отношении идеальны, поскольку в них понятия определяются исключительно в терминах общих характеристик их членов, а не в терминах специфических свойств человеческого понимания [4].

Авторский подход к когнитивным технологиям, включая Принцип предельных обобщений (ППО), изложен в работах [5, 6]. Ключевыми понятиями подхода являются Банк тестов $\{G(\tau(\}$, база прецедентов $\Omega = \{\alpha(\{\underline{\tau}/T\},\underline{z}/Z)\}$, синдромные $\{S\}$ и предельные синдромные модели знаний $\{S\ \}$, Z – задача, где Z – множество заключений.

Задачей настоящего исследования является применение когнитивного подхода к бикластеризации, который включает в себя три группы моделей:

- модели на основе замкнутых множеств;
- модели на основе синдромного подхода;
- модели на основе орграфов набросков.

Полученные результаты. Один из основных тезисов когнитивного подхода к АФП и бикластеризации заключается в том, что ключевое понятие «сходство» (задается операторами соответствия) имеет смысл только в рамках фиксированного уровня общности описания (совокупности доменов всех тестов). Различные уровни общности описания прецедентов формируются системой координат феноменологического пространства наблюдателя в виде Банка тестов. Другой важный тезис заключается в априорной «многоликости» понятия (категории, образа), в суперпозиции смыслов, вкладываемых в понятие, в существовании ядра и периферии системных признаков понятия (данный тезис реализуется в рамках синдромных моделей). Как следствие, все рассматриваемые модели обладают высокой устойчивостью к шумам за счет предельного обобщения и восстановления данных на основе моделей знаний.

Контекстом в АФП называют тройку $K = (G, M, I)$, где G – множество объектов, например, множество ситуаций действительности Ω (множество прецедентов); M – множество признаков (в нашем случае – Банк тестов), а бинарное отношение $I \subseteq G \times M$ говорит о том, какие объекты какими признаками обладают [2, 3].

При анализе прецедентов $\Omega = \{\alpha(\{\underline{\tau} / T\})\}$ под контекстом K в общем случае можно понимать тройку $<\Omega$, $\{G(\tau(\}$, $\{\tau / T\}>$, а в традиционной нотации $K_{\{\tau/T\}}$ =(Ω, $\{G(\tau(\}$, $I_{\{\tau/T\}}$), что означает фиксирование определенного уровня описания (доменов) по всем тестам. Без потери общности будем предполагать, что каждый тест входит в описание прецедента один раз, принимает одно значение и что для каждого прецедента известны значения всех тестов на всех уровнях (база прецедентов с полной информацией). Последнее требование фактически означает, что прецеденты должны быть описаны с помощью базовых доменов всех тестов: $\Omega = \{\alpha(\{\underline{\tau} / T_0\})\}$.

Главной особенностью контекста $K_{\{\tau/T\}}$ является зависимость результатов всех операций АФП от рассматриваемого уровня общности. Бинарное отношение $I_{\{\tau/T\}}$ в контексте $K_{\{\tau/T\}}$ можно представить следующим образом (все элементы всех доменов считаются различными):

$$I_{\{\tau/T\}} \subseteq \Omega \times M_{\{\tau/T\}}, \text{ где } M_{\{\tau/T\}} = T_{\tau I} \cup \dots \cup T_{\tau N}, \; N = |\{\tau\}|. \qquad (1)$$

Если уровень общности не фиксирован, то такой контекст будем называть свободным и обозначать $K_{\{\tau/T\}}$ = $<\Omega$, $\{G(\tau(\}>$.

Прецеденты могут быть описаны на уровне синдромов той или иной синдромной модели знаний $\{S\}$, т.е. $\Omega(Z) = \{\alpha(\{S\})\} \equiv \{\{S\}_\alpha\}$. Под контекстом $K_{\{S\}}$ в этом случае можно понимать пару $<\Omega(\{\{S\}_\alpha\})$, $\{S\}>$. В традиционной интерпретации можно записать: $K_{\{S\}}$ = ($\Omega(Z)$, $\{S\}$, $I_{\{S\}}$), где $I_{\{S\}} \subseteq \Omega \times \{S\}$. Ясно, что $K_{\{\tau/T\}}$, $Z \rightarrow K_{\{S\}} \rightarrow K_{\{\overset{\circ}{S}\}}$. В рамках данной цепочки происходит радикальное изменение объектно – признакового описания.

Контекст $K_{\{S\}}$ имеет значительно более высокий системный уровень, чем контексты $K_{\{\tau/T\}}$ и $K_{\{\tau/T\}}$ особенно, если речь идет о контекстах $K_{\{\overset{\circ}{S}\}}$. В рамках контекста $K_{\{S\}}$ категоризация осуществляется в два этапа: на первом этапе формируется синдромная модель знаний (формируются новые системные признаки), а на втором этапе формируются понятия уже на основе системных признаков. Контекст $K_{\{\overset{\circ}{S}\}}$ с полным основанием можно отнести к *базовому уровню* концептуальной организации. Следовательно, и понятия контекста $K_{\{\overset{\circ}{S}\}}$ также можно отнести к базовому уровню. К этому уровню относится и редуцированный Банк тестов $\{G\ (\tau\tau\}$, с помощью которого можно формировать контексты $K_{\{\tau/T\}}$ базового уровня («*пучок признаков*» является предшественником синдрома). Поскольку предельные синдромы являются гештальтами, то и идеализированные понятия базового уровня несут многие черты гештальтов.

База прецедентов может представлять собой множество образов $\Omega(\{w\})$. Для каждого образа w автоматизмами среды строятся орграфы набросков $Gs(w)$. Орграф набросков содержит слои набросков $\{p\}/Gs(w)$, где p – набросок, а также экстремальный пограничный слой набросков $\{p\ \}/Gs(w)$. Соответственно можно рассматривать множество разных контекстов, например: $K_{\{Gs\}} = <\Omega(\{w\}), \{Gs(w)\}, \{G(\tau(\}>$ – свободный контекст, $K_{\{p/Gs\}}$ – контекст фиксированного уровня общности, $K_{\{p/Gs\}}$ – контекст базового уровня и т.д. На основе орграфов набросков $\{Gs(w)\}$ могут быть построены (предельные) синдромные модели знаний $\{S_w\}$, следовательно, контекстом базового уровня является контекст $K_{\{S/Gs\}} = <\Omega(\{w\}), \{Gs(w)\}, \{S_w\}>$ [5].

Для произвольных $A \subseteq G$ и $B \subseteq M$ в АФП определены *операторы соответствия* Галуа [1-3]:

$$A' = \{m \in M \mid \forall g \in A\,(g\ I\ m)\};\ B' = \{g \in G \mid \forall m \in B\,(g\ I\ m)\}. \qquad (2)$$

Оператор '' (двукратное применение оператора ') является *оператором замыкания*: он идемпотентен ($A'''' = A''$), монотонен ($A \subseteq B$ влечет $A'' \subseteq B''$) и 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 $A \subseteq G$, $B \subseteq M$, $A' = B$ и $B' = A$, называется *формальным понятием контекста* K [1-3]. Множества A и B замкнуты и называются *объемом* и *содержанием* формального понятия (A, B) соответственно. Понятия, упорядоченные отношением

$$(A1, B1) \geq (A2, B2) \Leftrightarrow A1 \supseteq A2,$$

образуют полную решетку, называемую *решеткой понятий*.

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').

Для произвольного $B \subseteq M$ величина $|B'|$ называется *поддержкой* (support) B и обозначается $\sup(B)$. Нетрудно видеть, что множество B замкнуто тогда и только тогда когда для любого $D \supset B$ имеет место $\sup(D) < \sup(B)$. Множество $B \subseteq M$ называется k-частым если $|B'| > k$ (то есть множество признаков B встречается в более чем k объектах), где k – параметр.

Для произвольных $\{\alpha\} \subseteq \Omega$ и $\{\underline{b}/B\} \subseteq M_{\{\tau/T\}}$ в рамках контекста $K_{\{\tau/T\}}$ запишем *операторы соответствия* следующим образом:

$$\{\alpha\}'|_{\{\tau/T\}} = \{\underline{b}/B \in M_{\{\tau/T\}} \mid \forall \alpha \in \{\alpha\}\ (\alpha\ I_{\{\tau/T\}}\ \underline{b}/B)\}; \qquad (3)$$

$$\{\underline{b}/B\}'|_{\{\tau/T\}} = \{\alpha \in \Omega \mid \forall \underline{b}/B \in \{\underline{b}/B\}\ (\alpha\ I_{\{\tau/T\}}\ \underline{b}/B)\}. \qquad (4)$$

Каждый тест во множество $\{\underline{b}/B\}$ входит только один раз. Справедливо предложение.

***Предложение* 1**. Результат операции $\{\underline{b}/B\}'|_{\{\tau/T\}}$ будет одинаковым во всех контекстах $K_{\{b/B\}\cup\{a/A\}}$, где $\{a/A\}$ – произвольное дополнение $\{b/B\}$ до полного описания $\{\tau/T\}$.

Можно говорить *о кластере контекстов* с общим ядром $\{b/B\}$: $\{K\}_{\{b/B\}} = \cup_{\{a/A\}}\{K_{\{b/B\}\cup\{a/A\}}\}$. Другими словами, результат операции $\{\underline{b}/B\}'|_{\{\tau/T\}}$ будет одинаковым в рамках кластера контекстов $\{K\}_{\{b/B\}}$.

Операторы соответствия контекста $K_{\{S\}}$ имеют стандартный вид:

$$\{\alpha\}'|_{\{S\}} = \{S \in \{S\} \mid \forall \alpha \in \{\alpha\}\ (\alpha\ I_{\{S\}}\ S)\}; \qquad (5)$$

$$\{Si\}'|_{\{S\}} = \{\alpha \in \Omega \mid \forall S \in \{Si\}\ (\alpha\ I_{\{S\}}\ S)\}. \qquad (6)$$

Как известно любую синдромную модель знаний $\{S\}$ в рамках Z – задачи можно представить в виде $\{S\} = \cup_j \{S\}_j$, где $\{S\}_j$ отвечает $z_j \in Z$. Соответственно, $\Omega(Z) = \cup_j \Omega(z_j)$.

***Предложение* 2**. Пусть дано множество синдромов $\{Si\} \subseteq \{S\}$. Если существуют i, j такие, что $i \neq j$, $\{Si\} \cap \{S\}_j \neq \varnothing$, $\{Si\} \cap \{S\}_i \neq \varnothing$, то $\{Si\}'|_{\{S\}} = \varnothing$.

***Следствие* 1**. Если $\{Si\}'|_{\{S\}} \neq \varnothing$, то $\exists_j$: $\{Si\} \subseteq \{S\}_j$, а $\{\alpha\} \subseteq \Omega(z_j)$.

Дадим определения *обобщенных операторов соответствия* в рамках свободного контекста $K_{\cup\{\tau/T\}}$:

$$\{\alpha\}'|_{\cup\{\tau/T\}} = \{\underline{b}/B\}_+ = \cup_{b\in\{\tau\}}\,(\cap_{\alpha\in\{\alpha\}}(\underline{b}/B)^+_\alpha); \qquad (7)$$

$$\{\underline{b}/B\}'|_{\cup\{\tau/T\}} = \{\beta\},\ \text{таких что}\ \{\underline{b}/B\} \subseteq \cup_{c\in\{\tau\}}(\cap_{\beta\in\{\beta\}}(\underline{c}/C)^+_\beta), \qquad (8)$$

где $(\ )^+$ – замыкание множества значений тестов; нижний индекс «+» является маркером операции. Каждый тест во множество $\{\underline{b}/B\}$ может входить несколько раз, но с разными доменами (при этом не должно быть противоречивых данных). Ясно, что операторы (7) - (8) можно представить в традиционном виде, например:

$$\{\underline{b}/B\}'|_{\cup\{\tau/T\}} = \{\alpha \in \Omega \mid \forall \underline{b}/B \in \{\underline{b}/B\}\ (\alpha\ I_{\cup\{\tau/T\}}\ \underline{b}/B)\}, \qquad (9)$$

где $I_{\cup\{\tau/T\}} \subseteq \Omega \times M_{\cup\{\tau/T\}}$, а $M_{\cup\{\tau/T\}}$ объединяет все домены всех орграфов Банка тестов. Однако подобное представление нивелирует эффект обобщения, связанный с замыканием (эффект движения информации).

Приведем пример применения операторов соответствия (7) - (8). Пусть $\{G(\tau)\} = \{G^+(a),\ G(b)\}$, где $G^+(a)$ – структурно-завершенный орграф доменов для $G(a)$ [5]. Конфигураторы $G(a)$ и $G(b)$ имеют вид:

$$a\ \{A\ \{1;\ 2;\ 3;\ 4\}\};\quad b\ \{B\ \{1;\ 2;\ 3\}\};\quad G(b)=B;\quad G(a)=A.$$

Соответственно структурно-завершенный орграф $G^{+}(a)$ будет иметь вид:

a {A4 #A {4 ^4; ¬4 ^1 2 3} A3 #A {3 ^3; ¬3 ^1 2 4} A2 #A {2 ^2; ¬2 ^1 3 4}

A1 {1 ^1; ¬1 ^2 3 4} A {1; 2; 3; 4}};

$G^{+}(a)=\{A\to A1;\ A\to A2;\ A\to A3;\ A\to A4\}$.

Домены A и B являются базовыми. С их помощью описываются прецеденты.

Пусть $\Omega=\{\alpha,\beta,\gamma\}$, где

$\alpha=(a/A?1;b/B?1)$, $\beta=(a/A?2;b/B?2)$, $\gamma=(a/A?1;b/B?3)$.

Построим замыкания по всем тестам каждого прецедента:

$(\underline{a}/A)^{+}_{\alpha}=(\underline{a}/A)^{+}_{\gamma}=\{a/A?1;\ a/A1?1;\ a/A2?\neg 2;\ a/A3?\neg 3;\ a/A4?\neg 4\}$;

$(\underline{a}/A)^{+}_{\beta}=\{a/A?2;\ a/A1?\neg 1;\ a/A2?2;\ a/A3?\neg 3;\ a/A4?\neg 4\}$;

$(\underline{b}/B)^{+}_{\alpha}=\{b/B?1\}$; $(\underline{b}/B)^{+}_{\beta}=\{b/B?2\}$; $(\underline{b}/B)^{+}_{\gamma}=\{b/B?3\}$.

Применим обобщенные операторы соответствия контекста <Ω, {$G^{+}(a)$, $G(b)$}>:

$$\{\alpha,\beta\}'|_{\cup\{\tau/T\}}=\{a/A3?\neg 3;\ a/A4?\neg 4\};$$

$$\{a/A3?\neg 3;\ a/A4?\neg 4\}'|_{\cup\{\tau/T\}}=\{\alpha,\beta,\gamma\};$$

$$\{\alpha,\beta,\gamma\}'|_{\cup\{\tau/T\}}=\{a/A3?\neg 3;\ a/A4?\neg 4\}.$$

Как видим, пара множеств ($\{\alpha,\beta,\gamma\}$, $\{a/A3?\neg 3;\ a/A4?\neg 4\}$) образует формальное понятие контекста $K_{\cup\{\tau/T\}}$ = <Ω, {$G^{+}(a)$, $G(b)$}>.

Изменим контекст. Рассмотрим модернизированный Банк тестов $\{G(\tau)\}=\{G^{+}(a),\ G^{+}(b)\}$, где $G^{+}(b)$ – структурно-завершенный орграф для орграфа $G(b)$. По умолчанию он имеет такую же структуру, как и орграф $G^{+}(a)$, а именно:

b {B3 #B {3 ^3; ¬3 ^1 2} B2 #B {2 ^2; ¬2 ^1 3} B1 {1 ^1; ¬1 ^2 3} B {1; 2; 3}}; $G^{+}(b)=\{B\to B1;\ B\to B2;\ B\to B3\}$.

База прецедентов остается прежней. В рамках контекста $K_{\cup\{\tau/T\}}$ = <Ω, {$G^{+}(a)$, $G^{+}(b)$}> легко получить следующий результат

$$\{\alpha,\beta\}'|_{\cup\{\tau/T\}} = \{a/A3?\neg 3;\ a/A4?\neg 4;\ b/B3?\neg 3\};$$
$$\{a/A3?\neg 3;\ a/A4?\neg 4;\ b/B3?\neg 3\}'|_{\cup\{\tau/T\}} = \{\alpha,\beta\}.$$

Как видим, изменив в контексте лишь один орграф Банка тестов и не меняя базу прецедентов, мы получили совсем другой результат, в частности, формальное понятие образует иная пара множеств ($\{\alpha,\beta\}$, $\{a/A3?\neg 3;\ a/A4?\neg 4;\ b/B3?\neg 3\}$).

Пусть F – некоторый оператор вида

$$F:\ \{\alpha\},\ \{G(\tau)\} \to \{\underline{b}/B\}_F. \qquad (10)$$

С помощью оператора F определим *обобщенные операторы соответствия контекста* K_F:

$$\{\alpha\}'|_F = \{\underline{b}/B\}_F = F(\{\alpha\},\{G(\tau)\}), \qquad (11)$$

$$\{\underline{b}/B\}'|_F = \{\beta\},\ \text{ таких что } \{\underline{b}/B\} \subseteq F^+(\{\beta\},\ \{G(\tau)\}), \qquad (12)$$

где F^+ – замыкание множества значений тестов. Видно, что обобщенные операторы соответствия контекста $K_{\cup\{\tau/T\}}$ являются частным случаем обобщенных операторов соответствия контекста K_F. Приведем для примера еще одну интерпретацию оператора F:

$$\{\alpha\}'|_F = \{\underline{b}/B\}_F = \cup_{b\in\{\tau\}}(\cap_{\alpha\in\{\alpha\}}(\underline{b}/B)^+_\alpha)^-, \qquad (13)$$

где $\{\ \}^-$ – нижний предел множества значений тестов.

Приведем нотации операторов замыкания в разных контекстах:

$\{\alpha\}''|_{\{\tau/T\}} = \{\beta\}$ – замыкание по прецедентам, т.е. определение кластера сходных прецедентов по заданному исходному множеству прецедентов (ситуаций действительности). Уровень общности фиксирован;

$\{\alpha\}''|_{\cup\{\tau/T\}} = \{\beta\}$ – замыкание по прецедентам, т.е. определение кластера сходных прецедентов по заданному исходному множеству прецедентов в рамках свободного контекста $K_{\cup\{\tau/T\}}$ (уровень общности *не* фиксирован);

$\{\alpha\}''|_{\{S\}} = \{\beta\}$ – замыкание по прецедентам, т.е. определение кластера сходных прецедентов в рамках контекста $K_{\{S\}}$;

$\{\underline{b}/B\}''|_{\{\tau/T\}} = \{\underline{a}/A\}$ – замыкание по тестам, т.е. определение всех совпадающих результатов тестов у прецедентов, у которых совпадают некоторые априорно заданные результаты тестов. Уровень описания всех тестов фиксирован;

$\{\underline{b}/B\}''|_{\cup\{\tau/T\}} = \{\underline{a}/A\}$ – замыкание по тестам. Уровень описания не фиксирован;

$\{Si\}''|_{\{S\}} = \{S'\}$ – замыкание по синдромам в рамках контекста $K_{\{S\}}$.

Уточним формулировки идеализированных формальных понятий в рамках контекстов $K_{\{S\}}$, $K_{\{\tau/T\}}$ и $K_{\cup\{\tau/T\}}$.

Пара множеств ($\{\alpha\}$, $\{S\}_{\{\alpha\}}$), таких, что $\{\alpha\}\subseteq\Omega$, $\{S\}_{\{\alpha\}}\subseteq\{S\}$, $\{\alpha\}' = \{S\}_{\{\alpha\}}$ и ($\{S\}_{\{\alpha\}}$)' = $\{\alpha\}$, называется *формальным понятием контекста* $K_{\{S\}}$. В качестве модели знаний $\{S\}$ рекомендуется выбирать предельную синдромную модель $\{S^*\}$. Понятия контекста $K_{\{S^*\}}$ относятся к базовому уровню концептуализации. Справедливо утверждение.

***Предложение* 3**. Все формальные понятия контекста $K_{\{S\}}$ разбиваются на N непересекающихся классов, где $N = |Z|$. Каждый *j*–ый класс определяется подконтекстом $<\Omega(z_j)$, $\{S\}_j>$, где $\{S\}_j$ отвечает $z_j \in Z$.

Доказательство опирается на *предложение* 1.

Пару множеств ($\{\alpha\}$, $\{\underline{a}/A\}$), таких, что $\{\alpha\} \subseteq \Omega$, $\{\alpha\}'|_{\{\tau/T\}} = \{\underline{a}/A\}$ и $\{\underline{a}/A\}'|_{\{\tau/T\}} = \{\alpha\}$, назовем *формальным понятием контекста* $K_{\{\tau/T\}}$ (при условии, что каждый тест входит в описание один раз и база прецедентов с полной информацией).

Пару множеств ($\{\alpha\}$, $\{\underline{b}/B\}$), таких, что $\{\alpha\} \subseteq \Omega$, $\{\alpha\}'|_{\cup\{\tau/T\}} = \{\underline{b}/B\}$ и $\{\underline{b}/B\}'|_{\cup\{\tau/T\}} = \{\alpha\}$, назовем *формальным понятием контекста* $K_{\cup\{\tau/T\}}$. В данном варианте уровень общности заранее не фиксирован.

***Предложение* 4**. Если пара множеств ($\{\alpha\}$, $\{\underline{a}/A\}$) является формальным понятием контекста $K_{\{a/A\}\cup\{b/B\}}$, то она будет формальным понятием любого контекста $K_{\{a/A\}\cup\{b/B'\}}$, где $\{b/B\} \geq \{b/B'\}$ [6].

***Предложение* 5**. Если пара множеств ($\{\alpha\}$, $\{\underline{a}/A\}$) является формальным понятием свободного контекста $K_{\cup\{\tau/T\}}$ и каждый тест входит в $\{\underline{a}/A\}$ только один раз, то эта пара будет также формальным понятием фиксированных контекстов $K_{\{a/A\}\cup\{b/B\}}$, где $\{b/B\}$ – произвольное дополнение $\{a/A\}$ до полного описания $\{\tau/T\}$.

Действительно, если нашлось бы какое-то значение $\underline{b}/B$ принадлежащее всем $\{\alpha\}$, то оно обязательно вошло бы в замыкание $\{\underline{a}/A\}^+$, что противоречит условию. Следовательно, никакие значения $\underline{b}/B$ не могут быть добавлены к $\{\underline{a}/A\}$. При фиксированном контексте каждый тест входит в $\{\underline{a}/A\}$ только один раз (при свободном контексте это условие, как правило, не выполняется). Таким образом, выполнены все требования понятия фиксированного (по уровню общности) контекста.

Рассмотрим модели АФП и бикластеризации на основе синдромного подхода. Зададим контекст $K_Z = <\Omega(\{\alpha(z_\alpha)\})$, Z, $I_Z >$ следующим образом:

$$\alpha\ I_Z\ \underline{z}/Z \Leftrightarrow z_\alpha = \underline{z}/Z\,.$$

Справедливо предложение.

***Предложение* 6**. Полное множество формальных понятий контекста K_Z образуют пары $(\{\alpha\}_j,\ z_j)$, где $z_j \in Z$, а $\{\alpha\}_j = \Omega(z_j)$. Всего имеется N понятий, где $N = |Z|$.

Каждому понятию $(\{\alpha\}_j, z_j)$ контекста K_Z однозначно соответствует сопряженная пара множеств $(\{\alpha\}_j,\ \{S\}_j)$ контекста $K_{\{S\}}$, где $\{S\}_j$ отвечает $z_j \in Z$, следовательно, между объектами *j*-го понятия (бикластера) имеется лишь *семейное сходство*: они связаны общим заключением z_j, при этом у них может не быть общих свойств (синдромов). Действительно, в рамках любого $\{S\}_j$ может существовать *директивная зона* (общее ядро синдромов)

$$\{S\}_{\Omega j} = \ \cap_{\alpha \in \Omega j} \{S\}_\alpha, \quad j \in Z. \tag{14}$$

и *зона возможности*: $\{S\}_j^{\perp} = \{S\}_j \setminus \{S\}_{\Omega j}$. Если $\{S\}_{\Omega j} = \varnothing$, то это означает, что у объектов *j*-го бикластера нет общих свойств (кроме семейного сходства).

Директивная зона и зона возможности формируют структуру понятия «*центр - периферия*».

Благодаря наличию зоны возможности процедура распознавания семейного класса $\underline{z}/Z$ для нового или знакомого прецедента обладает максимальной *пластичностью* и *устойчивостью* к шумам. Устойчивость к пропускам данных повышает *процедура восстановления данных на основе моделей знаний*: для каждого дискретного домена теста, входящего в описание прецедентов, по требованию строится синдромная модель знаний (домен теста выступает в качестве Z); с использованием моделей знаний производится восстановление значения требуемого уровня общности.

Учитывая также предвестники $\{R\}_j$, сопряженность пар множеств $(\{\alpha\}_j, z_j)$ и $(\{\alpha\}_j,\ \{S\}_j,\ \{R\}_j)$ будем обозначать следующим образом ($j \in Z$):

$$(\{\alpha\}_j, z_j) \leftrightarrow (\{\alpha\}_j, \{S\}_j, \{R\}_j),\ \ \{\alpha\}_j = \Omega(z_j),\ \{S\}_j \subset \{S\},\ \{R\}_j \subset \{R\} \tag{15}$$

или в сокращенной нотации: $(\{\alpha\}_j, z_j,\ \{S\}_j,\ \{R\}_j)$, $j \in Z$. Структура (15) означает, что любое эмпирическое понятие должно быть понимаемо как *микро-теория* со своей внутренней логической структурой. Только обращение к внутренней структуре понятий или образов позволяет иерархически сопрягать их с мыслительными образами и теоретическими положениями более высоких структурных уровней (с суждениями, умозаключениями, предложениями).

Важно отметить, что пары $(\{\alpha\}_j, z_j)$ относятся к априорной информации, в то время как системные свойства $\{S\}_j$ и $\{R\}_j$ вычисляются, а точнее говоря, формируются в результате естественной самоорганизации. Системные свойства базового уровня (признаки) представляют собой отражение сущности (природных) способностей наблюдателя и опыта его функционирования в материальном и социальном окружении. Подобные понятия, категории, по-видимому, не осознаются, они используются автоматически, бессознательно и без заметных усилий, воспринимаясь просто как часть нормальной жизнедеятельности. Используемые таким образом понятия обладают иным, более важным, психологическим статусом по сравнению с теми понятиями, которые обязательно осознаются [4].

Пусть $Z = \{1, 2\} \equiv$ {фрукт+; фрукт-} $\equiv$ {образ+; образ-}. При такой оппозиции можно говорить о *понятии положительного контекста*. В общем случае речь будет идти о *понятии Z-контекста*. На основе контекста $<\Omega(Z), \{G(\tau)\}>$ формируется предельная синдромная модель знаний $\{S^*\}$ и бикластеры $\{\{\alpha\}_j, z_j, \{S^*\}_j\}$. Банк тестов $\{G(\tau)\}$ выполняет роль опытного знания существующего автономно от текущей Z-задачи. Он формируется в процессе решения большого числа разнообразных Z-задач и обучения. Данный пример показывает, что развиваемый подход кардинально отличается от других методов классификации, в частности, ДСМ-метода [2].

Выводы. Предлагаемый подход показывает, как понятия и категории связаны с доконцептуальной структурой опытного знания (орграфами доменов тестов, орграфами набросков, экстремальными слоями набросков, синдромами, ассоциативными связями и т.д.). Другими словами, когнитивные модели приобретают фундаментальную значимость благодаря своей способности органично вписываться в рамки доконцептуальной структуры опытного знания.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Игнатов Д. И. Бикластеризация объектно-признаковых данных на основе решеток замкнутых множеств / Д. И. Игнатов, С. О. Кузнецов // Труды 12-ой национальной конференции по искусственному интеллекту с международным участием. Т.1. – М. : Физматлит, 2008. – С. 175-182.
2. Кузнецов С. О. Методы теории решеток и анализа формальных понятий в машинном обучении / С. О. Кузнецов // Новости искусственного интеллекта. – 2004. – № 3. – С. 19-31.
3. B. Ganter Formal Concept Analysis / B. Ganter, R. Wille // Mathematical Foundations. – Springer, 1999.
4. Лакофф Джордж. Женщины, огонь и опасные вещи: Что категории языка говорят нам о мышлении. – М. : Языки славянской культуры, 2004. – 792 с.
5. Прокопчук Ю. А. Когнитивное моделирование на основе принципа предельных обобщений: методология, задачи, приложения / Ю. А. Прокопчук // Искусственный интеллект. – 2011. – №3. – С. 82-93.
6. Прокопчук Ю. А. Предельные синдромные и вероятностные модели знаний / Ю. А. Прокопчук // Науковий вісник ХДМІ : науковий журнал. – Херсон : Видавництво ХДМІ, 2011. – № 2 (5). – С. 322-333.

Прокопчук Ю.О. КОГНІТИВНИЙ ПІДХІД ДО АНАЛІЗУ ФОРМАЛЬНИХ ПОНЯТЬ І БІКЛАСТЕРІЗАЦІЇ

У роботі розглядається специфіка когнітивного підходу до аналізу формальних понять і бікластеризації. Когнітивний підхід показує, як поняття та категорії пов'язані з доконцептуальною структурою дослідного знання. Розглянуто моделі на основі замкнутих множин різного рівня спільності, на основі синдромного підходу і на основі орграфів начерків.

Ключові слова: аналіз формальних понять, бікластрерізація, когнітивний підхід.

Prokopchuk Iu.A. COGNITIVE APPROACH TO THE ANALYSIS OF FORMAL NOTIONS AND BICLUSTERIZATION

The article considers the specific cognitive approach to the analysis of formal concepts and biclusterization.The cognitive approach shows how notions and categories are related to the pre-conceptual structure of empirical knowledge. Models based on closed sets of different levels of generality, on the syndromic approach and sketch digraphs.

Keywords: formal concept analysis, biclusterization, cognitive approach.