

Ревунков Г.И., кандидат технических наук, доцент

Гапанюк Ю.Е., кандидат технических наук, доцент

Нардид А.Н., аспирант

(Московский государственный технический университет им. Н.Э. Баумана)

СТРУКТУРА СИСТЕМЫ СБОРА И АНАЛИЗА ДАННЫХ ИНТЕРНЕТ-ИСТОЧНИКОВ НА ОСНОВЕ МЕТАГРАФОВОГО ПОДХОДА

В статье рассмотрено применение метаграфового подхода для описания структуры системы сбора и анализа данных Интернет-источников. Предложены основные элементы структуры системы и варианты их комбинирования.

Ключевые слова: сбор и анализ данных Интернет-источников, структура системы, метаграф.

THE STRUCTURE OF THE SYSTEM OF COLLECTION AND ANALYSIS OF INTERNET DATA SOURCES ON THE BASIS OF METAGRAPH APPROACH

In the article, the use metagraph approach to describe the structure of the system of collection and analysis of Internet data sources is discussed. The main elements of the system structure and variants for their combining are proposed.

Keywords: collection and analysis of Internet data sources, system structure, metagraph.

В настоящее время анализ информации в Интернет-пространстве является актуальной задачей. Сетевое информационное пространство весьма велико, и часто возникают задачи автоматизированного сбора информации из различных Интернет-источников. К этим задачам можно отнести задачи сбора и мониторинга данных, построения ситуаций и прогнозирования. В данной статье авторы предлагают унифицированный подход к решению перечисленных задач на основе метаграфовой модели.

Основные элементы структуры системы сбора и анализа информации в Интернет-пространстве представлены на рис. 1.

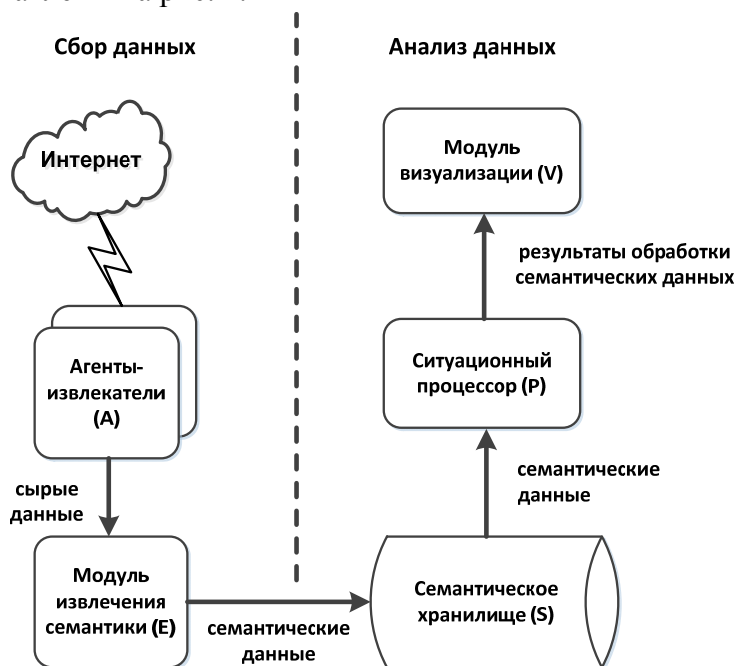


Рис. 1. Основные элементы структуры системы сбора и анализа информации в Интернет-пространстве.

Сбор информации осуществляется с помощью агентов-извлекателей (А), которые осуществляют извлечение информации с сайтов, из новостных лент, социальных сетей и других источников.

В результате работы агентов-извлекателей формируется поток сырых данных, который поступает в модуль извлечения семантики. Сырые данные могут представлять собой как текстовые, так и графические данные (изображения, видео).

Модуль извлечения семантики (Е) осуществляет извлечение семантики из данных (как текстовых, так и графических) и формирование единой обогащенной семантической модели на основании объединения семантики, извлеченной из различных данных.

В настоящее время и алгоритмы извлечения семантики, и способы ее хранения являются предметом активных исследований. Для извлечения семантики используются методы машинного обучения – данные подвергаются классификации и кластеризации, разбиваются на группы связанных тем, сообщениям приписываются определенные теги. В рамках семантического обогащения выстраиваются связи между различными данными, извлеченными из разных источников, в том числе между источниками различных форматов – текстовыми и графическими.

Извлеченные семантические данные помещаются в семантическое хранилище (S) на основе метаграфовой модели данных [1], которое должно представлять собой СУБД с графовой или гибридной моделью данных. В статье [2] показано, что на сегодняшний день не существует СУБД, которые бы в полной мере обеспечивали работу со сложными графовыми моделями, в частности, с метаграфовой моделью. Поэтому в статье [3] предлагается использование метаграфового хранилища на основе предикатного описания.

Данные из семантического хранилища поступают на вход ситуационного процессора (Р), который в зависимости от задачи реализует различные алгоритмы ситуационного анализа и прогноза на основе метаграфовой модели. Основой таких алгоритмов является использование метаграфовых агентов, которые детально рассмотрены в [4]. Поскольку метаграфовая модель также содержит числовые данные, то ситуационный процессор может использовать классические методы обработки числовых данных, а также методы на основе искусственного интеллекта и машинного обучения: символьную регрессию, генетические алгоритмы, разновидности нейросетевых алгоритмов.

Результаты обработки семантических данных могут быть представлены эксперту с использованием модуля визуализации (V). Данный модуль предназначен для удобного отображения результатов ситуационного анализа и прогноза. В частности, одним из методов визуализации являются картограммы (анаморфированные карты) [5].

Рассмотрев основные элементы системы сбора и анализа информации в Интернет-пространстве, разберем различные варианты комбинации элементов в структуру возможной системы.

Простейшая структура 1 предполагает однократное использование каждого модуля: $STR_1 = \langle A, E, S, P, V \rangle$.

Вариант структуры 2 предполагает комбинацию агентов-извлекателей и модулей извлечения семантики в зависимости от вида извлекаемой информации. Такую комбинацию назовем модулем-извлекателем: $EXT_i = \langle A, E \rangle$. Каждый модуль-извлекатель отвечает за отдельный вид извлекаемой информации: текстовая, графическая, видеоинформация. В этом случае структура системы 2 включает множество модулей-извлекателей: $STR_1 = \langle \{EXT_i\}, S, P, V \rangle$.

Наибольший интерес представляет комбинация семантического хранилища и ситуационного процессора. Назовем такой элемент семантическим преобразователем: $SP_i = \langle S, P \rangle$. Семантический преобразователь является базовым элементом для построения систем обработки сложных событий (Complex Event Processing – CEP) на основе метаграфового подхода.

В классической СЕР-системе на модули-обработчики событий и структуру событий не накладывается никаких ограничений. В СЕР-системе на основе метаграфового подхода (МСЕР) события представляются в виде метаграфов (MG), а модули-обработчики являются семантическими преобразователями (SP). При этом, используя терминологию статьи [4], МСЕР является контейнерным агентом, то есть представляет собой метаграф, вершины и метавершины которого являются семантическими преобразователями, а обмен данными между семантическими преобразователями производится в форме метаграфов:

$$МСЕР = MG, v_i \equiv SP_i, v_i \in V, mv_i \equiv SP_i, mv_i \in MV,$$

где МСЕР – СЕР-система на основе метаграфового подхода; MG – метаграф; v_i – вершина метаграфа; SP_i – семантический преобразователь; V – множество вершин метаграфа; mv_i – метавершина метаграфа; MV – множество метавершин метаграфа.

Тогда вариант структуры 3 на основе МСЕР может быть представлен как $STR_3 = \langle A, E, МСЕР, V \rangle$.

Вариант структуры 4, который комбинирует подходы 2 и 3, может быть представлен как $STR_4 = \langle \{EXT_i\}, МСЕР, V \rangle$. В этом случае для каждого вида информации используется свой модуль-извлекатель, после чего извлеченные семантические описания обрабатываются метаграфовой СЕР-системой.

Таким образом, использование метаграфового подхода позволяет унифицировать как элементы, так и возможные варианты структуры системы сбора и анализа данных Интернет-источников.

ЛИТЕРАТУРА

1. Самохвалов Э.Н., Ревунков Г.И., Гапанюк Ю.Е. Использование метаграфов для описания семантики и прагматики информационных систем. Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2015. Выпуск №1. С. 83-99.
2. Березкин Д.В., Ревунков Г.И., Гапанюк Ю.Е. Подходы к построению гибридных хранилищ данных для информационных систем. Гибридные и синергетические интеллектуальные системы: материалы III международного Пospelовского симпозиума / под. ред. д-ра техн. наук, проф. А.В. Колесникова. – Калининград : Изд-во БФУ им. И.Канта, 2016. – 458с.
3. Гапанюк Ю.Е., Ревунков Г.И., Федоренко Ю.С. Предикатное описание метаграфовой модели данных. Информационно-измерительные и управляющие системы. № 12, 2016.
4. Черненький В.М., Терехов В.И., Гапанюк Ю.Е. Структура гибридной интеллектуальной информационной системы на основе метаграфов. Нейрокомпьютеры: разработка, применение. 2016. Выпуск №9. С. 3-14.
5. Терехов В.И., Черненький И.М. Разработка варианта принятия решения с помощью метода анаморфирования. Информационно-измерительные и управляющие системы. № 12, 2016.

Ревунков Г.И., кандидат технических наук, доцент

Гапанюк Ю.Е., кандидат технических наук, доцент

Федоренко Ю.С., аспирант

(Московский государственный технический университет им. Н.Э. Баумана)

ОПИСАНИЕ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ МЕТАГРАФОВОГО ПОДХОДА

В статье рассмотрено применение метаграфового подхода для описания нейронной сети. Рассмотрено описание персептрона с использованием агента-функции и описание нейронной сети с использованием метаграфовых агентов. Показано, что предложенный подход позволяет реализовать режимы создания, изменения, обучения и штатной работы нейронной сети.

Ключевые слова: *нейронная сеть, персептрон, метаграф.*

DESCRIPTION OF NEURAL NETWORK USING METAGRAPH APPROACH

In the article, the description of neural network using metagraph approach is discussed. The description of the perceptron using the agent-function and description of neural network using metagraph agents are discussed. It is shown that the proposed approach allows implementing the routines of creation, restructuring, learning and the normal operation of a neural network.

Keywords: *neural network, perceptron, metagraph.*

В настоящее время комплексные графовые модели находят все более широкое применение в различных областях информатики. Одной из таких моделей является метаграфовая модель. Метаграфовую модель предлагается применять как средство для описания сложных сетей [1], как средство для описания семантики и прагматики информационных систем [2], как средство для описания гибридных интеллектуальных информационных систем [3].

Одним из основных элементов метаграфовой модели является метавершина. Метавершина представляет собой вложенный фрагмент метаграфа, который также включает собственные атрибуты и связи с другими вершинами (метавершинами). Этот подход соответствует принципу эмерджентности, то есть приданию элементу нового качества, несводимости элемента к сумме его составных частей. Фактически, как только вводится новое понятие в виде метавершины, оно «получает право» на собственные свойства, связи и т.д., так как в соответствии с принципом эмерджентности новое понятие обладает новым качеством и не может быть сведено к подграфу базовых понятий.

Целью данной статьи является применение метаграфового подхода для описания нейронной сети.

В соответствии с [3] по уровню активности метаграфовые структуры можно разделить на три уровня:

1. Представление данных в виде метаграфа; такой элемент не обладает самостоятельным поведением и предназначен для описания данных.
2. Агент-функцию, который, в частности, реализует поведение элементарного нейрона.
3. Метаграфовый агент, который реализует создание, изменение и обучение нейросети.

Для описания персептрона в виде агента-функции необходимо предварительно рассмотреть описание персептрона с использованием метаграфового подхода, что показано на рис. 1.

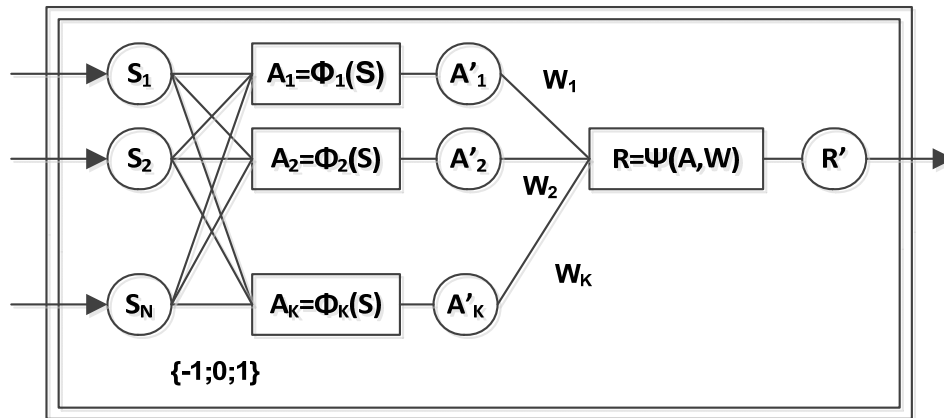


Рис. 1. Описание персептрона с использованием метаграфового подхода.

В соответствии с моделью Ф. Розенблатта [4] классический персептрон состоит из S , A и R элементов.

Слой сенсоров (S) представляет собой набор входных сигналов. Ассоциативный слой (A) включает набор промежуточных элементов, которые активизируются, если одновременно активизируется некоторый набор (образ) входных сигналов. Сумматор (R) активизируется, если одновременно активизируется некоторый набор A -элементов.

В соответствии с обозначениями, принятыми в [5], значение сигнала на A -элементе может быть представлено в виде предиката $\varphi(S)$, а значение сигнала на сумматоре в виде предиката $\psi(A, W)$. Под предикатом в [5] понимается функция, принимающая только два значения «0» и «1».

В зависимости от конкретного вида персептрона вид предикатов $\varphi(S)$ и $\psi(A, W)$ может быть различным. Как правило, с помощью предиката $\varphi(S)$ проверяется, что суммарный входной сигнал от сенсоров не превышает некоторый порог. Также с помощью предиката $\psi(A, W)$ проверяется, что взвешенная сумма от A -элементов не превышает некоторого порога, где W – вектор весов.

В нашем случае конкретный вид предикатов не важен, важно то, что $\varphi(S)$ и $\psi(A, W)$ являются обычными функциональными зависимостями, которые на программном уровне могут быть представлены в виде абстрактного синтаксического дерева.

С точки зрения метаграфов, описание персептрона представляет собой метавершину (показана на рис. 1 в виде двойного прямоугольника). Входные сигналы S , промежуточные сигналы A' и выходной сигнал R' (пассивные данные) показаны на рис. 1 в виде окружностей. Агенты-функции, соответствующие A -элементам и R -элементам, показаны на рис. 1 в виде прямоугольников.

Представим рассмотренную структуру персептрона в виде комбинации агентов-функций, что показано на рис. 2.

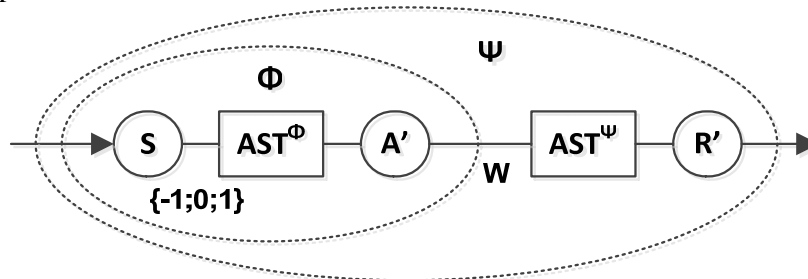


Рис. 2. Представление персептрона в виде агентов-функций.

Предикаты $\varphi(S)$ и $\psi(A, W)$ представлены в виде метавершин (показаны на рис. 2 в виде пунктирных овалов). Определим агент-функцию следующим образом:

$$ag^F = \langle MG_{IN}, MG_{OUT}, AST \rangle, \quad (1)$$

где ag^F – агент-функция; MG_{IN} , MG_{OUT} – метаграфы, выполняющие роль входного и выходного параметра агента-функции соответственно; AST – абстрактное синтаксическое дерево агента-функции, которое может быть представлено в виде метаграфа.

На основе рис. 2 можно описать персептрон в виде комбинации агентов-функций, каждая из которых соответствует виду (1):

$$\varphi^F = \langle S, A', AST^\phi \rangle, \psi^F = \langle \langle \{ \varphi^F \}, W \rangle, R', AST^\psi \rangle.$$

А-элемент может быть представлен в виде агента-функции φ^F . Входным параметром является вектор значений S , выходным параметром вектор значений A' .

Описание персептрона эквивалентно описанию агента-функции ψ^F . Входным параметром является метаграфовое представление кортежа, содержащего описание А-элементов в виде агентов-функций φ^F и вектора W . Выходным параметром является значение выходного сигнала R' .

Описание функций может содержать и другие параметры, например, пороги, но мы предполагаем, что эти параметры входят в описание абстрактного синтаксического дерева.

Для динамических действий с нейронной сетью используется метаграфовый агент, который мы определим следующим образом:

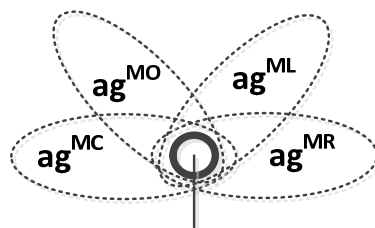
$$ag^M = \langle MG_D, RL, AG^{ST} \rangle, RL = \{rl_j\}, r_i : MG_j \rightarrow OP^{MG},$$

где ag^M – метаграфовый агент; MG_D – метаграф данных, на основе которого выполняются правила агента; RL – набор правил (множество правил rl_j); AG^{ST} – стартовое условие выполнения агента (фрагмент метаграфа, который используется для стартовой проверки правил, или стартовое правило); MG_j – фрагмент метаграфа, на основе которого выполняется правило; OP^{MG} – множество операций, выполняемых над метаграфом.

Метаграфовый агент работает на основе продукционных правил. Антецедентом правила является фрагмент метаграфа, консеквентом правила является множество операций, выполняемых над метаграфом.

Для метаграфового агента структура персептрона представляется в виде метаграфа данных. В нашем случае используются следующие агенты:

1. ag^{MC} – агент создания нейросети;
2. ag^{MO} – агент изменения нейросети;
3. ag^{ML} – агент обучения нейросети;
4. ag^{MR} – агент запуска нейросети.



Представление нейронной сети в виде метаграфа

Рис. 3. Структура метаграфовых агентов для действий с нейронной сетью.

На рис. 3 агенты показаны в виде метавершин с помощью пунктирных овалов.

Агент создания нейросети (ag^{MC}) реализует правила создания начальной топологии нейросети. Данный агент содержит как правила создания отдельных нейронов, так и правила соединения нейронов в нейросеть. Данный агент, в частности, создает абстрактные синтаксические деревья агентов-функций ϕ^F и ψ^F .

Агент изменения нейросети (ag^{MO}) содержит правила изменения топологии сети в процессе работы. Это особенно важно для сетей с изменяемой топологией таких, как SOINN.

Агент обучения нейросети (ag^{ML}) реализует один из алгоритмов обучения. При этом в результате обучения измененные значения весов записываются в метаграфовое представление нейросети. Возможна реализация нескольких алгоритмов обучения с использованием различных наборов правил для агента ag^{ML} .

Агент запуска нейросети (ag^{MR}) реализует запуск и работу обученной нейросети в штатном режиме.

Отметим, что агенты могут работать как независимо, так и совместно. При обучении сети SOINN агент ag^{ML} может вызывать правила агента ag^{MO} для изменения топологии в процессе обучения.

Каждый агент на основе заложенных в него правил фактически реализует специфическую программную «машину». Использование метаграфового подхода позволяет реализовать принцип «мультимашинности», когда несколько агентов с различными целями реализуют различные действия на одной и той же структуре данных.

Таким образом, с помощью метаграфового подхода реализуется описание нейронной сети, а также создание, изменение, обучение и штатная работа нейронной сети.

ЛИТЕРАТУРА

1. Черненький В.М., Терехов В.И., Гапанюк Ю.Е. Представление сложных сетей на основе метаграфов // Нейроинформатика-2016. XVIII Всероссийская научно-техническая конференция. Сборник научных трудов. Ч. 1. М.: НИЯУ МИФИ, 2016.
2. Самохвалов Э.Н., Ревунков Г.И., Гапанюк Ю.Е. Использование метаграфов для описания семантики и прагматики информационных систем. Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2015. Выпуск №1. С. 83-99.
3. Черненький В.М., Терехов В.И., Гапанюк Ю.Е. Структура гибридной интеллектуальной информационной системы на основе метаграфов. Нейрокомпьютеры: разработка, применение. 2016. Выпуск №9. С. 3-14.
4. Розенблатт Ф. Принципы нейродинамики. М.: Мир, 1965. — 480 с.
5. Минский М., Пейперт С. Перцептроны. М.: Мир, 1971. — 262 с.