

ПРИМЕНЕНИЕ СЕМИОТИЧЕСКИХ МОДЕЛЕЙ ПРЕДМЕТНОЙ ОБЛАСТИ В ЗАДАЧАХ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

А.А. Кулинич

Аннотация. Предложен подход к построению систем поддержки принятия решений, основанный на применении семиотической модели предметной области и методов обработки естественного языка. В качестве базы знаний семиотической модели используется корпус текста, релевантный предметной области, в которой построена субъективная семиотическая модель ситуации и полученный из сети Интернет. Предложен метод решения обратной задачи в семиотической системе. Для интерпретации получаемых решений в предметной области предложен семантический калькулятор, в котором применяются метод извлечения родовидовых отношений из корпуса текста на основе лексико-синтаксических шаблонов и метод определения частоты совместной встречаемости слов в решении на основе дистрибутивного анализа корпуса текста. Предложены обобщенные структуры подсистем мониторинга и принятия решений, основанные на семиотической модели ситуации и методах обработки естественного языка. Разработан программный макет подсистемы принятия решений. Эксперименты показали эффективность предложенного подхода.

Ключевые слова: принятие решений, семиотическая система, субъективная модель, обработка естественного языка, дистрибутивный анализ.

ВВЕДЕНИЕ

В настоящее время методы поддержки принятия решений в сложных социально-экономических и политических системах в условиях неопределенности можно разделить на ряд классов. Первый класс методов (Data Mining) основан на получении закономерностей предметной области и представляет собой извлечение знаний из данных, представленных в числовых шкалах.

Другой класс методов основан на непосредственном извлечении знаний экспертов – это их представление лучшего решения в моделях выбора или в моделях прогноза развития ситуаций. Эти методы используют субъективные предпочтения экспертов, их оценки, знания закономерностей предметной области и др. Однако при этом возникают сложности построения математической модели объекта и измерения его параметров. В условиях неопределенности такая модель – это концептуальная имитационная модель, построенная экспертом и качественно отражающая основные закономерности ситуации. В таких условиях модель ситуации трудно верифицировать и поэтому ре-

зультаты моделирования, трудно интерпретируемые в терминах предметной области, оказываются ненадежными.

Методы принятия решений с использованием лингвистической информации об объекте управления исследовались в рамках ситуационного управления [1]. Здесь описание объекта на естественном языке представляется на ограниченном естественном языке с помощью ядерных конструкций, включающих в себя элементы языка и разнообразные отношения между ними. Такое описание называется состоянием объекта и управление возможно, если для некоторого целевого состояния существует описание управляющего воздействия на естественном языке. В ситуационном управлении необходимо перечислить все возможные состояния объекта управления и привязать к каждому состоянию управляющее воздействие на естественном языке. Для сложных объектов такая задача оказывается труднорешаемой, требующей большой экспертной работы.

Идеи ситуационного управления получили развитие в рамках прикладной семиотики [2]. Здесь модель объекта строится с использованием знаков-



символов. Знак-символ определен немецким логиком Г. Фреге и представляется тройкой [3]: имя, смысл и значение знака. Знак-символ связывает знания эксперта (это имя и смысл) с объектом реального мира (это значение знака).

В работе [4] знак определяется как четверка: имя, образ (перцепт), значение и личностный смысл. Здесь определена математическая модель знака, операторы связывания всех его элементов, операции на множествах элементов знака. В этой работе акценты сделаны на распознавание перцептивных образов в виде связанной знаковой картины мира, определяющей поведение субъекта, основанное на опыте (личностном смысле).

В работе [2] определена модель семиотической системы, известная как квадрат Поспелова. Квадрат Поспелова включает в себя: метазнак, определяющий имя семиотической системы (множество знаков-символов); синтаксис, определяющий правила построения знаковой системы; семантику, определяющую смысл – основные свойства семиотической системы; прагматику, определяющую основные действия, которые возможно выполнить в рамках этой семиотической системы.

Основные аспекты семиотических систем – синтаксис, семантика и прагматика семиотической системы – были сформулированы в классических работах логиков Ч. Пирса [5] и Дж. Морриса [6].

Семиотический подход применяется: при проектировании информационных систем [7]; при проектировании компьютерных систем (*computer semiotics*) [8]; для представления интерфейсов систем в разных, но эквивалентных знаковых системах (*algebraic semiotic*) [9]; в концептуальном моделировании в базах данных [10], основанном на расширении модели «сущность – связь», с помощью алгебры фреймов и образов данных.

Применение семиотического подхода для решения сложных стратегических задач в энергетической сфере и других критических инфраструктурах рассматривалось в работах [11, 12].

В настоящей работе рассматриваются вопросы построения семиотических систем поддержки принятия решений в условиях неопределенности. Для этого совместно используются качественная семиотическая модель ситуации и технологии получения релевантной информации из сети Интернет с применением различных методов обработки естественного языка. Качественная субъективная семиотическая модель используется как шаблон для получения релевантной информации из сети

Интернет. Рассматриваются задачи: мониторинга состояния ситуации и получения прогнозов ее развития; поддержки принятия решений для управления ситуацией с интерпретацией решений и поиска их прецедентов в сети Интернет.

1. АРХИТЕКТУРА СЕМИОТИЧЕСКОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Архитектура систем управления, основанная на прикладной семиотике [1, 2, 13], формировалась в рамках исследований систем ситуационного управления сложными объектами.

Архитектура семиотической системы поддержки принятия решений ориентирована на работу с объектами (ситуациями), описанными на естественном языке, и включает в себя следующие подсистемы (рис. 1):

- интерпретатор входного языка, который переводит неструктурированную лингвистическую информацию об объекте управления на естественном языке на внутренний язык системы;
- анализатор, который осуществляет предварительную классификацию текущей ситуации на ситуации, требующие и не требующие управления;
- классификатор, осуществляющий обобщение и редукцию текущей ситуации к одному или нескольким классам типовых ситуаций, включенных в базу знаний и допускающих одношаговые управляющие воздействия;
- коррелятор, который организует процесс формирования управляющих воздействий на управляемый объект, адекватных текущим ситуациям и текущей модели мира в базе знаний системы;
- экстраполятор, который оценивает полученное решение на основе прогноза развития ситуации в текущей модели мира и, возможно, корректирует его, подбирая модель мира в цикле «анализатор – классификатор – коррелятор»;
- интерпретатор внутреннего языка, который представляет полученное решение на естественном языке.

Основным элементом этой архитектуры является база знаний, включающая в себя множество классов типовых ситуаций и моделей возможных миров. Последовательная передача результатов работы от интерпретатора анализатору, классификатору, а затем коррелятору и экстраполятору образует цикл работы семиотической архитектуры системы.

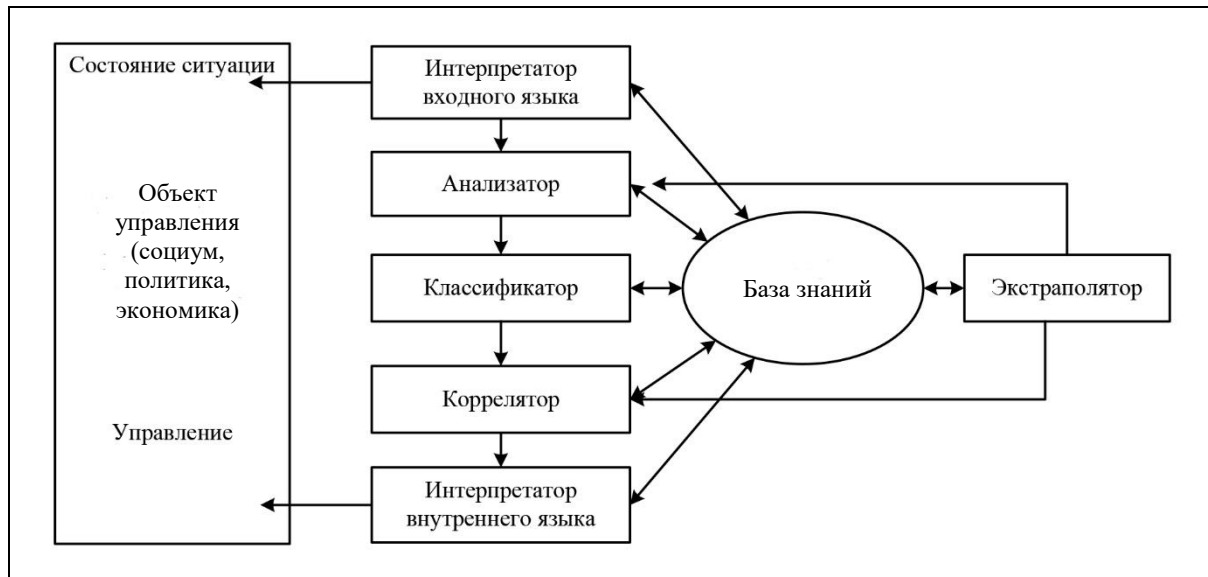


Рис. 1. Архитектура семиотической системы [13]

В системах управления, построенных на принципах прикладной семиотики, происходит чередование шагов математического моделирования ситуаций и синтеза управления в рамках фиксированных формальных моделей мира, представленных в базе знаний. Семиотическое моделирование приводит к перестройке формальных моделей мира на основе знаний о предметной области и реальных ситуациях, возникающих в процессе функционирования объекта и системы управления [13].

В этой семиотической архитектуре модели и возможные миры для принятия решений, представленные в базе знаний, замкнуты в отношении знаний о проблемной области.

В условиях неопределенности семиотическая система должна быть открыта, т. е. должна иметь возможность пополнения знаний. Поэтому основное отличие рассматриваемой здесь семиотической системы поддержки принятия решений заключается в том, что для ее работы в качестве базы знаний используется информация сети Интернет.

В международных и российских стандартах даются различные определения информации. Например, в области информационных технологий информация – это знания о фактах, событиях, вещах, идеях и понятиях, которые в определенном контексте имеют конкретный смысл (ISO/IEC 2382:2015), а в области обработки информации это любые данные, представленные в электронном или любом другом виде, предназначенные для обработки информационными системами и системами принятия решений. Федеральный закон «Об информации, информационных технологиях и о защите информации» № 149-ФЗ трактует информа-

цию как «сведения (сообщения, данные) независимо от формы их представления».

При этом данные определяются как представление информации формализованным способом, подходящим для коммуникации, интерпретации или коммуникации или обработки (ISO/IEC 2382-1:2015).

По форме представления выделяют структурированные и неструктурированные данные. Структурированные данные (*structured data*) – это данные, которые организованы на основе заранее определенного набора правил. Предопределенный набор правил, регулирующий основу, на которой структурированы данные, должен быть четко установлен и обнародован и может быть использован для управления структурированием данных (ИСО/МЭК 20546:2019(ru), 3.1.35). Примеры структурированных данных – это данные, представленные в таблицах баз данных, или же размеченный вручную или автоматически текст. Разметка предполагает привязку к словам текста определенных меток (тегов), что позволяет представить информацию в табличном виде и обработать.

Неструктурированные данные (*unstructured data*) – это данные, характеризующиеся отсутствием какой-либо структуры, кроме структуры на уровне записи или файла. Примером неструктурированных данных является свободный текст (ИСО/МЭК 20546:2019(ru), 3.1.37). Формально свободный текст имеет синтаксическую структуру и передает конкретный смысл. Однако для получения информации из текста необходима его предварительная обработка методами обработки естественного языка и интеллектуального анализа.



Отметим, что архитектура, предложенная в работе [13], – абстрактная. Здесь сформулированы основные идеи: как такая семиотическая система и ее подсистемы могут функционировать, но ее реализация невозможна. В настоящей работе предлагается возможная реализация семиотической системы поддержки принятия решений в условиях неопределенности. Предложенная архитектура включает в себя семиотическую модель, подсистему мониторинга ситуации, подсистему выработки альтернатив решений и их объяснение. Внутренний язык задается субъективной семиотической моделью, построенной экспертом. Интерпретатор входного языка – это подсистема мониторинга состояния. Эта подсистема позволяет представить свободный текст, полученный из сети Интернет, на внутреннем языке семиотической системы. Подсистема выработки альтернатив решений основана на решении обратной задачи в семиотической системе. Кроме того, в этой подсистеме реализуется интерпретатор внутреннего языка. Интерпретатор переводит решения, полученные в семиотической системе, на естественный язык, определяя имя класса решений и объясняя контекст, в котором имя класса используется (это предложение – свободный текст).

В условиях неопределенности в качестве базы знаний предложено использовать неструктурированные данные сети Интернет (свободный текст), который предварительно структурируется с помощью методов обработки естественного языка.

2. СЕМИОТИЧЕСКАЯ МОДЕЛЬ СИТУАЦИИ

В роли качественной субъективной модели ситуации выступает семиотическая модель предметной области, предложенная в работе [14]. Семиотическая модель ситуации [14] является субъективной качественной моделью представления знаний эксперта о динамических системах, элементами которой является модель знака Г. Фреге [3], связывающая реальный мир (денотат) с ментальными представлениями о мире (знаниями) в виде имени знака (символа) и смысла, определяющего его основные признаки (свойства). Семиотическая модель описывает ситуацию в трех аспектах: синтаксическом, семантическом, прагматическом. Синтаксис отвечает за представление отношений между знаками, описывающими реальность. Семантика изучает отношения между знаками и тем, что они обозначают в реальном мире, а прагматика отвечает за отношения между знаками и теми, кто ими пользуется. В нашем случае пользователь зна-

ковой системы – это лицо, принимающее решение (ЛПР).

Синтаксическая модель. При построении синтаксической модели будем основываться на логико-лингвистическом представлении основных элементов исследуемой системы [15]. В логико-лингвистических моделях основные элементы, отношения между ними, их возможные состояния (значения) представляются на естественном языке. В синтаксической модели представлены основные части анализируемой системы в виде множества их имен $D = \{d_i\}$, $i = 1, \dots, M$. На этом множестве определено отношение «часть – целое» – $\Theta \subseteq D \times D$. Для каждой составной части d_i моделируемой ситуации определено множество имен его параметров $F_i = \{f_{ij}\}$, где j – номер параметра i -й части. Для каждого параметра задано множество значений, $Z_i = \{Z_{ij}\}$ в виде упорядоченного множества лингвистических значений, т. е. $Z_{ij} = \{z_{ij1}, \dots, z_{ijq}\}$, $z_{ijq+1} \succ z_{ijq}$, $q = 0, \dots, n - 1$. Вектор значений всех параметров ситуации в момент времени t , $Z(t) = (z_{1je}, \dots, z_{njqs})$, называется ее состоянием.

Логико-лингвистическая модель строится для представления динамики изменения параметров ситуации и решения обратной задачи для нахождения значений параметров, способных изменить текущее состояние ситуации на целевое. Для моделирования динамики необходимо определить причинно-следственные связи между параметрами и их силу.

Сила причинно-следственной связи определяется на естественном языке из множества возможных значений: например, $RF = \{\text{«Сильно усиливает»}, \text{«Усиливает»}, \text{«Слабо усиливает»}\}$. Сила причинной связи определяет бинарное отношение между множествами возможных значений параметров. Например, $RF(\text{«Сильно усиливает»}) \subseteq Z_{1j} \times Z_{2u}$ означает, что параметр j первой подсистемы (причина) связан с параметром u второй подсистемы (следствие). Силу влияния определяют пары значений из множества $Z_{1j} \times Z_{2u}$: например, для отношения «Сильно усиливает» пары значений этого отношения $(z_{1j2}, z_{2u3}; \dots, z_{1jn}, z_{2um})$ означают, что изменение значения параметра j первой подсистемы до второго элемента (z_{1j2}) множества значений Z_{1j} (причины) увеличит значение параметра u второй подсистемы до третьего элемента (z_{2u3}) множества Z_{2u} и т. д.

В синтаксической модели экспертным способом определяются причинно-следственные отношения между параметрами и представляются в виде отношения W , заданного на множествах значений всех параметров. Для определения прогноза разви-

тия ситуации это отношение в виде логико-лингвистических уравнений для системы, имеющей f_i параметров ($i = 1, \dots, n$), каждый из которых имеет множество значений Z_i , запишем в виде отображения [15]

$$W: Z(t) \rightarrow Z(t+1), \quad (1)$$

где $Z(t) \in \times_i Z_i$, $Z(t) = (z_{1t}, \dots, z_{nt})$ – вектор состояния ситуации; $\times_i Z_i$ – множество векторов возможных состояний; W – причинно-следственные отношения; $i = 1, \dots, n$.

Для вычисления прогноза развития ситуации в модели, заданной логико-лингвистическими уравнениями, обычно применяют теорию нечетких множеств и систем [16]. В этом случае для всех параметров должны быть определены функции принадлежности и нечеткое причинно-следственное отношение. Несмотря на то, что все теоретические вопросы для вычисления прогноза в нечетком виде решены, процесс построения функций принадлежности и определения нечетких отношений требует большой экспертной работы.

В работе [17] Б. Коско предложил нечеткую каузальную алгебру, которая не требует построения функций принадлежности, достаточно упорядоченных лингвистических значений силы причинных связей. Он предлагает способ вычисления влияния одного параметра на другой по цепочкам причинно-следственных связей. Сила влияния цепочки определяется минимумом силы, включенной в цепочку, а агрегированное значение влияния всех цепочек определяется как максимум из всех сил, влияющих на параметр цепочек влияния.

В настоящей работе применяется метод получения прогноза развития ситуации, предложенный в статье [18].

Здесь строится шкала лингвистических значений параметра $Z_{ij} = \{z_{ij1}, \dots, z_{ijq}\}$ как отображение в числовое множество $X_{ij} = \{x_{ij1}, \dots, x_{ijq}\}$, элементы которого определены на отрезке числовой оси $[0, 1]$, т. е. $x_{ij1}, \dots, x_{ijq} \in [0, 1]$. Точки числовой оси образуют упорядоченное множество числовых точек X_{ij} , элементы которого упорядочены так же, как упорядочены лингвистические значения, т. е. если $z_{ijq+1} \succ z_{ijq}$, то $x_{ijq+1} \succ x_{ijq}$. Таким образом, определено отображение $\varphi: Z_{ij} \rightarrow X_{ij}$. Определено также и обратное отображение $\varphi^{-1}: X_{ij} \rightarrow Z_{ij}$, позволяющее интерпретировать любое значение $x_{ijq} \in [0, 1]$ в лингвистическое значение $z_{ijq} \in Z_{ij}$, $\forall q$.

В этом случае сила причинно-следственной связи может быть интерпретирована как коэффициент передачи, определенный как действительное

число. Значения параметра причины и параметра следствия связаны соотношением $x_{ijq} = w_{ijpt}^* x_{pto}$, где $x_{ijq} \in X_{ij}$ – значение параметра следствия, $x_{pto} \in X_{pt}$ – значение параметра причины. Коэффициент передачи $w_{ijpt}^* = 1$, если моделируется отношение «Усиливает», $w_{ijpt}^* > 1$, если моделируется отношение «Сильно усиливает», и $w_{ijpt}^* < 1$ при отношении «Слабо усиливает». Подробно вопросы определения силы причинной связи (коэффициентов передачи) описаны в работе [18].

Линейная связь значений параметров причины и следствия оправдана в условиях неопределенности, при субъективных оценках силы причинно-следственного влияния.

Вычисление прогноза развития ситуации в числовой системе определится с помощью конечно-разностного уравнения

$$X(t+1) = W^{\circ} X(t),$$

где $X(t)$ и $X(t+1)$ – векторы параметров ситуации в последовательные моменты времени; $W^* = |w_{ij}^*|_{n \times n}$ – матрица коэффициентов передачи; \circ – правило вычисления прогнозных значений.

В качестве правила вычисления прогнозных значений вектора $X(t+1)$ используется правило агрегации значений max-product (умножение и взятия максимума). Таким образом, элемент прогнозного вектора определится из соотношения: $x_i(t+1) = \max_{r=1, \dots, n} x_r(t) w_{ir}^*$, $\forall i$.

Прогнозный вектор может быть представлен в лингвистическом виде: $Z(t+1) = \varphi^{-1}(X(t+1))$.

Таким образом, синтаксическая модель определена кортежем

$$\langle F, Z, W, Z(t) \rangle, \quad (2)$$

где F – множество параметров; Z – множество множеств значений параметров; W – причинно-следственное отношение на множестве значений параметров; $Z(t)$ – состояние (вектор значений всех параметров).

Семантическая модель. В семантической модели предметной области представлены возможные состояния синтаксической модели (2) в виде частично упорядоченного множества именованных классов состояний.

Такое представление основано на интерпретации пространства возможных состояний динамической системы (1) $SS = \times_i Z_i$ как признакового семантического пространства (*Semantic Space*).

В признаковом семантическом пространстве состояния ситуации представляются как понятия. Реальные ситуации (состояния-денотаты), опреде-

ляются именами и векторами значений признаков, характеризующих их содержание (смысл). В семантических пространствах ситуации с близкими значениями параметров образуют классы состояний, а между классами определяются отношения («класс – подкласс» или «род – вид»), т. е. определяется понятийная структура.

В работе [19] был предложен метод структуризации пространства состояний SS динамической системы (1) на вложенные области возможных состояний $SS(d^H) \subset SS$, имеющие искусственные имена d^H , определяющие класс состояний системы (1): $SS(d^H) \Leftrightarrow d^H, H = 0, \dots, 3^N$, где N – число параметров.

На рис. 2–5 дана иллюстрация предложенного метода структуризации пространства состояний на примере системы с двумя параметрами (признаками).

На рис. 2 показан пример семантического пространства для ситуации с двумя признаками $F = \{f_1, f_2\}$, значения которых определяются из множеств значений соответственно $Z_1 = \{Z_{11}, \dots, Z_{1n}\}$ и $Z_2 = \{Z_{21}, \dots, Z_{2m}\}$. Начальное состояние ситуации $Z(0)$ показано как точка с координатами $Z(0) = (Z_{1q}, Z_{2s})$, в пространстве $SS = Z_1 \times Z_2$.

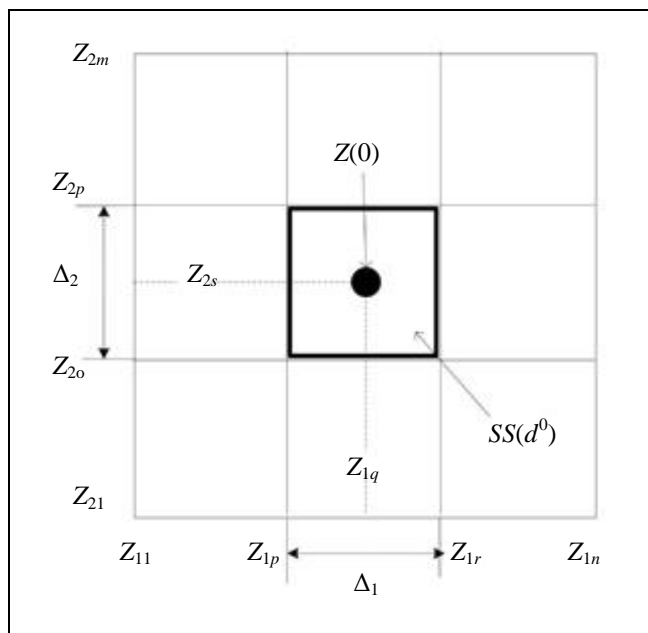


Рис. 2. Область базового класса состояний

Экспертным путем определяются окрестности точки начального состояния по признаку 1 (это $\Delta_1 = \{Z_{1p}, \dots, Z_{1r}\}$) и по признаку 2 ($\Delta_2 = \{Z_{2o}, \dots, Z_{2p}\}$), которые называются интервалами толерантности начального состояния по признакам 1 и 2. Область семантического пространства $SS(d^0) \subseteq SS$, полу-

ченная прямым произведением интервалов толерантности всех признаков состояния $Z(0)$, определяет базовый класс состояний $SS(d^0) = \Delta_1 \times \Delta_2$. Любое состояние $Z(t)$ системы (1), которое принадлежит базовому классу $SS(d^0)$, $Z(t) \in SS(d^0)$, будет иметь имя d^0 , т. е. базовый класс определяет класс малоразличимых, эквивалентных состояний.

На рис. 3 показаны области семантического пространства $SS(d^1), SS(d^2), SS(d^3), SS(d^4)$. Например, область $SS(d^3)$ определяется так: $SS(d^3) = \{Z_{2o}, \dots, Z_{2p}\} \times \{Z_{11}, \dots, Z_{1r}\} = \Delta_2 \times \Delta_3$. Поскольку области $SS(d^1), SS(d^2), SS(d^3), SS(d^4)$ включают в себя область базового класса состояний $SS(d^0)$, то считается, что они обобщают базовый класс состояний. Причем области $SS(d^1), SS(d^3)$ обобщают область $SS(d^0)$ по признаку 2, а $SS(d^2), SS(d^4)$ – по признаку 1. Областям $SS(d^1), SS(d^2), SS(d^3), SS(d^4)$ соответствуют имена этих областей, соответственно имена классов состояний d^1, d^2, d^3, d^4 . Далее мы будем оперировать именами классов состояний.

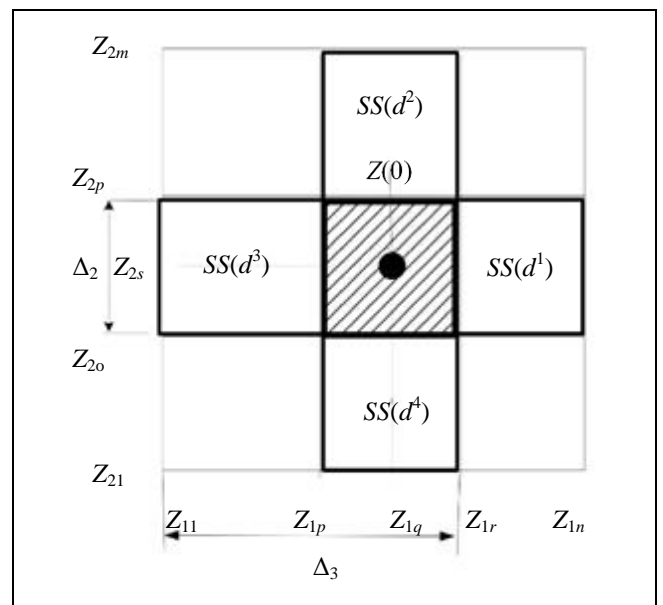


Рис. 3. Области обобщенных классов состояний по признакам 1, 2

На рис. 4 показаны области семантического пространства $SS(d^5), SS(d^6), SS(d^7), SS(d^8)$. Например, область $SS(d^5)$ определяется так: $SS(d^5) = \{Z_{2o}, \dots, Z_{2m}\} \times \{Z_{1r}, \dots, Z_{1n}\} = \Delta_4 \times \Delta_3$. Эти области включают в себя область базового класса и обобщенные области по одному из признаков. Например, $SS(d^5)$ включает в себя области $SS(d^1), SS(d^2)$ и область базового класса $SS(d^0)$. Такая вложенность позволяет говорить, что эти области обобщают обобщенные области по одному признаку, а базовый класс состояний – по двум признакам. Эти области обозначаются именами d^5, d^6, d^7, d^8 .

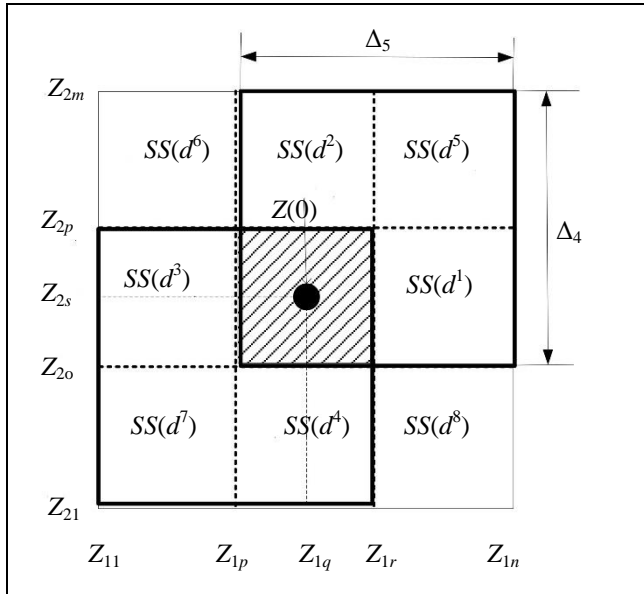


Рис. 4. Области обобщенных классов состояний по двум признакам

В работе [16] было показано, что имена d^H по вложению областей состояний $SS(d^H)$ образуют частично упорядоченное множество имен $\{d^H\}$ классов состояний $CF = (\{d^H\}, \leq)$, называемое качественным концептуальным каркасом (*Conceptual Framework*), который определяет качественную онтологию предметной области, для которой, в свою очередь, была построена синтаксическая модель (2).

На рис. 5 показана диаграмма Хассе, определяющая частичный порядок имен классов состояний. Первый уровень имен классов обобщает базовый класс, а второй обобщает имена первого уровня. Такая качественная онтология определяет понятийную систему плохо определенной предметной области.

Таким образом, семантика синтаксической модели (2) определена качественным концептуальным каркасом

$$CF = (\{d^H\}, \leq), \quad (3)$$

где $d^H \Leftrightarrow SS(d^H)$ – имена классов состояний, которые однозначно определяют области семантического пространства.

Отметим, что состояния синтаксической модели $Z(t)$ в разные моменты времени могут принадлежать разным областям $SS(d^H)$ и, следовательно, иметь разные имена d^H , $Z(t) \in SS(d^H)$.

Отметим также, что концептуальный каркас (3) является идеализированной семантической моделью предметной области. Здесь явно определено только имя базового понятия (класса) d^0 , а для всех

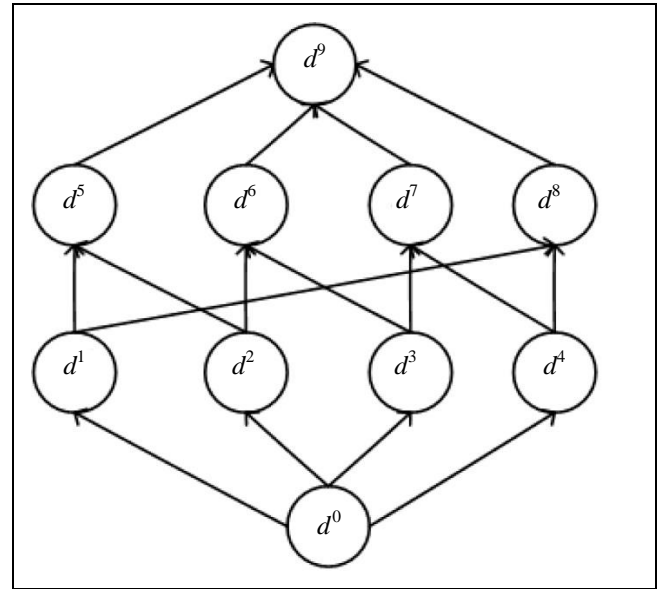


Рис. 5. Концептуальный каркас предметной области

остальных классов состояний определены искусственные имена d^H и области семантического пространства $SS(d^H)$, которое они определяют. В условиях неопределенности, когда онтологию предметной области построить невозможно, концептуальный каркас задает «опорные точки» возможных классов состояний. Эти «опорные точки» связаны с базовым классом отношением «класс – подкласс». Реальные (не искусственные) имена классов состояний, которые будут соответствовать решениям обратной задачи, мы будем искать с помощью методов обработки неструктурированных данных (обработки свободного текста) в сети Интернет.

Прагматическая модель. В семиотике прагматика отвечает за отношения между знаками и теми, кто ими пользуется. В прагматической модели рассматривается оценка полезности состояния системы для ЛПР. Оценка основана на определении значений коэффициентов предпочтений эксперта α_j относительно значений параметров в векторе состояний $(Z(t))$.

Оценка состояния $O(Z(t))$ представляется линейной сверткой

$$O(Z(t)) = \sum_j \alpha_j x_j(t), \quad j = 1, \dots, n, \quad (4)$$

$$\sum_j \alpha_j = 1.$$

где $x_j(t) \in [0, 1]$ – отображение φ лингвистических значений параметров $z_j(t)$ на отрезок числовой оси $[0, 1]$, т. е. $\varphi: z_j(t) \rightarrow x_j(t) \in [0, 1]$, $z_j(t) \in Z(t)$, $O(Z(t)) \in [0, 1]$.



Все три модели связаны между собой общими параметрами, и поэтому изменение состояния любой из них приведет к изменению состояний других моделей.

Общая постановка задачи принятия решения в семиотической системе следующая. Для семиотического описания сложной системы (2)–(4) необходимо найти новое описание системы в синтаксической и семантической модели, такое чтобы их прагматическая оценка O^* была бы лучше существующей O . В работе [20] был предложен метод решения этой задачи. Он основан на решении обратной задачи в семиотической модели путем определения цели в прагматической модели, решения обратной задачи в этой модели и последовательной передачи результатов решения из прагматической модели в синтаксическую, а затем в семантическую модель.

Для решения обратной задачи в синтаксической модели применялись методы решения обратных задач для систем, параметры которых определены качественно или нечетко. Эти методы рассматривались в работах [21, 22].

3. МЕТОД РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ

Рассмотрим алгоритм решения обратной задачи (*inverse problem*) в общем виде. Считается заданным отношение $W = |w_{ij}|_{n \times n}$ между признаками ситуации и целевой вектор значений признаков ситуации $G = (g_1, g_2, \dots, g_n)$. Задача заключается в нахождении множества векторов входных воздействий $\Omega = \{U_k\}$ таких, что для всех $U_k = (u_{k1}, u_{k2}, \dots, u_{kn})$ выполняется равенство $U_k \circ W = G, \forall k$. Задача решается в числовой системе, т. е. элементы отношения W , вектора цели $g_i \in G$ и вектора управляющих воздействий $u_{ij} \in U$ определяются как действительные числа: $w_{ij} \in \mathbb{R}, g_i \in [0, 1], u_{ki} \in [0, 1]$.

Поскольку для получения прогнозов развития ситуации использовалась композиция *max-product*, то в дальнейшем будем рассматривать алгоритмы решения обратных задач для композиционного правила *max-product*. Задача здесь заключается в нахождении обратного отображения к композиции *max-product*.

В работе [22] разработан итерационный алгоритм, позволяющий получить множество решений обратной задачи $\Omega = \{U_{\max}, U_{\min}\}$ – одно максимальное U_{\max} и множество минимальных решений $U_{\min} = \{U_1, U_2, \dots, U_q\}$, где $U_{\max}, U_1, \dots, U_q$ – векторы значений параметров ситуации. Итерационный алгоритм такой.

1. Определяется вектор $U_{\max} = (u_{1\max}, \dots, u_{n\max})$ максимального решения $U_{\max} = \min(W \alpha G^T)$, где

$$w_{ij} \alpha g_j = \begin{cases} 1, & \text{если } g_j \geq w_{ij}; \\ \frac{g_j}{w_{ij}} & \text{в противном случае.} \end{cases}$$

Элемент вектора U_{\max} определяется из соотношения $u_{i \max} = \min_{r=1}^n w_{ir} \alpha g_r$. При определении вектора U_{\max} , в отличие от обычного умножения матриц, для композиции *max-product* вместо обычного умножения используется операция α , а вместо суммы – взятие минимума.

2. Множество минимальных решений определяется из соотношения

$$U_{\min} = \{ \max \Phi [((W \beta G^T) \gamma (U_{\max})^T)] \}.$$

2.1. Здесь операция β определяется так:

$$U_{\beta} = w_{ij} \beta g_j = \begin{cases} 0, & \text{если } g_j > w_{ij} \text{ или } g_j = w_{ij} = 0; \\ \frac{g_j}{w_{ij}} & \text{в противном случае.} \end{cases}$$

Элемент вектора строки $U_{\beta} = (u_{\beta 1}, \dots, u_{\beta n})$ определяется из соотношения $u_{\beta i} = \max_{r=1}^n w_{ir} \alpha g_r$.

2.2. Операция γ для матриц U_{β} и U_{\max}^T определяется так:

$$U_{\gamma} = u_{\beta i} \gamma u_{i \max} = \begin{cases} 0, & \text{если } u_{i \max} \neq u_{\beta}; \\ u_{i \max}, & \text{если } u_{i \max} = u_{\beta}. \end{cases}$$

Элементы матрицы $U_{\gamma} = |u_{\gamma ij}|_{n \times n}$ определяются из соотношения:

$$u_{\gamma ij} = (u_{i \max} \gamma u_{\beta j})_{\substack{j=1, \dots, n, \\ i=1, \dots, n}}$$

2.3. Операция $\Phi(U_{\gamma})$ образует множество матриц $\Phi(U_{\gamma}) = \{\phi(U_{\gamma k})\}$, полученных из U_{γ} по определенному правилу:

- матрицы $\phi(U_{\gamma k})$ формируются таким образом, чтобы в каждом столбце матрицы содержался только один ненулевой элемент, а все остальные элементы обнуляются. В этом случае сумма элементов для каждого столбца матрицы $\phi(U_{\gamma k})$ будет равна ненулевому элементу;
- любая матрица $\phi(U_{\gamma k}) \in \Phi(U_{\gamma})$ содержит уникальное сочетание ненулевых элементов столбцов исходной матрицы $U_{\gamma k}$.

2.4. Операция $\max(\Phi(U_{\gamma}))$ формирует все минимальные решения обратной задачи путем взятия максимума по строкам для каждой матрицы $\phi(U_{\gamma k}) = |u_{\gamma kij}|$. Минимальные решения определяются из соотношения $U_k = \max_{i=1, \dots, n} \phi(U_{\gamma k}), U_k = (u_{k1}, \dots, u_{kn})$.

Таким образом, получаем векторы-столбцы решений (U_1, \dots, U_k) .

Применяя к элементам векторов решений обратное отображение φ^{-1} , получаем вектор лингвистических значений решения обратной задачи, т. е. $Z_k^* = \varphi^{-1}(U_k) = (z_{k1}, \dots, z_{kn}), \forall z_{ki} \in Z_i$.

Все решения обратной задачи представляются в виде точек с координатами элементов векторов решений в структурированном семантическом пространстве $CF(3)$. Тогда решения (точки в семантическом пространстве) попадают в разные области $SS(d^H)$, характеризующие классы состояний с разными именами d^H . Таким образом, формальное решение обратной задачи дает множество имен классов решений, которые структурируются по вложению областей $SS(d^H)$, соответствующих этим именам в виде качественной онтологии – концептуального каркаса решений.

Формальные имена классов решений определены математическими символами d^H . Для определения имен классов решений в предметной области в работе [20] были рассмотрены методы, основанные на исследованиях процессов классификации [23] и категоризации [24], выполненных психологами. Предложен метод определения составного имени нового класса, суть которого заключается в добавлении к имени базового класса d^0 оценки отличительного признака или признаков. Поясним этот метод.

В семантической модели определена область базового класса $SS(d^0) = \times_i \Delta_i$, где $\Delta_i = z_{i0} \pm \varepsilon_i$ – интервал толерантности по признаку i , $z_{i0} \in Z_i$ – значение признака в начальном состоянии, $\pm \varepsilon_i$ – определяет границы интервала толерантности по признаку f_i . Решение обратной задачи – это вектор $Z_k = (z_{k1}, \dots, z_{kn}), z_{ki} \in Z_i$. Результат решения обратной задачи в семантической модели (3) будем представлять в виде вектора $A_k = (a_{k1}, \dots, a_{kn}), a_{ki} \in \{-1, 0, 1\}$, где $a_{ki} = -1$, если $u_{ki} < z_{i0} - \varepsilon_i$; $a_i = 1$, если $u_{ki} > z_{i0} + \varepsilon_i$; $a_i = 0$, если $u_{ki} \in z_{i0} \pm \varepsilon_i$.

В векторе решения $A_k = (a_{k1}, \dots, a_{kn}), a_{ki} \in \{1, 0, -1\}$, компонента a_{ki} качественно характеризует оценку значения i -го параметра (f_i) в решении обратной задачи: т. е. если $a_{ki} = 1$, то параметр имеет большое значение, а если $a_{ki} = -1$, то параметр имеет малое значение. Этот вектор можно представить в виде вектора с лингвистическими оценками $L_k = (l_{k1}, \dots, l_{kn})$, где $l_{ki} = \langle \text{Большой} \rangle$, если $a_{ki} = 1$, $l_{ki} = \langle \text{Малый} \rangle$, если $a_{ki} = -1$. Тогда составное искусственное имя класса решений будет иметь вид

$$d_k^H = d^0 \ \& \ \underset{i(a_{ki} \neq 0)}{l_{ki}}$$

Например, если имя базового класса $d^0 = \langle \text{Инфляция} \rangle$, то возможные составные имена новых классов по признаку «Уровень инфляции» могут быть такими: $d_k^1 = \langle \text{Инфляция высокая} \rangle$ или $d_k^2 = \langle \text{Инфляция низкая} \rangle$.

Этот метод определения составного искусственного имени дает для каждого решения из множества $Z_k, \forall k$, решение, выраженное на ограниченном естественном языке – это имя базового класса решений d^0 и перечисление оценок (l_{ki}) значений признаков отличных от значений признаков базового класса.

В психологии возможен и другой метод определения имени класса решения. Он основан на психологической теории прототипов (категорий) [24]. В этой теории считается, что имя прототипа определяется именем наиболее характерного и часто употребляемого имени представителя этого класса. Отметим, что прототип имеет социокультурный контекст. Это означает, что имя прототипа в разных социальных или культурных сообществах может иметь разный смысл. Смысл в определении Г. Фреге [3] – это информация об объекте, иначе – множество его свойств.

В нашем случае прототипом класса решений с именем d_k^H считаются слова (знаки-символы), которые часто употребляются с составным именем класса решений d_k^H в контексте предметной области.

Особенность семиотических моделей заключается в том, что в условиях неопределенности и неполноты знаний они позволяют представить множество альтернативных синтаксических моделей ситуации в виде частично упорядоченного множества имен классов состояний семантической модели – концептуального каркаса предметной области.

Недостаток субъективных качественных семиотических моделей – это сложность их верификации и множественность субъективных интерпретаций результатов моделирования. Суть этой проблемы и методы ее решения описаны в работах [25, 26]. В настоящей работе для поддержки интерпретации решений в предметной области предлагается применить методы обработки естественного языка корпуса текста, релевантного предметной области и полученного из сети Интернет.

4. МОНИТОРИНГ СОСТОЯНИЯ СИТУАЦИИ

Под мониторингом понимают процесс наблюдения и регистрации состояния некоторого объекта или ситуации. Когда речь идет о наблюдении за

состоянием технического объекта, параметры которого могут быть измерены и представлены в числовых шкалах, реализация процесса мониторинга, как правило, не вызывает затруднений. При мониторинге социальных и политических ситуаций, множество параметров которых представлено в лингвистических шкалах, отражающих мнения, точки зрения, поведение социальных групп или отдельных политических деятелей, программная реализация наблюдения за состоянием ситуации значительно усложняется.

Применение для определения состояния ситуации полного лингвистического анализа (*Knowledge Acquisition*), предполагающего морфологический, синтаксический и семантический анализ текста, результатом которого является семантическая сеть концептов предметной области, из-за теоретических сложностей обработки естественного языка оказывается нецелесообразным.

Подсистему мониторинга состояния ситуации можно рассматривать как интерпретатор входного языка в семиотической архитектуре системы поддержки принятия решений [13].

Обобщенная структура подсистемы для решения задачи мониторинга состояния ситуации с использованием технологий обработки текстовой информации, основанная на субъективной семиотической модели, показана на рис. 6.

Здесь есть две основные подсистемы:

- подсистема получения и обработки неструктурированных данных из сети Интернет,
- подсистема семиотической модели ситуации.

Мониторинг состояния ситуации основан на субъективной семиотической модели ситуации (2)–(4). Параметры этой модели используются как параметры в подсистеме информационного поиска (*Information Retrieval*) для получения неструктурированных данных о текущем состоянии ситуации из сети Интернет. Под текущим состоянием понимается состояние ситуации в момент наблюдения $Z(t)$. Текущее состояние может отличаться от начального состояния $Z(0)$. Анализируются неструктурированные данные новостных лент.

Для работы этой подсистемы имена параметров ситуации семиотической модели $f_i \in F$ и их лингвистические значения Z_i используются для построения базы образцов $\langle \{f_i\}, \{Z_i\} \rangle$. Имена параметров ситуации семиотической модели $f_i \in F$ и имя базового класса d^0 используются при формировании запроса для подсистемы *Information Retrieval*. Из полученной информации извлекается текстовая информация о значениях параметров, т. е. о текущем состоянии ситуации.

В нашем случае будем ориентироваться на технологию извлечения из текста информации (*Information Extraction*), представленную в работах [27–29]. Эта технология ориентирована на работу с семиотической моделью ситуации и использует параметры синтаксической модели: $\{f_i\}$ – множество имен параметров и $\{Z_i\}$ – множество их возможных значений. Метод основан на построении образцов на корпусе текста предметной области. При построении образцов в предложениях предметной области определяется опорный элемент – это имя

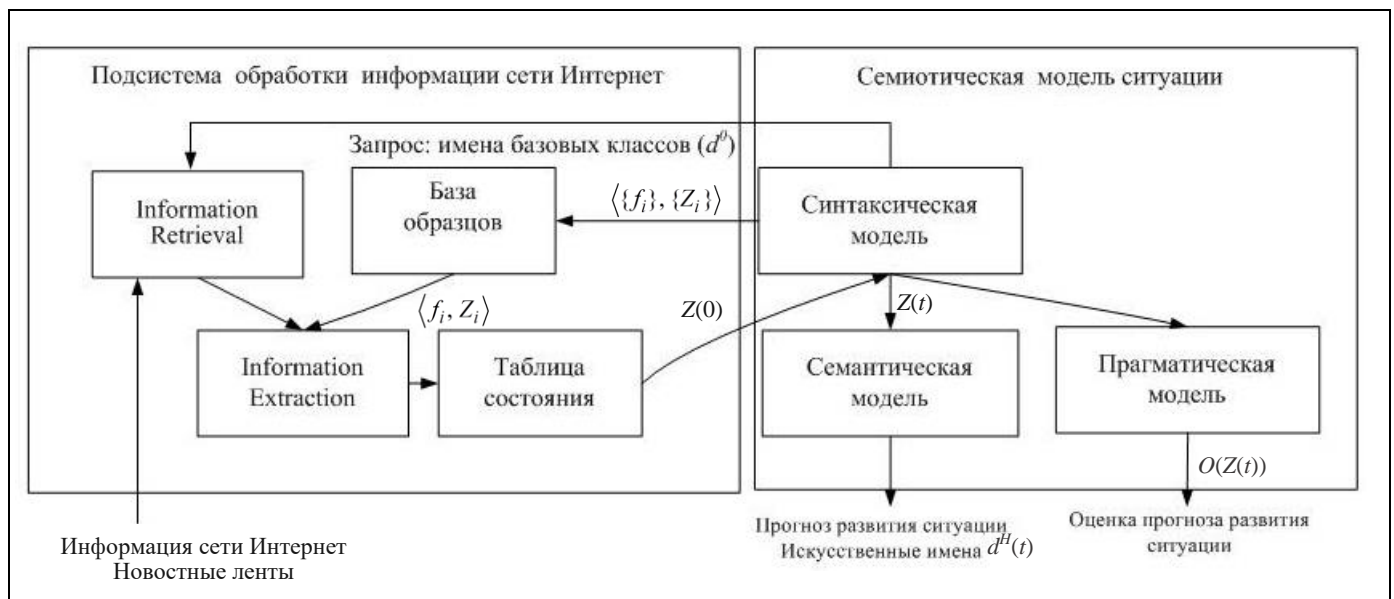


Рис. 6. Обобщенная структура подсистемы для решения задачи мониторинга состояния ситуации

параметра семиотической системы f_i . В предложении слева или справа от опорного элемента эксперт с помощью программы разметки текста определяет текстовые значения этого параметра.

Например, для параметра «Социальная напряженность» возможные текстовые значения таковы: «Единичный пикет», «Многочисленный митинг» и т. д. Этим текстовым значениям эксперт ставит оценку из множества возможных лингвистических оценок: «Очень сильно растет», «Сильно растет», «Растет», «Слабо растет», «Не меняется», «Слабо падает», «Падает», «Сильно падает», «Очень сильно падает». Например, текстовое значение «Единичный пикет» относительно текущего состояния может иметь экспертную оценку «Слабо растет», а «Многочисленный митинг» – оценку «Сильно растет». В этой технологии шкала значений каждого параметра Z_i также имеет шкалу оценок лингвистических значений, которые совпадают с ранее перечисленными оценками. В этом случае любой текстовой оценке, выявленной экспертом в корпусе текста, ставится в соответствие лингвистическое значение параметра шкалы Z_i , если совпадают их оценки. Таким образом, формируется текстовый образец – это имя параметра, текстовое значение в корпусе текста и соответствующее ему лингвистическое значение шкалы этого параметра.

Качество метода Information Extraction зависит от количества построенных образцов. Но построение каждого образца требует большой экспертной работы. При этом каждый текстовый образец уникален и может быть распознан только в конкретном тексте. Для сокращения объема экспертной работы и повышения качества выявления фактографической информации авторы предлагают использовать интеллектуальный тезаурус, выявляющий синонимы возможных текстовых значений параметров.

Итоговый образец формируется в результате концептуальной сборки текстового образца с учетом синонимов текстовых значений, а также с учетом концептов онтологии предметной области (учитывается связь «общее – частное»). В базе образцов представляются концептуальные образцы, которые с учетом синонимии и связей «часть – целое» в онтологии позволяют распознавать классы схожей фактографической информации.

Таким образом, результат работы подсистемы Information Extraction – это вектор параметров ситуации, полученный в результате извлечения лингвистических значений параметров: $Z^* = (z^*_{1h}, \dots, z^*_{nq}), z^*_{ij} \in Z_i$. Этот вектор нормализуется,

$\varphi: Z^* \rightarrow X^*, X^* = (x^*_{1h}, \dots, x^*_{nq}), x^*_{ij} \in [0, 1]$, после чего передается в таблицу состояния ситуации.

Если вновь полученное состояние Z^* отличается от текущего $Z(0)$, то в семиотической модели рассчитывается прогноз развития ситуации: в синтаксической модели это вектор значений $Z(n)$; в семантической модели – имя класса состояния ($d^H \in CF$), в прагматической – оценка нового состояния $O(Z^*)$.

Основные элементы подсистемы мониторинга – это технологии Information Retrieval и Information Extraction. Эти технологии исследуются множеством ученых, предлагающих методы и алгоритмы их реализации. Анализируя оценки качества этих технологий, приведенные в литературе, можно понять и качественно оценить эффективность предлагаемого подхода мониторинга, основанного на семиотической модели.

Технология Information Retrieval представляется в виде множества поисковых интернет-сервисов, позволяющих извлечь информацию из сети Интернет на основе запросов. Запросы включают параметры семиотической модели – это названия параметров и имена базовых классов.

Качество этой технологии – полнота и релевантность получаемой по запросу информации – обеспечивают разработчики этих сервисов, а возможность использования результатов поддерживается библиотеками для разных языков программирования. Это позволяет применять данную технологию в программных разработках конечных пользователей, решающих конкретные задачи.

Технология Information Extraction выделяет из текста информацию, шаблон которой представлен в базе образцов, включающих в себя параметры семиотической модели. Для структурированных данных, в которых явным образом удается распознать шаблон, качество работы этой подсистемы считается удовлетворительным.

Однако с неструктурированными данными (когда шаблон отсутствует) для работы этой системы приходится решать ряд специфических лингвистических задач: определение именованных существительных, разрешение кореферентности, построение связей, сценариев. Все эти задачи сложны: например, даже приближенное решение задачи кореферентности возможно только в некоторых предметных областях при условии использования базы знаний этой области [29]. Так, по результатам конференции Message Understanding Conference (MUC-6; Колумбия, Мэриленд, США, 1995 г.) лучшие результаты при решении задачи кореферентности достигали 59 % полноты при 72 % точ-



ности. Результаты работы человека при этом были оценены в 80 % [30]. Эти показатели считаются некоторым пределом качества этой технологии при анализе неструктурированных данных, отражающих свойства естественного языка. Дальнейшее повышение качества работы этой подсистемы с неструктурированными данными связаны с большими затратами [30].

Интерес представляют методы извлечения родовидовых отношений из текста для пополнения таксономий, тезаурусов и онтологий предметной области. Были организованы многочисленные международные конференции (соревнования) алгоритмов извлечения гиперонимов из текста для автоматического или автоматизированного пополнения существующих таксономий для английского и других западноевропейских языков [31–33].

В рамках конференции «Диалог-2020» (Москва, 2020 г.) была поставлена задача извлечения гиперонимов для автоматизированного пополнения тезауруса русского языка RuWordNet [34]. Задача заключалась в нахождении гиперонимов для некоторого целевого слова (существительного или глагола) на основе анализа корпуса текста [35]. Разработчиками предложены комбинированные методы, основанные на вычислении векторов совместной встречаемости целевого слова и множества слов корпуса текста [35]. Выделение множества кандидатов-гиперонимов из вектора совместной встречаемости осуществляется с применением различных приемов (взвешивание слов на основе эвристики, нахождение близости векторов слов корпуса текста и векторов известных таксономий и тезаурусов, размеченных вручную).

Для выделения гиперонимов также используются словари (например, Викисловарь), лексические шаблоны, предварительно обученные мультязычные нейросети (R-BERT) [36] и др.

В настоящее время имеется большое количество коммерческих систем, реализующих технологию Information Extraction, большинство из которых предполагает предварительную подготовку, структуризацию корпуса текста и извлечение числовой информации о значениях некоторых параметров.

Интерес представляет модульная система обработки естественного языка Gate (General Architecture for Text Engineering), разработанная на факультете информатики Шеффилдского университета (Великобритания) [37]. Для задачи Information Extraction на базе архитектуры Gate было разработано приложение ANNIE (ANearly-NewIESystem) [38].

В настоящее время применение технологии Information Extraction в задачах мониторинга состояния ситуации совместно с субъективной семиотической моделью целесообразно для предварительной структуризации корпуса текста и построения образцов для выделения фактов изменения состояния ситуации. В этом случае снимается значительная часть рутинной работы аналитика в задачах мониторинга состояния ситуации.

5. ПРИНЯТИЕ РЕШЕНИЙ В СЕМИОТИЧЕСКОЙ СИСТЕМЕ

Результатом работы подсистемы мониторинга является оценка прогноза развития ситуации $O(Z(n))$. Если в прогнозе развития ситуации оценка $O(Z(n))$ хуже текущей $O(Z(0))$, то в этом случае возникает задача принятия решений. В работе [20] задача принятия решений сводилась к решению обратной задачи в семиотической модели. При этом задается вектор цели $G = (g_1, g_2, \dots, g_n)$ (см. § 3), элементы которого содержат значения параметров, которые с точки зрения эксперта позволят улучшить прагматическую оценку развития ситуации, т. е. $O(G) > O(Z(n))$. В результате решения обратной задачи для заданной цели в синтаксической модели получается множество решений – $U_{\max}, U_1, \dots, U_k$. Данные решения – это управляющие воздействия (альтернативы решений), применение которых позволит достичь цели и, следовательно, улучшить прагматическую оценку. Эти решения в семантической модели представлены составными именами классов решений d_k^H . Эти составные имена представлены на внутреннем формальном языке эксперта, создавшего субъективную экспертную модель. В условиях неопределенности аргументация, обоснование и выбор экспертом приемлемого решения, полученного с помощью субъективной модели, возможны в рамках его знаний, которые могут быть неполными и противоречивыми. В этом случае для поддержки работы эксперта необходима внешняя база знаний, в качестве которой могут выступать неструктурированные данные (свободный текст) сети Интернет.

Задача подсистемы принятия решений, основанной на семиотической модели, заключается в поиске в сети Интернет интерпретаций искусственных составных имен классов решений и их объяснение.

Обобщенная структура подсистемы для решения задачи принятия решений с использованием технологий обработки текстовой информации, основанная на субъективной семиотической модели, показана на рис. 7.

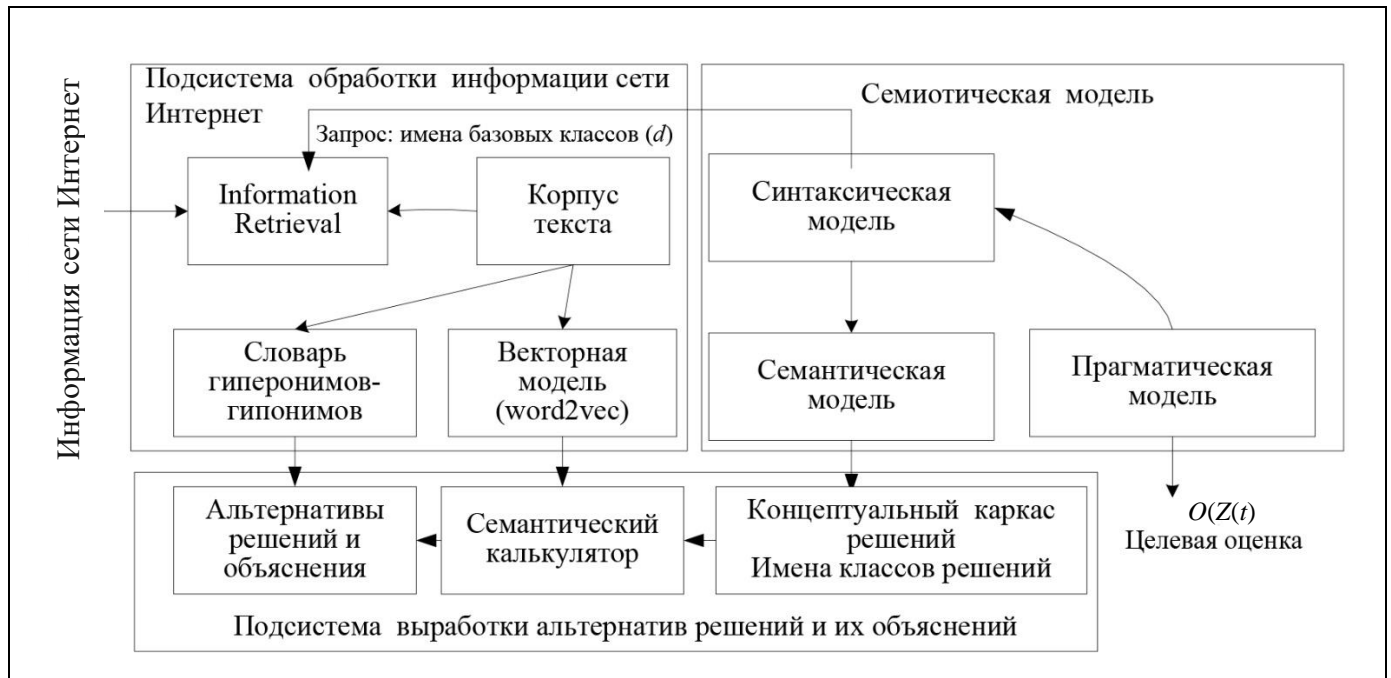


Рис. 7. Обобщенная структура подсистемы для принятия решений

Система принятия решений включает в себя основные подсистемы:

- подсистему семиотической модели,
- подсистему обработки неструктурированных данных сети Интернет,
- подсистему выработки альтернатив решений и их объяснений.

5.1. Подсистема обработки неструктурированных данных сети Интернет

Подсистема обработки неструктурированных данных сети Интернет включает в себя следующие блоки: информационного поиска информации в сети Интернет; корпус текста; словарь гиперонимов – гипонимов; векторную модель.

Корпус текста. Задача блока информационного поиска в сети Интернет (Information Retrieval) заключается в получении как можно большей по объему информации, релевантной запросу. Запрос на поиск информации в сети Интернет включает в себя имена базовых понятий семиотической модели. В результате получаем корпус текстов, релевантный имени базового понятия семиотической модели. Однако для построения векторной модели word2vec [39] корпус текста необходимо нормализовать с учетом синтаксических связей родовидовых отношений. Для этого используется синтаксическое окно, выделяющее в корпусе текста существительные, которые приводятся к нормальной форме (именительный падеж единственного числа).

Векторная модель. Векторная модель корпуса текста основана на технологии дистрибутивного анализа текста. В работе [40] дистрибутивный анализ определен как метод исследования языка, основанный на изучении окружения (дистрибуции, распределения) отдельных единиц в тексте и не использующий сведений о лексическом или грамматическом значении этих единиц. При дистрибутивном анализе каждое слово (словосочетание) в некотором тексте представляется в виде вектора слов, употребляемых совместно с этим словом в этом контексте. Каждое из слов в этом векторе характеризуется числом, отражающим частоту их совместной встречаемости в этом контексте. В работе [41] сформулирована гипотеза, согласно которой лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения векторов слов, употребляемых совместно.

Формально эту технологию можно представить следующем виде: задан корпус текстов в виде множества предложений, характеризующих предметную область. При дистрибутивном анализе синтаксические связи в предложениях игнорируются, тогда предметная область характеризуется множеством слов всех предложений $Tr = \{v_{gh}\}$, где g – номер предложения, h – номер слова в предложении (корпус текста). На множестве всех слов определены слова без их повторов – подмножество слов $V \subseteq Tr$, которое называется словарем предметной области, $V = \{v_r\}$, $r = 1, \dots, q$, – индексы слов в словаре. Между словами словаря определя-



ется характеристика их совместной встречаемости в данном контексте, т. е. определяется отношение

$$R_{w2v}: V \times V \rightarrow r_{ij},$$

где $r_{ij} \in [0, 1]$ характеризует совместную встречаемость слов v_i и v_j в рассматриваемой предметной области.

Для каждого слова словаря $v_r \in V$ определяется вектор $R_p = (v_1/r_{p1}, \dots, v_q/r_{pq})$, $p = 1, \dots, q$, который характеризует его совместную встречаемость r_{pi} с другими словами (v_1, \dots, v_q) рассматриваемой предметной области и называется контекстным вектором. Косая черта в контекстном векторе используется как разделитель. Перед разделителем стоит слово из словаря, а после разделителя – характеристика совместной встречаемости с другими словами.

В настоящее время в качестве метода построения отображения R_{w2v} применяется метод, основанный на машинном обучении искусственной нейронной сети [39], при котором обучающая выборка – это корпус текста Tr .

В технологии word2vec определены операции с векторами слов, позволяющие определить новые векторы, определяющие частоту совместного употребления сочетания отдельных слов словаря V с другими словами предметной области. Операции в технологии word2vec можно представить в виде отображения

$$w2v:(\bullet)(v_1, \dots, v_p) \rightarrow R_w^*,$$

где R_w^* – результирующий вектор, который характеризует частоту совместного употребления слов (v_1, \dots, v_n) с другими словами предметной области.

Здесь (\bullet) – операции, определенные в word2vec как `positive()` и `negative()`. Операция `positive()` определяет результирующий вектор R_w^* , характеризующий частоту совместного употребления слов в аргументе этой операции с другими словами предметной области, а операция `negative()` определяет вектор R_w^* , характеризующий частоту слов предметной области, которые не употребляются со словами, включенными в аргумент операции `negative()`.

Как работают такие векторные модели, можно понять на примере браузера Google: в строку поиска вводится одно слово, с помощью векторной модели получаем подсказку в виде вектора слов, часто употребляемых с этим словом; добавляем еще слово – получаем подсказку в виде вектора слова (слов), часто употребляемых с двумя словами, и т. д. Все вводимые слова – это аргументы операции `positive()`. Но в технологии word2vec можно искать слова, которые не употребляются

совместно. Эти слова добавляют в аргумент операции `negative()`. Результат – тоже вектор слов.

Контекстные векторы в этой технологии включают в себя слова и характеристику их совместной встречаемости с другими словами предметной области. В настоящей работе, считается, что слова – это имена знаков (Г. Фреге [3]), которые обозначают реальный объект или ситуацию и определяют его свойства – смысл объекта или ситуации. В определении Г. Фреге смысл – это информация об объекте, иначе – множество его свойств в контексте употребления слова. Зная слово, обозначающее объект, можно определить его свойства – смысл.

Другими словами, в технологии word2vec определена функция

$$w2v(\text{positive}(v_1, \dots, v_s); \text{negative}(v_q, \dots, v_n)) = R_w^*, \quad (5)$$

где $(v_1, \dots, v_s) \in V$ – слова в аргументе операции `positive`, $(v_q, \dots, v_n) \in V$ – слова в аргументе операции `negative`.

Ранее было определено, что именем класса решений можно считать имя прототипа, которое определяется по частоте употребления для обозначения категории. Контекстный вектор содержит информацию о совместной встречаемости слов, поэтому здесь предполагается, что слова, имеющие большие значения совместной встречаемости со словами, включенными в аргумент функции (5), могут быть именами прототипов.

Словарь гиперонимов – гипонимов. В словаре гиперонимов – гипонимов отражены родовидовые отношения, выделяемые из корпуса текста предметной области. Для выделения таких отношений используются лексико-синтаксические шаблоны [42]. Лексико-синтаксический шаблон – это структурный образец языковой конструкции, отражающий ее лексические и поверхностно-синтаксические свойства. В общем случае шаблон определяет последовательность элементов, из которых должна состоять описываемая языковая конструкция, и задает условия грамматического согласования этих элементов.

В научной литературе встречается много работ, посвященных выделению родовидовых отношений в английских, русских текстах, оценкам качества обнаружения отношений, построения и отладки самих шаблонов и их приложений. В работе [43] предложены шаблоны, которые, по словам авторов, позволяют выделить из текста 78,5 %, содержащихся в них родовидовых отношений.

Предложенные в работе [43] шаблоны взяты за основу для построения алгоритмов выделения родовидовых отношений из текста, характеризующего предметную область. При разработке алгоритма

мов, реализующих шаблоны, использовались морфологические характеристики слов в анализируемых предложениях. В частности, считается, что гипероним и гипоним – это существительные, которые должны иметь одинаковую характеристику одушевленности. В шаблонах использовалось правило согласования падежных окончаний гиперонимов и гипонимов в предложении.

Использование правил, основанных на морфологическом анализе, позволяет повысить качество выделения родовидовых отношений в предложениях.

В дистрибутивном анализе слова, находящиеся слева и справа от анализируемого слова, равнозначны, поскольку исключаются синтаксические связи между ними. Для устранения этого недостатка дистрибутивного анализа в работах [44, 45] предложено использовать отдельные словари для слов модели и контекстов. В словарь гиперонимов – гипонимов будем включать контекст, поясняющий родовидовую связь.

Применение лексико-синтаксических шаблонов позволяет сформировать словарь гиперонимов – гипонимов предметной области и связанный с этими словами контекстный словарь [46]. В этой работе контекстный словарь расширен до толкового словаря, куда включены предложения, в которых были выделены родовидовые отношения в результате применения шаблонов. Добавление контекста в словарь гиперонимов – гипонимов помогает выбрать гипероним в качестве прототипа, ориентируясь на контекст его употребления в предметной области.

Так формируется словарь гиперонимов – гипонимов и предложений, описывающих родовидовые отношения. Таким образом, определена структура в виде кортежа

$$\langle \text{HYPER}, \text{HYPO}, \text{Contex} \rangle. \quad (6)$$

Элементы кортежа – HYPER, HYPO – это множества гипонимов и гиперонимов предметной области, приведенных к нормальной форме, Contex – это множество предложений, в которых родовидовые отношения были определены.

Отметим, что родовидовые отношения, включенные в словарь, отражают структуру знаний о предметной области, по сути это элементы онтологии предметной области. Ранее мы определили в семантической модели качественную онтологию плохо определенной области в виде идеализированной концептуальной структуры с искусственными именами – концептуальным каркасом (3). Отметим, что выявленные гиперонимы могут заменить искусственные имена класса состояний концептуального каркаса, а гипонимы – служить

экземплярами этого класса состояний. Фактически выявляя родовидовые отношения, мы пытаемся определить возможные имена классов состояний идеализированного концептуального каркаса, полученные из корпуса текста предметной области.

5.2. Подсистема выработки альтернатив имен классов решений и их объяснений

Подсистема выработки альтернатив решений и их объяснений включает в себя: концептуальный каркас решений, семантический калькулятор, блок определения и объяснение альтернатив имен классов решений.

Концептуальный каркас решений. Формальное решение обратной задачи, полученное в семиотической модели, дает множество имен классов решений, структурированное в виде качественной онтологии – концептуального каркаса решений. Эти решения – знаки-символы, имеющие имя и содержание, выражены на внутреннем языке семиотической модели. Эти решения необходимо интерпретировать в предметной области.

Результат решения обратной задачи в семантической модели (3) представляется в виде вектора $A_k = (a_{k1}, \dots, a_{kn})$, $a_{ki} \in \{1, 0, -1\}$, где a_{ki} – качественно характеризует оценку значения i -го параметра (f_i) в решении обратной задачи, т. е. $a_{ki} = 1$ – параметр имеет большое значение, $a_{ki} = -1$ – параметр имеет малое значение.

Например, если вектор решения равен $(1, 0, 0)$, то это означает, что первый признак при решении обратной задачи получил значение, существенно превышающее значение этого признака в базовом понятии d^0 , а два других признака не изменились. Этот вектор определяет область семантического пространства $SS(d_k^H)$ и соответственно имя класса решения. Все имена классов решений (d_k^H) могут быть представлены в виде частично упорядоченного множества по вложениям областей семантического пространства, которые они определяют, (d_k^H, \leq) . Ранее было показано, что все имена классов решений – это составные имена, которые мы далее будем использовать для поиска в корпусе текста имен прототипов с помощью семантического калькулятора.

Семантический калькулятор. Семантический калькулятор предназначен для определения вектора совместного употребления слов, включенных в составное имя класса решений обратной задачи, полученного в семиотической модели, и слов корпуса текста о предметной области. Работа семантического калькулятора основана на обученной дистрибутивной модели предметной области и



представляется в виде функции (5), аргументами которой являются искусственные имена классов решений обратной задачи. Составное имя решения содержит имя базового понятия и различные оценки изменения различных параметров: «*Большой*», «*Малый*» или синонимы этих оценок

В семантическом калькуляторе решение обратной задачи в семантической модели $A_k = (a_{k1}, \dots, a_{kn})$ представляется так:

$$\begin{aligned} &w2v(\text{positive}(d^0, f_i|a_{ki} = 1, \dots, f_s|a_{ks} = 1); \\ &\text{negative}(f_q|a_{kq} = -1, \dots, f_n|a_{kn} = -1)) = R_w^* \end{aligned}$$

где f_i, f_s – имена параметров модели, для которых элемент вектора решений $a_{ki} = 1$ включается в аргумент операции $\text{positive}()$, а f_q, \dots, f_n – имена, для которых $a_{kq} = -1$, включаются в аргумент операции $\text{negative}()$. В аргумент операции $\text{positive}()$ добавляется также и имя базового класса (d^0).

В результате вычисления получаем вектор слов $R_w^* = (v_i/r_{i1}, \dots, v_n/r_{in})$, в котором упорядочено совместное употребление слов (r_{ij}), параметров модели, которые определены в решении обратной задачи, и всех слов корпуса текста о предметной области, включенных в словарь $v_i \in V$. Ранее говорилось, что слова с большой частотой встречаемости могут рассматриваться как прототипы имени класса решений.

Альтернативы имен классов решений и их объяснение. Решения в семиотической системе – это возможные имена классов решений. Гиперонимы, которые были выделены в корпусе текста, могут быть именами классов решений, так как они определяют элементы онтологии предметной области. Поэтому для получения альтернатив классов решений находим пересечение гиперонимов из словаря (6) слов в векторе решения R_w^* . Возможные имена класса решений представим в виде кортежа

$$\langle (V \cap \text{HYPER}); \text{Context} \rangle,$$

где пересечение множества $V = \{v_i\} \in R_w^*$ и множества гиперонимов из словаря (5) дает множество имен классов решений, а Context (текст предложения) поможет выбрать нужное имя.

6. ПРИМЕР И ЭКСПЕРИМЕНТЫ

Экспериментальная проверка предложенной семиотической системы проводилась для подсистемы поддержки принятия решений [46]. Была разработана семиотическая модель социально-политической ситуации. В синтаксической модели определены: базовые понятия – «*Власть*» (d_1^0), «*Население*» (d_2^0), «*Экономика*» (d_3^0), «*Олигархи*»

(d_4^0); признаки этих понятий, $f_i \in F$; возможные значения признаков $Z_i \in Z$ и причинно-следственная сеть W .

Для заданной цели $O(Z(t))$ при решении обратной задачи у базового понятия «*Олигархи*», имеющего признаки «*Уровень недовольства*» и «*Уровень патриотизма*», увеличилось значение признака «*Уровень недовольства*». Таким образом, получили новое понятие d_4^1 «*Олигархи*» с высоким значением признака «*Уровень недовольства*». В семантической модели это решение формально представлено вектором $A_4 = (1, 0)$, которое обозначается искусственным именем $d_4^1 =$ «*Недовольные олигархи*».

Задача заключается в том, чтобы найти интерпретацию этого решения в предметной области.

Для этого на языке Python3 был разработан программный макет.

Для построения корпуса текста был разработан WebScraper, позволяющий извлечь из сети Интернет релевантную информацию. Считывалась информация со 150 URL (сайтов), кроме этого, в корпус текста включен текст книги, посвященной олигархам России [47]. В качестве поисковой системы использовалась библиотека Google – это googlesearch с параметрами поиска: имя базового понятия и количество ссылок на сайты в ответе, а для синтаксического анализа html-кода полученных страниц сайтов (парсинг) использовалась библиотека Python3-а -BeautifulSoup.

Для построения словаря гиперонимов – гипонимов были разработаны и отлажены лексико-синтаксические шаблоны. При построении шаблонов использовалась программа морфологического анализа русского языка Rymorphy2 [48].

Для получения векторной модели корпуса текста использовалась программа word2vec. Параметры для обучения модели word2vec следующие: модель обучения – skipgram, окно обучения – 5, количество итераций обучения – 10, метод агрегации – softmax, порог встречаемости слов – 3, размерность вектора слов – 150.

Словарь и векторная модель сохраняются в базе данных SQLite-3.

Для выработки альтернатив решений и их объяснений были разработаны SQL-запросы к базе данных SQLite-3, которые позволяют получить имена классов решений с комментариями.

Пример. Для нахождения интерпретации решения обратной задачи $d_4^1 =$ «*Недовольные олигархи*» нормализуем и подставляем его в семантический калькулятор:

$$w2v(\text{positive}(\langle \text{«Олигархи»}, \text{«Недовольство»} \rangle)) = R_w^*.$$

В результате на обученной векторной модели word2vec получаем вектор слов R_w^* , отражающий ча-

стоту совместного употребления слов «Олигархи» и «Недовольство»:

$R_w^* = (\text{Вред}/0,904; \text{Факт}/0,885; \text{Респондент}/0,873; \text{Эксперт}/0,872; \text{Аннексия}/0,866; \text{Сожаление}/0,863; \text{Положение}/0,852; \text{Фактор}/0,844; \text{Тренд}/0,833; \text{Претензия}/0,830; \text{Недоверие}/0,817; \text{Эффективность}/0,813; \text{Преимущество}/0,808; \text{Характер}/0,806; \text{Запрос}/0,805; \text{Причина}/0,805; \dots;)$

Далее получаем возможные имена классов решений путем пересечения вектора слов R_w^* с гиперонимами предметной области:

$(W \cap \text{HYPER}) = (\text{Вред}/0,904; \text{Факт}/0,885; \text{Сожаление}/0,863; \text{Положение}/0,852; \text{Претензия}/0,830; \text{Недоверие}/0,817; \text{Характер}/0,806; \dots;)$

Из примера видно, что к решению обратной задачи с именем «Недовольные олигархи» в векторе слов ближе всего находится слово «Вред» – совместная встречаемость 0,904. Поэтому возможный новый класс решений – это класс с именем «Вредные олигархи».

Для выбора имен классов решений экспертным способом анализируется контекст кандидата-гиперонима. Например, для слова «Недоверие» с совместной встречаемостью 0,817 имеется следующий контекст: «Между этой категорией бизнеса и условным «коллективным Путиным» сложилось устойчивое взаимное недоверие: первые всегда опасались отъема собственности, а «коллективный Путин» – нелояльности»¹. ♦

Таким образом, предложенная семиотическая архитектура системы поддержки принятия решения позволяет получить альтернативы имен классов решений и выбрать решение, анализируя текст, аргументирующий этот выбор.

Отметим, что корпус текста примера включал в себя порядка 30 000 предложений, характеризующих разные аспекты предметной области. Применение этого подхода позволило определить возможные имена классов решений, анализируя значительно меньшее число предложений. Это определяет эффективность метода, которая заключается в сокращении рутинной аналитической работы, повышает интеллектуальную продуктивность эксперта.

ЗАКЛЮЧЕНИЕ

Предложена семиотическая система поддержки принятия решений в сложных динамических системах в условиях неопределенности. Поддержка основана на извлечении, обработке и структуризации информации из сети Интернет, релевантной семиотической модели ситуации, включающей в

себя три связанные параметрами модели: синтаксическую, семантическую и прагматическую. Решение обратной задачи в семиотической модели представляется в виде качественной онтологии классов решений (концептуального каркаса). Методы определения и интерпретации имен классов решений основаны на извлечении из сети Интернет релевантной информации. С помощью лексико-семантических шаблонов в корпусе текстов предметной области определяются словари: родовидовых отношений (гиперонимов – гипонимов) и контекстный. На основе методов дистрибутивной семантики (word2vec) построен семантический калькулятор, позволяющий определить смыслы имен классов решений концептуального каркаса.

Экспериментальная проверка предложенной архитектуры показала ее работоспособность. Дальнейшие экспериментальные исследования будут направлены на повышение качества предложенного подхода с помощью методов увеличения объема корпуса текстов предметной области, применения словарей гиперонимов – гипонимов, имеющихся в открытом доступе, и дополнительного семантического анализа предложений, включающих в себя слова вектора решений, имеющих высокий уровень совместной встречаемости с составным именем класса решений.

ЛИТЕРАТУРА

1. *Поспелов Д.А.* Ситуационное управление: теория и практика. – М.: Наука, 1986. – 288 с. [*Pospelov, D.A.* Situatsionnoe upravlenie: teoriya i praktika. – М.: Nauka, 1986. – 288 s. (In Russian)]
2. *Поспелов Д.А., Осипов Г.С.* Прикладная семиотика // Новости искусственного интеллекта. – 1999. – № 1. – С. 9–35. [*Pospelov, D.A., Osipov, G.S.* Prikladnaya semiotika // Novosti iskusstvennogo intellekta. – 1999. – No. 1. – P. 9–35. (In Russian)]
3. *Frege, G.* On Sense and Nominatum // Readings in Philosophical Analysis. – 1949. – P. 85–102.
4. *Осипов Г.С., Чудова Н.В., Панов А.И., Кузнецова Ю.М.* Знаковая картина мира субъекта поведения. – М.: Физматлит, 2018. – 264 с. [*Osipov, G.S., Chudova, N.V., Panov, A.I., Kuznetsova, Yu.M.* Znakovaya kartina mira sub"ekta povedeniya. – М.: Fizmatlit, 2018. – 264 s. (In Russian)]
5. *Пирс Ч.С.* Что такое знак? // Вестник Томского государственного университета. Философия. Социология. Политология. – 2009. – № 3(7). – С. 88–95. [*Pirs, Ch.S.* Chto takoe znak? // Vestnik Tomskogo Gosudarstvennogo Universiteta. Filosofiya. Sotsiologiya. Politologiya. – 2009. – No. 3(7). – P. 88–95. (In Russian)]
6. *Моррис Ч.У.* Основания теории знаков. – Семиотика: антология. Изд. 2-е, перераб. и доп. М. – Екатеринбург. – 2001. – С. 45–97. [*Morris, Ch.U.* Osnovaniya teorii znakov. Semiotika: antologiya. Izd. 2-e, pererab. idop. – М.-Ekaterinburg. – 2001. – P. 45–97. (In Russian)]

¹ <https://carnegie.ru/commentary/76115> (Дата обращения 24.02.2022).



7. *Barron, T.M., Chiang Roger, H.L., Storey, V.C.* A Semiotics Framework for Information Systems Classification and Development // *Decision Support Systems*. – 1999. – No. 25. – P. 1–17.
8. *Andersen, P.B.* A Theory of Computer Semiotics: Semiotic Approaches to Construction and Assessment of Computer Systems. – Cambridge University Press, 2006. – 460 p.
9. *Goguen, J.* An Introduction to Algebraic Semiotics, with Applications to User Interface Design / In: *Computation for Metaphor, Analogy and Agents*, ed. by Ch. Nehaniv. – Springer Lecture Notes in Artificial Intelligence. – 1999. – Vol. 1562. – P. 242–291.
10. *Furtado, A.L., Casanova, M.A., Barbosa, S.D.J.* A Semiotic Approach to Conceptual Modelling / In: *ER 2014: Conceptual Modeling*, ed. by Yu E., Dobbie G., Jarke M., Puro S. – Lecture Notes in Computer Science. – 2014. – Vol. 8824. – P. 1–12.
11. *Массель Л.В., Массель А.Г.* Ситуационное управление и семантическое моделирование в энергетике // IV Международная конференция OSTIS: Тр. – Минск. – 2014. – С. 111–116. [*Massel', L.V., Massel', A.G.* Situatsionnoe upravlenie I semanticheskoe modelirovanie v ehnergetike // IV Mezhdunarodnaya konferentsiya OSTIS: Trudy. – Minsk. – 2014. – P. 111–116. (In Russian)]
12. *Массель Л.В.* Проблемы создания интеллектуальных систем семиотического типа для стратегического ситуационного управления в критических инфраструктурах // Информационные и математические технологии в науке и управлении. Научный журнал. – 2016. – № 1. – С. 7–27. [*Massel', L.V.* Problemy sozdaniya intellektual'nykh system semioticheskogo tipa dlya strategicheskogo situatsionnogo upravleniya v kriticheskikh infrastrukturakh // Informatsionnye I matematicheskie tekhnologii v nauke I upravlenii. Nauchnyi zhurnal. – 2016. – No. 1. – P. 7–27. (In Russian)]
13. *Эрлих А.И.* Прикладная семиотика и управление сложными объектами // Программные продукты и системы. – 1997. – № 3. – P. 3–11. [*Ehrlikh, A.I.* Prikladnaya semiotika i upravlenie slozhnymi ob"ektami // Programmnye produkty i sistemy. – 1997. – No. 3. – P. 3–11. (In Russian)]
14. *Кулинич А.А.* Семиотический подход в моделировании и принятии решений в плохо определенных сложных ситуациях // Матер. одиннадцатой международной конференции «Управление развитием крупномасштабных систем MLSD'2018». – ИПУ РАН. – 2018. – Т. 2. – С. 298–301. [*Kulinich, A.A.* Semioticheskii podkhod v modelirovanii i prinyatii reshenii v plokhooopredelennykh slozhnykh situatsiyakh // Materialy odinnadsatoi mezhdunarodnoi konferentsii «Upravlenie razvitiem krupnomasshtabnykh system MLSD'2018». – IPU RAN. – 2018. – Vol. 2. – P. 298–301. (In Russian)]
15. *Поспелов Д.А.* Логико-лингвистические модели в системах управления. – М.: Энергоиздат, 1981. – С. 232. [*Pospelov, D.A.* Logiko-lingvisticheskie modeli v sistemakh upravleniya. – M.: Ehnergoizdat. – 1981. – S. 232. (In Russian)]
16. *Борисов В.В., Круглов В.В., Федулов А.С.* Нечеткие модели и сети. – 2-е изд., стереотип. – М.: Горячая линия – Телеком, 2012. – 284 с. [*Borisov, V.V., Kruglov, V.V., Fedulov, A.S.* Nchetkie modeli i seti. – 2-e izd., stereotip. – M.: Goryachaya liniya – Telekom, 2012. – 284 s. (In Russian)]
17. *Kosko, B.* Fuzzy Cognitive Maps // *International Journal of Man-Machine Studies*. – 1986. – Vol. 24, iss. 1. – P. 65–75.
18. *Кулинич А.А.* Компьютерные системы моделирования когнитивных карт: подходы и методы // Проблемы управления. – 2010. – № 3. – С. 2–16. [*Kulinich, A.A.* Cognitive Maps Modelling Computer Systems (Approaches and Methods) // *Control Sciences*. – 2010. – No. 3. – P. 2–16. (In Russian)]
19. *Кулинич А.А.* Концептуальные каркасы онтологий слабо структурированных предметных областей // Искусственный интеллект и принятие решений. – 2014. – № 4. – С. 31–41. [*Kulinich, A.A.* Kontseptual'nye karkasy ontologii slabostrukturirovannykh predmetnykh oblastei // Iskusstvennyi intellekt i prinyatie reshenii. – 2014. – No. 4. – С. 31–41. (In Russian)]
20. *Кулинич А.А.* Решение обратной задачи в семиотических системах в условиях неопределенности // Сб. науч. тр. IX Междунар. науч.-практ. конф. «Интегрированные модели и мягкие вычисления в искусственном интеллекте». – (Коломна, 2019 г. – Т. 2. – С. 572–582. [*Kulinich A.A.* Reshenie obratnoi zadachi v semioticheskikh sistemakh v usloviyakh neopredelennosti // Sb. nauch. tr. IX Mezhdunar. nauch.-prakt. conf. «Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte». – Kolomna, 2019 g. – 2019. – Vol. 2. – P. 572–582. (In Russian)]
21. *Pedrycz, W.* Fuzzy Models and Relational Equations // *Math. Modeling*. – 1987. – No. 9. – P. 427–434.
22. *Pedrycz, W.* Algorithms for Solving Fuzzy Relational Equations in a Probabilistic Setting // *Fuzzy Sets and Systems*. – 1990. – No. 38. – P. 313–327.
23. *Выготский Л.С.* Мышление и речь // Собр. соч.: – Т. 2. – М.: Педагогика, 1982. – С. 5–361. [*Vygotskij, L.S.* Myshlenie i rech' // Sobr. soch.: – M.: Pedagogika, 1982. – Vol. 2. – P. 5–361. (In Russian)]
24. *Rosch, E.* Principles of Categorization / In: *Rosch, E., Lloyd B.D.* Cognition and categorization. – N.-Y.: Lawrence Erlbaum association, 1978.
25. *Абрамова Н. А.* Экспертная верификация при использовании формальных когнитивных карт. Подходы и практика // Управление большими системами. Специальный выпуск 30.1 «Сетевые модели в управлении». – М.: ИПУ РАН, 2010. – С. 371–410. [*Abramova, N.A.* Expert Verification in Formal Cognitive Map Application. Approaches and Practices / Large-Scale Systems Control. Special'nyj vypusk 30.1 «Setevye modeli v upravlenii». – M.: IPU RAN, 2010. – S. 371–410. (In Russian)]
26. *Abramova, N.A.* Interdisciplinary Approach to Verification in Decision-Making with Formal Methods / In: *Handbook on Psychology of Decision-Making: New Research. Series: Psychology of Emotions, Motivations and Actions*. – Hauppauge, USA: Nova Science Publishers, 2012. – P. 89–111.
27. *Рубашкин В.Ш.* Онтологии - проблемы и решения. Точка зрения разработчика // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конф. «Диалог 2007». – Бекасово, 2007 г. – С. 481–485. [*Rubashkin, V.Sh.* Ontologii – problemy i resheniya. Tochka zreniya razrabotchika // Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Trudy mezhdunarodnoj konferentsii «Dialog 2007». – Bekasovo, 2007. – S. 481–485. (In Russian)]
28. *Пивоварова Л.М.* Фактографический анализ текста в системе поддержки принятия решений // Вестник СПбГУ. Язык и литература. – 2010. – № 4. – С. 190–197. [*Pivovarova, L.M.* Faktograficheskij analiz teksta v sisteme podderzhki prinyatiya reshenij // Vestnik SPbGU. Yazyk i literatura. – 2010. – № 4. – S. 190–197. (In Russian)]. URL: <https://cyberleninka.ru/article/n/faktograficheskij-analiz-teksta-v-sisteme-podderzhki-prinyatiya-resheniy> (дата обращения: 19.05.2022).

29. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах. – М.: Наука, 1989. [Rubashkin, V.Sh. Predstavlenie i analiz smysla v intellektual'nykh informatsionnykh sistemakh. – М.: Nauka, – 1989. (In Russian)]
30. Grishman, R. Information Extraction: Techniques and Challenges // In: Information Extraction, ed. by Pazienza, M.T. – Lecture Notes in Artificial Intelligence. – Rome: Springer-Verlag, 1997. – Vol. 1299. – P. 10–27.
31. Camacho-Collados, J., Delli Bovi, C., Espinosa-Anke, L., et al. SemEval-2018 task 9: Hypernym Discovery // Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018). – New Orleans, Stroudsburg, 2018. – P. 712–724.
32. Jurgens, D., Pilehvar, M.T. SemEval-2016 task 14: Semantic taxonomy enrichment // Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). – San Diego, California, 2016. – P. 1092–1102.
33. Bernier-Colborne, G., Barriere, C. CRIM at SemEval-2018 task 9: A Hybrid Approach to Hypernym Discovery // Proceedings of the 12th International Workshop on Semantic Evaluation. – New Orleans, Louisiana, 2018. – P. 725–731.
34. Loukachevitch, N.V., Dobrov, B.V. RuThes Linguistic Ontology vs. Russian Wordnets // Proceedings of the Seventh Global Wordnet Conference. – Tartu, 2014. – P. 154–162.
35. Nikishina, I., Logacheva, V., Panchenko, A., Loukachevitch, N. RUSSE'2020: Findings of the First Taxonomy Enrichment Task for the Russian language // Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2020. – Moscow, 2020. – P. 579–595. – DOI: 10.28995/2075-7182-2020-19-579-595.
36. Tikhomirov, M.M., Loukachevitch, N.V., Parkhomenko, E.A. Combined Approach to Hypernym Detection for Thesaurus Enrichment // Computational Linguistics and Intellectual Technologies: Papers from the Annual Conference «Dialogue». – Moscow, 2020. – P. 736–746. – DOI: 10.28995/2075-7182-2020-19-736-746.
37. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications // Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), – Philadelphia, 2002.
38. Gate's ANNIE system. URL: <http://gate.ac.uk/ie/annie.html>. (Дата обращения 24.02.2022). [Accessed February 24, 2022]
39. Mikolov, T., Sutskever, I., View, M., et al. Distributed Representations of Words and Phrases and Their Compositionality // Advances in Neural Information Processing Systems: 27th Annual Conference on Neural Information Processing Systems. Proceedings of a meeting held December 5-8, 2013. – Lake Tahoe, Nevada, United States, 2013. – P. 3111–3119.
40. Лингвистический энциклопедический словарь // Под ред. Ярцева В.Н. – М.: Советская энциклопедия, 1990. [Lingvisticheskiy ehntsiklopedicheskiy slovar' // pod red. Yartseva, V.N. – М.: Sovetskaya ehntsiklopediya, 1990. (In Russian)]
41. Sahlgren, M. The Distributional Hypothesis. From context to meaning // Distributional Models of the Lexicon in Linguistics and Cognitive Science (Special issue of the Italian Journal of Linguistics), Rivista di Linguistica. – 2008. – Vol. 20, no. 1. – P. 33–53.
42. Большакова Е.И., Васильева Н.Э., Морозов С.С. Лексико-синтаксические шаблоны для автоматического анализа научно-технических текстов // 10-я Национальная конференция по искусственному интеллекту (КИИ-2006). Тр. конф. – М.: Физматлит, – 2006. – Т. 2. – С. 506–524. [Bol'shakova, E.I., Vasil'eva, N.Eh., Morozov, S.S. Leksiko-sintaksicheskie shablony dlya avtomaticheskogo analiza nauchno-tekhnicheskikh tekstov // 10-ya Natsional'naya konferentsiya po iskusstvennomu intellektu (KI-2006). Trudy konferentsii. – М.: Fizmatlit. – 2006. – Т. 2. – P. 506–524 (In Russian)]
43. Sabirova, K., Lukanin, A. Automatic Extraction of Hypernyms and Hyponyms from Russian Texts // CEUR Workshop Proceedings. – 2014. – Vol. 1197. – P. 35–40.
44. Levy, O., Goldberg, Y. Dependency-Based Word Embeddings // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Short Papers. – Baltimore, Maryland, USA, 2014. – P. 302–308.
45. Трофимов И.В., Сулейманова Е.А. Дистрибутивно-семантическая модель для выявления категориального сходства // Программные системы: теория и приложения. – 2018. – Т. 9, № :4(39). – С. 443–460. [Trofimov, I.V., Suleimanova, E.A. Distributivno-semanticheskaya model' dlya vyyavleniya kategorial'nogo skhodstva // Programmnye sistemy: teoriya i prilozheniya. – 2018. – Vol.9, no. 4(39). – P. 443–460. (In Russian)]
46. Kulnich, A. Semiotic Models in Monitoring and Decision Support Systems // In: RCAI 2021: Artificial Intelligence, ed. by Kovalev, S.M., Kuznetsov, S.O., Panov, A.I. – Lecture Notes in Computer Science. – Cham: Springer, 2021. – Vol. 12948. – P. 303–316.
47. Хоффман Д. Олигархи. Богатство и власть в новой России. – Изд-во Corpus, 2010. – 816 с. [Hoffman, D. Oligarkhi. Bogatstvoivlast' v novej Rossii. – Iz-vo Corpus, 2010. – 816 s. (In Russian)]
48. Korobov, M. Morphological Analyzer and Generator for Russian and Ukrainian Language's // Analysis of Images, Social Networks and Texts. – 2015. – P. 320–332.

Статья представлена к публикации членом редколлегии
В.Г. Лебедевым

Поступила в редакцию 1.03.2022,
после доработки 1.11.2022.
Принята к публикации 22.11.2022.

Кулинич Александр Алексеевич – канд. техн. наук, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, ✉ alexkul@rambler.ru.



APPLICATION OF SEMIOTIC MODELS TO DECISION-MAKING

A.A. Kulinich

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ alexkul@rambler.ru

Abstract. This paper introduces an approach to building decision support systems based on a semiotic domain model and natural language processing methods. The knowledge base of this model is a text corpus of linguistic information obtained from the Internet. The text corpus is relevant to the subject domain in which the subjective semiotic model of the situation is constructed. A method for solving the inverse problem in a semiotic system is proposed. The obtained solutions are interpreted in the subject domain using a semantic calculator. The semantic calculator extracts generic relations from the text corpus based on lexico-syntactic patterns and determines the frequency of joint occurrence of words in the solution based on the distributive analysis of the text corpus. The generalized structures of monitoring and decision-making subsystems with the semiotic model of the situation and natural language processing methods are described. A software layout of the decision-making subsystem is developed. The effectiveness of this approach is demonstrated by experiments.

Keywords: decision-making, semiotic system, subjective model, natural language processing, distributive analysis.