

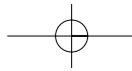
Перестройка модели мира, формируемой на материале анализа текстовой информации с использованием искусственных нейронных сетей, в условиях динамики внешней среды

А.А. Харламов,
кандидат технических наук

В.В. Раевский,
доктор биологических наук

Правополушарная образная модель мира у человека как ассоциативная семантическая сеть имеет левополушарный языковой эквивалент (также в виде сети): многообразным образом объектов мира в их взаимосвязях сопоставлены понятия (слова, словосочетания) с их соответствующими связями. Поэтому операции на метаязыковой семантической сети можно считать отражением операций на образной сети. В связи с этим, под классификацией входной ситуации можно понимать отнесение семантической сети ситуации как отражения последней в модели мира к образной (а значит, и языковой) семантической сети одной из предметных областей. Изменения внешней среды приводят к нарушению адекватности между входными ситуациями и сформированной ранее моделью мира. Восстановление адекватности требует замены устаревающих семантических сетей предметных областей модели мира новыми их версиями. Всё вышеперечисленное можно промоделировать на манипуляциях с текстовой информацией с формированием ассоциативных семантических сетей, отражающих структуру текстов. Возможна классификация текстов путём отнесения их к соответствующим текстовым выборкам предметных областей. Корректность такой классификации зависит от соответствия текстовых выборок





Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

предметных областей реальному положению дел в предметной области в текущий момент. Корректная классификация достигается своевременной заменой ассоциативных сетей — моделей предметных областей — по мере изменения информации, составляющей предметные области.

Abstract

The dextrocerebral image world model in human beings, as an associative semantic network, has a sinistrocerebral language equivalent (also in the form of a network): notions (words, word combinations) with their respective links are confronted with multimodal images of world objects in their correlation. That is why operations on the metalinguistic semantic network can be considered as a reflection of operations on the image network. In this regard, an entry situation can be classified as such ascribing the semantic network of situation as its mapping into the world model to the image (which means linguistic as well) semantic network of one of object domains. Changes in the external environment lead to the disturbance of adequacy between entry situations and the previously formed world model. Restoration of the adequacy requires that outdated semantic networks of object domains of the world model be replaced by their new versions. All the above-mentioned can be modeled on manipulations with textual information, with associative semantic networks mapping the text structure being formed.

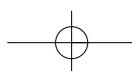
Введение

В привычных ситуациях человек оперирует образной правополушарной моделью мира [1]. Образная модель мира может быть представлена иерархией семантических сетей, в которой элемент заданного уровня представляет собой события более низкого уровня сложности, объединённые в рамках этого элемента в единое целое. Логика связей этих элементов отражает структуру моделируемого объекта мира.

Структуре образной модели мира соответствует структура описывающей её языковой модели мира, где образам соответствуют понятия (слова и словосочетания) в их взаимосвязях, а попарная сочетаемость понятий определяет семантику предметной области [2].

Поскольку текст описывает некоторую действительность, в результате анализа текстовой информации можно построить ассоциативную сеть, похожую на семантическую сеть — языковую модель мира человека [3]. Сформированная таким образом семантическая сеть содержит понятия, ранжированные их семантическими весами, в их взаимосвязях, также с весами. Существует класс искусственных нейронных сетей, позволяющий автоматически сформировать семантические представления, подобные тем, которые формируются человеком. Такой нейронной сетью является сеть на основе нейроподобных элементов с временной суммацией сигналов [4, 3].

Нейроподобный элемент с временной суммацией сигналов имеет цепочку задержек на входе. Его архитектура имеет вид, изображённый на [рисунке 1](#).



Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

Нейроподобный элемент с временной суммацией сигналов является последовательным развитием нейроподобного элемента А.Н. Радченко [4], возникшего на основе модели W. Rall [5], которая, в свою очередь, возникла на основе представлений D.A. Sholl [6].

В эквивалентной форме его можно представить в виде, изображённом на рис. 2.

Такой нейрон выполняет свёртку фрагмента бинарной последовательности длины n : $(x_{t-n+1}, x_{t-n+2}, \dots, x_t)$, $x_i \in \{0, 1\}$ — с последовательностью весовых коэффициентов (w_1, w_2, \dots, w_n) , $w_i \in \{-1, 1\}$:

$$S = \sum_{i=1}^n x_{t-n+i} w_i \quad (1)$$

Свёртка будет иметь наибольшее значение, если n -членный фрагмент входной последовательности соответствует последовательности весовых коэффициентов нейрона, то есть: если $w_i = -1$, то $x_i = 0$, а если $w_i = +1$, $x_i = 1$. Этот фрагмент последовательности называется адресом нейрона. Наибольшее значение свёртки равно числу единиц в адресе — $\Sigma_{ед}$.

В качестве нелинейной функции используется пороговое преобразование $f(*) = H_{адп}$ с порогом $h_{адп}$. Если порог преобразования $h_{адп}$ равен числу единиц в адресе — $\Sigma_{ед}$, то нейрон будет откликаться строго на свой адрес. То есть он моделирует одну из точек n -мерного пространства R^n .

В случае бинарной входной последовательности это вершина n -мерного единичного гиперкуба G_e . Нейроны с разными распределениями синапсов на дендритах можно избирательно возбуждать, подавая на них различные фрагменты бинарных последовательностей. При этом пороги возбуждения нейронов, которым эти дендриты принадлежат, должны быть равны числу возбуждающих синапсов на своих обобщённых дендритах (такова может быть максимальная накопленная в каждом дендрите сумма, если весовые коэффициенты всех синапсов равны +1 или -1).

Объединение множества из 2^n нейронов с разными адресными комбинациями в единую структуру порождает модель n -мерного пространства R^n (точнее, единичного гиперкуба $G_e \in R^n$, если веса синапсов равны +1 и -1). В этом случае каждый отдельный нейрон моделирует одну из вершин гиперкуба. Такая структура позволяет отобразить любую бинарную последовательность A в последовательность вершин G_e — траекторию.

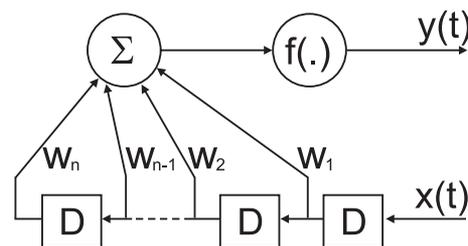


Рис. 1. Нейроподобный элемент с цепочкой задержек на входе

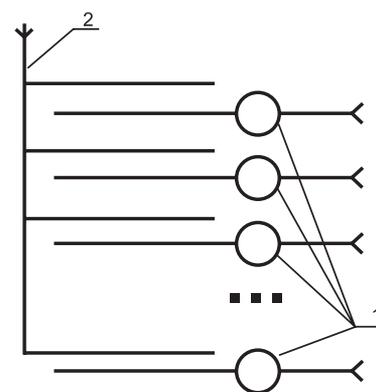


Рис. 2. Обычный нейроподобный элемент с пространственной суммацией сигналов, входы которого объединяются бинарным сдвиговым регистром

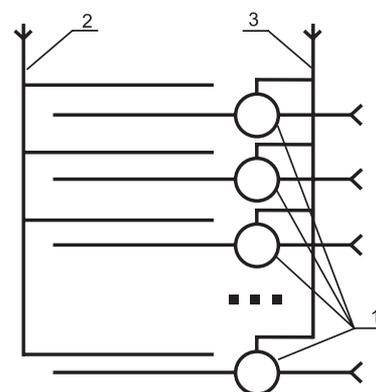


Рис. 3. Нейронный пучок. Здесь 1 — нейроны пучка, имеющие обобщённые дендриты с разными адресами от (000...0) до (111...1), 2 — общее афферентное волокно, 3 — управляющий вход

Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

В общем случае среди n -членных фрагментов информационной последовательности может встретиться уже ранее встречавшийся n -членный фрагмент, и траектория в этом случае пройдет через вершину, уже принадлежащую ей, то есть пересечётся с самой собой. В этой точке возможно более одного продолжения траектории. Ассоциативность преобразования F позволяет сохранить топологию структуры преобразуемой информации. Действительно, одинаковые фрагменты последовательности преобразуются в одну и ту же траекторию, разные — в разные траектории.

Введение в вершинах гиперкуба G_e функции памяти M ставит в соответствие каждой вершине $\hat{a}(t) \in \hat{A}$, соответствующей t -му символу последовательности A , двоичную переменную $a(t+1)$, являющуюся $(t+1)$ -м символом этой последовательности. При усложнении функции памяти M реализуется механизм, фиксирующий число проходов траекторией заданной точки в заданном направлении. Такой механизм позволяет характеризовать каждую точку траектории с точки зрения частоты появления во входной информации сочетания $(\hat{a}(t), a(t+1))$. Применение порогового преобразования позволяет выявлять траектории, точки которых превышают порог по частоте прохождения. Такой механизм позволяет анализировать входную последовательность с точки зрения появления повторяющихся частей, так как одинаковые фрагменты последовательности отображаются преобразованием F в одну и ту же часть траектории.

Если имеется класс последовательностей $\{A\}$, в которых в разных комбинациях встречаются подпоследовательности $\{B_i\}$, то отображение последовательности класса $\{A\}$ в n -мерное пространство и применение к ним порогового преобразования приводит к формированию множества траекторий $\{\hat{B}_i\}$, соответствующего множеству подпоследовательностей $\{B_i\}$, — словаря. Можно сказать, что преобразование $HM^{-1}MF$ при взаимодействии с входным классом $\{A\}$ формирует словарь, характеризующий траектории, соответствующие последовательностям входного класса в пространстве данной мерности:

$$\{\hat{B}\} = HM^{-1}MF(\{A\}) . \quad (4)$$

Сформированный словарь может быть использован для детектирования старой информации в потоке новой. Для этого необходимо реализовать поглощение фрагментов входной последовательности A , соответствующих словам словаря, и пропускание новой, относительно словаря, информации. В результате появляется предпосылка для реализации структурного подхода к обработке информации: имеются механизмы для выявления как событий, так и их связей. В целях решения задачи детектирования преобразование F^{-1} модифицируется для придания ему детектирующих свойств. Преобразование F_c^{-1} взаимодействует с входной последовательностью \hat{A} , которая содержит, наряду со старой, некоторую новую информацию. Если на основании множества входных последовательностей A ранее был сформирован словарь $\{\hat{B}\} = HM^{-1}MF(\{A\})$, то использование преобразования F_c^{-1} позволяет сформировать так называемую синтаксическую последовательность, или последовательность аббревиатур, — C , характеризующую связи слов B словаря $\{B\}$ в последовательности A . Здесь $\{B\}$ есть множество подпоследовательностей, соответствующих всем цепям слов \hat{B} словаря $\{\hat{B}\}$. В результате взаимодействия происходит формирование последовательности C , в которой заменяются нулями те части последовательности \hat{A} , соответствующие которым части траектории $\hat{A} = F(\tilde{A})$ совпадают с частями траектории \hat{A} :

$$C = F_c^{-1}(F(\tilde{A}), HM^{-1}(\{\hat{A}\})). \quad (5)$$

Отображение F^{-1} устраняет из входной последовательности \tilde{A} некоторую информацию, содержащуюся в словаре $\{\hat{B}\}$. Тем самым создаётся предпосылка построения иерархической структуры для лингвистической (структурной) обработки входной информации. Синтаксическая

последовательность C , содержащая только новую, по отношению к данному уровню, информацию, становится входной для следующего уровня. На следующем уровне, подобно описанному выше, из множества синтаксических последовательностей $\{C\}$ формируется словарь $\{\hat{D}\}$ и множество синтаксических последовательностей следующего уровня $\{E\}$. Мы имеем стандартный элемент многоуровневой иерархической структуры: обработка с выделением поуровневых словарей может происходить на всех уровнях. Словарь следующего уровня является в этом случае грамматикой для предыдущего уровня, так как его элементами, при соответствующем выборе размерностей пространств этих уровней, являются элементы связей слов предыдущего уровня.

Из таких модулей можно сформировать иерархическую структуру для обработки информации заданной модальности, где на каждом уровне формируется ряд параллельно включённых процессов обработки информации, описанных выше, связанных с процессами следующего уровня по типу «каждый с каждым». Внешние события отображаются в иерархическую структуру данной модальности и формируют в ней иерархию словарей, слова которых оказываются связанными между собой синтаксическими последовательностями. Эта иерархия словарей, где слова словарей более высоких уровней содержат информацию о связях слов более низких уровней, является моделью мира в терминах данной модальности. Модель мира заданной модальности m представляет собой многократно вложенный суперграф \hat{M}_m . На верхнем уровне K он объединяет все графы-слова \hat{B}_{ijk} словаря (подсловарей) верхнего уровня, а на всех более низких уровнях включает в себя (в соответствующие места — по ассоциативному принципу) графы-слова \hat{B}_{ijk} словарей нижних уровней:

$$\hat{M}_m = \bigcup_{ij} \bigvee_{ijk < K} \hat{B}_{jlk} \quad (6)$$

Здесь U — операция объединения, V обозначает включение на своё место слова словаря более высокого уровня. Это включение аналогично логическому сложению фрагмента последовательности, соответствующего слову словаря нижнего уровня, с синтаксической последовательностью (вложению слова словаря в соответствующую купюру синтаксической последовательности).

Объединение несколько иерархических структур, соответствующих разным модальностям, со сформированными на них модальными моделями мира \hat{M}_m позволяет сформировать многомодальный объединённый суперграф — семантическую сеть:

$$\hat{M} = \bigcup_m \hat{M}_m = \bigcup_{ijm} \hat{B}_{jlk} \bigvee_{ijk < Km} \hat{B}_{jlk} \quad (7)$$

В иерархической структуре, реализующей структурную обработку информации, может сосуществовать одновременно две системы знаний о мире: статическая (парадигматическая), потенциально хранящая всю доступную информацию об элементах и связях структур событий мира, — она описана выше в виде ассоциативной семантической сети, — и динамическая (синтагматическая), в той или иной степени соответствующая конкретным ситуациям, отображаемым на иерархическую структуру. Вторая имеет общие черты с фреймовыми представлениями.

Динамическая компонента формируется отображением текущей входной ситуации на семантическую сеть. События мира в той или иной комбинации связываются в ситуации, которые отображаются на модели мира в виде цепочек слов словарей

Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

разных уровней, разворачивающихся во времени. В каждый момент времени в каждом графе-слове данной модальности может высветиться только одна цепь \hat{A} . Эти цепи на вложенном суперграфе (семантической сети) составляют динамическое представление (назовём их компонентами фрейма \hat{f}_m данной модальности), соответствующее некоторой текущей входной ситуации (синтагме):

$$\hat{f}_m = \hat{A}_{ijKm} \bigvee_{ijk < K} \hat{A}_{ijkm} \quad (8)$$

Компоненты фрейма могут включать в себя конкатенации цепей \hat{A}_i (также по ассоциативному принципу). Здесь $[*]$ означает конкатенацию по индексу i . Каждой конкретной ситуации на семантической сети \hat{M} соответствует некоторое объединение компонент фрейма отдельных модальностей — субфрейм:

$$\hat{S} = \bigcup_m \hat{f}_m = \bigcup_m \left(\hat{A}_i * \hat{A}_{ijKm} \bigvee_{ijk < K} \hat{A}_{ijkm} = \hat{f}_1 \bigcup_{m \neq 1} \hat{f}_m \right), \quad (9)$$

где \hat{f}_1 — имя субфрейма — слово, например, в первой (вербальной) модальности. Последовательность субфреймов формирует фрейм:

$$\hat{F} = \hat{A}_1 * \hat{S}_p = \hat{f}_1 * (\hat{A}_p * \hat{S}_p) = \hat{I} * (\hat{A}_p * \hat{S}_p),$$

где \hat{I} — имя фрейма (так же, как в случае имени субфрейма) — слово, например (но не обязательно), в вербальной модальности.

1. Ассоциативная семантическая сеть

Формирование семантических представлений может быть продемонстрировано на примере формирования металингвистической ассоциативной семантической сети некоторого текста или множества текстов, представляющих собой определённую предметную область. Такая семантическая сеть строится с учётом сочетаемости слов. Имеется в виду однородная (ассоциативная) семантическая сеть N , учитывающая только ассоциативные отношения [7]. Такая семантическая сеть представляется как множество понятий (слов и словосочетаний), объединённых связями.

Для формирования семантической сети используется иерархия процессов, в которых текстовая информация последовательно обрабатывается на морфологическом, лексическом, синтаксическом и семантическом уровнях с формированием поуровневых словарей.

Формирование поуровневых словарей на основе текстовой информации может происходить следующим образом. На морфологическом уровне формируется несколько словарей, но в первую очередь нас интересует словарь флективных морфем. На лексическом уровне представлены, в том числе, словарь основ слов (префикс + корневая морфема + аффикс) и словарь устойчивых сочетаний корневых основ. На синтаксическом уровне формируется словарь синтаксем, представляющих собой флективную структуру синтаксических узлов с выколотыми корневыми основами.

Наконец, формируется словарь семантического уровня. Этот словарь формируется фильтрацией текстов входной выборки через словарь синтаксем. Он представляет собой словарь попарной встречаемости корневых основ в текстах. Полученные при этом цепочки корневых основ



формируют граф семантической сети. При достаточно большой текстовой выборке представление в многомерном пространстве разбивается на пары корневых основ (назовём это словарём семем). Пары корневых основ могут быть объединены в сеть. В дальнейшем мы воспользуемся именно таким представлением семантического уровня. Семантика, таким образом, характеризуется попарной сочетаемостью корневых основ [2].

Прагматика может быть рассмотрена как надстройка над семантическим уровнем. При условии ранее сформированной ассоциативной семантической сети прагматическая информация характеризуется последовательностями корневых основ понятий семантической сети, полученными в результате отображения на семантическую сеть конкретной ситуации (текста). Объединение таких цепочек в единое представление, являющееся сценарием (алгоритмом, фреймом) этой ситуации в терминах понятий семантической сети, и формирует прагматический уровень знаний.

Таким образом, далее рассматриваются ассоциативные семантические сети, в которых связи между понятиями описываются только отношением ассоциативности. Рассматривается их подкласс, где связи направлены только в одну сторону. Такая ассоциативная сеть образует метрическое пространство, на котором определены логические операции вложения и пересечения, позволяющие вычислить степень совпадения двух сетей.

По аналогии с семантическим уровнем представления текстовой информации в виде ассоциативной сети ключевых понятий с их взаимосвязями можно говорить о таком же по виду семантическом уровне представления многомодальной информации в многомодальной модели мира: в виде ассоциативной сети ключевых событий в их взаимосвязях.

А прагматический уровень многомодальных представлений является графом-сценарием (фреймом, алгоритмом) на семантической сети многомодальной модели мира.

2. Пересечение семантических сетей

Опишем более формально ассоциативную семантическую сеть N , а также рассмотрим её как некое подмножество метрического пространства.

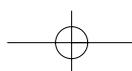
Определение 1. Под семантической сетью N понимается совокупность несимметричных пар понятий $\langle\langle c_i c_j \rangle\rangle$:

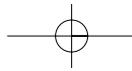
$$N \cong \langle\langle c_i c_j \rangle\rangle \quad (10)$$

Выполнение условия несимметричности предполагает неравенство двух пар: $\langle c_i c_j \rangle \neq \langle c_j c_i \rangle$.

Определение 2. Семантическая сеть, описанная таким образом, может быть переопределена как множество так называемых звёздочек $\langle\langle c_i \langle c_j \rangle \rangle\rangle$:

$$N \cong \langle\langle c_i \langle c_j \rangle \rangle\rangle \quad (11)$$





Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

Определение 3. Под звёздочкой $\langle c_i \langle c_j \rangle \rangle$ понимается конструкция, включающая главное понятие c_i , связанное с множеством понятий-ассоциантов $\langle c_j \rangle$, которые являются семантически признаками главного понятия, отстоящими от главного понятия на одну связь. Связи направлены от главного понятия к понятиям-ассоциантам.

В одной из звёздочек семантической сети главным понятием может быть одно из понятий-ассоциантов другой звёздочки, и наоборот. Но связи всегда направлены от главного понятия к понятию-ассоцианту.

Определение 4. Звёздочка с единичными значениями весов понятий-ассоциантов называется единичной звёздочкой (звёздочкой-ортом).

Определение 5. Звёздочкой-подпространством называется звёздочка, полученная на единичной звёздочке введением семантических весов понятий (w_i, w_j) :

$$S \cong \langle \langle w_i c_i \langle w_j c_j \rangle \rangle \rangle \quad (12)$$

Семантическая сеть в терминах этих определений представляет собой декартово произведение подпространств, порождаемых всеми звёздочками, входящими в семантическую сеть, полученными на единичных звёздочках за счёт введения весовых характеристик понятий-ассоциантов:

$$N = S_1 \times S_2 \times \dots \times S_i \quad (13)$$

Для введения в метрическом пространстве, определённом на семантической сети, операций вложения и пересечения разобьём звёздочку на пары понятий $\langle c_i c_j \rangle$. Здесь c_i — главное понятие звёздочки, а $c_j \in \langle c_j \rangle$ — множество понятий-ассоциантов понятия c_i . Связь направлена от c_i к c_j .

Определение 6. Введём векторное произведение на векторах $\bar{c}_i \bar{c}_j$, где угол w_{ij} между векторами понятий c_i, c_j пропорционален весу связи от c_i к c_j : $w_{ij} \in (0 \dots 90^\circ)$. Площадь треугольника S_j , построенного на векторах $\bar{c}_i \bar{c}_j$, развёрнутых на угол w_{ij} относительно друг друга, будет использована для вычисления степени пересечения сначала звёздочек, а потом семантических сетей как совокупностей звёздочек.

Определение 7. Под пересечением двух звёздочек понимается сумма (по всем понятиям-ассоциантам данного главного понятия звёздочки) пересечений площадей двух треугольников, построенных в плоскости векторов $\bar{c}_i \bar{c}_j$, один из которых построен на векторах, развёрнутых на угол, пропорциональный связи $(w_{ij})_1$ между понятиями в одной звёздочке, а другой — на векторах, развёрнутых на угол, пропорциональный связи $(w_{ij})_2$ между теми же понятиями в другой, сравниваемой с первой, звёздочке. В случае если в одной из звёздочек пары, для которой считается пересечение, не нашлось соответствующего понятия-ассоцианта, пересечение считается равным 0.

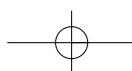
$$S_{12} = \langle c_{i_1} \langle c_{j_1} \rangle \rangle \cap \langle c_{i_2} \langle c_{j_2} \rangle \rangle = \sum_{j=1}^{\max(N_1, N_2)} S_{j_1} \cap S_{j_2} \quad (14)$$

N_1, N_2 — число ассоциантов, соответственно, в звёздочках 1 и 2.

Определение 8. Под пересечением семантических сетей понимается сумма пересечений звёздочек, включённых в эти сети (считая по главным понятиям):

$$S_{12} = \langle c_{i_1} c_{j_1} \rangle \cap \langle c_{i_2} c_{j_2} \rangle = \sum_{k=1}^{\max(M_1, M_2)} (S_{k_1} \cap S_{k_2}) \quad (15)$$

M_1, M_2 — число звёздочек, входящих, соответственно, в семантические сети 1 и 2.



3. Модель мира. Языковой и образный компоненты. Классификация

Правополушарная образная модель мира у человека $\hat{M} = \bigcup_m \hat{M}_m$ как объединение модальных моделей мира \hat{M}_m представляет собой граф — ассоциативную сеть, вершины которой являются подграфами, описывающими ключевые ситуации (события), которые связаны между собой ассоциативными связями, соответствующими логике структуры мира. Наряду с многомодальной правополушарной моделью мира и, параллельно ей, в левом полушарии формируется изоморфная многомодальной модели металингвистическая модель мира \hat{M}_1 — также ассоциативная сеть ключевых понятий (здесь индекс 1 означает первую — текстовую — среди возможных модальностей), соответствующих ключевым ситуациям многомодальной модели, с их связями, аналогичными связям ситуаций в многомодальной модели. Таким образом, образная модель, представленная в виде семантической сети, имеет левополушарный языковой эквивалент (также в виде сети): многомодальным образам объектов мира в их взаимосвязях там сопоставляются понятия (слова, словосочетания) с их соответствующими связями. Поэтому операции на метаязыковой семантической сети можно считать отражением операций на образной сети (мы можем оценить операции над многомодальной моделью, моделируя их на металингвистической модели).

Определение 9. В связи с этим, под классификацией входной ситуации можно понимать отнесение метаязыковой семантической сети ситуации N к сети N_n (где $n=1\dots N$ — число предметных областей) одной из предметных областей метаязыковой модели мира. В идеальном случае семантическая сеть ситуации вкладывается в сеть соответствующей предметной области.

Используя операцию пересечения сетей, определённую выше, мы можем оценить степень подобия двух сетей и, тем самым, сравнивать по смыслу (по структуре) ситуации (их метаязыковые модели). Имея модели предметных областей в виде ассоциативных семантических сетей, мы можем классифицировать входные ситуации (описывающие их метаязыковые модели) вычислением степени совпадения (вложения) сети входной ситуации и сетей предметных областей, относя входную ситуацию к той предметной области, для сети которой степень совпадения с сетью входной ситуации окажется больше.

Далее мы можем абстрагироваться от входных ситуаций. Будем рассматривать их метаязыковые модели. Пойдём ещё дальше: перейдём к анализу текстов, на основе которых эти метаязыковые модели формируются. Как и в случае классификации входных ситуаций, используя операцию пересечения сетей, мы можем оценить степень подобия двух сетей и, тем самым, сравнивать по смыслу уже тексты. Имея модели предметных областей в виде ассоциативных семантических сетей соответствующих тематических текстовых выборок, мы можем классифицировать входные тексты вычислением степени совпадения (вложения) сети входного текста и сетей предметных областей, относя входной текст к той предметной области, для сети которой степень совпадения с сетью входного текста окажется больше.

Динамика внешней среды приводит к нарушению адекватности между входными ситуациями и сформированной ранее моделью мира. И, как следствие, к необходимости замены устаревающих семантических сетей предметных областей модели мира их новыми версиями. В противном случае, начиная с некоторого момента, становится невозможной их корректная классификация.

Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

4. Динамика внешней среды и перестройка модели

Кластеризация (рубрикация) информационного поля текстовой выборки. Поскольку мы рассматриваем текст как модель описываемой им ситуации, а множество подмножеств текстов — как множество моделей предметных областей, можно считать, что динамика изменений семантических сетей текстовых выборок предметных областей отражает динамику изменений входных ситуаций. Рассмотрим динамику таких изменений на примере текстовых выборок новостной ленты.

Множество текстов новостной ленты, относящихся к некоторому периоду времени, может быть разбито на подмножества, соответствующие некоторым предметным областям. Для каждой предметной области можно построить ассоциативную семантическую сеть, которая характеризует состав предметной области в терминах её ключевых понятий и их связей.

Изменение во времени семантических сетей предметных областей. Для разных (в частности, последовательных) промежутков времени состав текстов, отнесённых к разным предметным областям, будет различным. Следовательно, будут различными соответствующие им семантические сети. Будут различаться словари ключевых понятий и их связи. На разных промежутках времени в текстах конкретных предметных областей новостной ленты будут исчезать одни ключевые понятия, появляться другие, может меняться их семантический вес. То же можно сказать и об их связях.

Классификация входных событий. Как описывалось выше, можно классифицировать входной текст новостной ленты отнесением его к одной из предметных областей. Степень пересечения семантических сетей текста и предметных областей предоставляет такую возможность. Текст относится к той из предметных областей, для которой степень пересечения ассоциативной сети с ассоциативной сетью текста наибольшая.

Входной текст классифицируется корректно, если состав ассоциативных сетей предметных областей соответствует текущему положению дел на новостной ленте; в противном случае классификация будет неправильной. Поэтому для правильной классификации входных текстов необходимо контролировать модели предметных областей: перечень ключевых понятий (с их весами) и их связей (также с их весами) ассоциативных сетей предметных областей должен не более чем на заданную величину отличаться от их текущих состояний в текстах новостной ленты. В случае превышения указанного заданного предела несоответствия для правильной классификации входных текстов требуется приведение моделей предметных областей в соответствие с текущей ситуацией путём их перестройки.

Заключение

В статье показано наличие соответствия многомодальной модели мира, формируемой у человека, и параллельно формируемой у него метаязыковой модели мира. Показано, что в силу этого соответствия возможно моделирование процессов классификации событий на многомодальной модели классификацией соответствующих понятий на метаязыковой модели. В том числе на метаязыковых моделях (ассоциативных семантических сетях), полученных на текстовой информации. Представлен способ и механизм классификации текстовой информации с использованием автоматически извлечённых из текста ключевых понятий (с их весами) и их связей (также с их весами) — ассоциативной сети текста — путём вычисления степени пересечения этой сети с аналогичными заранее сформированными сетями — моделями предметных областей (текстовых выборок). Показано, что корректная классификация



Харламов А.А., Раевский В.В. Перестройка модели мира, формируемой на материале анализа текстовой информации...

текста возможна только в случае соответствия текущего состояния моделей предметных областей (их ассоциативных сетей) реальному составу текстового потока в исследуемый момент времени. В противном случае требуется перестройка моделей предметных областей с целью приведения их в соответствие с текущей ситуацией.

Литература

1. Глезер В.Д. Зрение и мышление. Л.: Наука. 1985.
2. Рахилина Е.В. Когнитивный анализ предметных имён: семантика и сочетаемость. М.: Русские словари. 2000.
3. Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). М.: Радиотехника. 2006.
4. Радченко А.Н. Моделирование основных механизмов мозга. Л.: Наука. 1969.
5. Rall W. Electrophysiology of a dendritic neuron model. Biophys. J., 2. (Suppl.), 1962. Pp. 145–167.
6. Sholl D.A. Dendritic organization in the neurons of the visual and motor cortices of the cat. J. Anat., 87, 1953. Pp. 387–406.
7. Осипов Г.С. Приобретение знаний интеллектуальными системами: Основы теории и технологии. М.: Наука, Физматлит, 1997.

Харламов Александр Александрович —

окончил теплоэнергетический факультет МЭИ, в 1970 г. — механико-математический факультет МГУ. С 1996 г. работает старшим научным сотрудником Института высшей нервной деятельности и нейрофизиологии РАН. С 2005 г. — начальником отдела обработки звуковых сигналов ООО «Юникор микросистемы». С 2007 г. — начальником лаборатории электронных образовательных ресурсов Федерального института развития образования. Специалист в области нейроинформатики. Область интересов: распознавание речи, распознавание изображений, анализ текстов, нейронные сети, семантические представления, моделирование информатики мозга.

Раевский Владимир Вячеславович —

доктор биологических наук, профессор. Родился в 1942 г., в 1965 г. окончил 1 Московский медицинский институт им И.М.Сеченова. В настоящее время руководит лабораторией нейроонтогенеза в Институте высшей нервной деятельности и нейрофизиологии РАН. Основные работы посвящены исследованию морфологических основ, нейрофизиологических и нейрохимических механизмов формирования адаптивного поведения на ранних стадиях развития организма.

