

МЕТОД СИНТЕЗА СЛОВ ДЛЯ НЕЙРОСЕТЕВОГО ЛИНГВИСТИЧЕСКОГО ПРОЦЕССОРА

Национальный авиационный университет

В статье предлагается метод синтеза слов для нейросетевого лингвистического процессора, который учитывает грамматику естественного языка, использует количественные оценки естественно-языковых объектов, формализует взаимосвязи между ними и операции над ними, и осуществляет синтез слов на основе модели нейросети

Постановка проблемы. В настоящее время актуальной является проблема повышения эффективности естественно-языкового взаимодействия оператора и технического объекта посредством компьютерной системы, связанной с задачами анализа и синтеза естественно-языковых объектов (ЕЯО).

Анализ исследований и публикаций. Семантическая нейронная сеть [1], основанная на алгебре конечных предикатов [2-3], не учитывает правила грамматики естественного языка [4-5] и морфологические явления [4-5]. Расширенные сети переходов [6] и сети RAAM [7-8] используются только для анализа структуры предложения. Кроме того, семантическая нейронная сеть, расширенные сети переходов и RAAM не используют количественные оценки ЕЯО и операции над ними и связи между ними.

Цель и постановка задачи. Целью настоящей работы является формирование методических положений синтеза слов для нейросетевого лингвистического процессора. Для достижения поставленной цели решаются следующие задачи:

- определение вектора признаков ЕЯО и его взвешенной нормы;
- формализация взаимосвязей между ЕЯО;
- формализация операций над ЕЯО;
- формирование правил нелинейных морфологических преобразований;
- разработка структура синтеза слов;

- разработка метода построения модели синтеза слов;

- формирование логико-формальных правил оценивания результата синтеза слов.

1. Определение вектора признаков естественно-языкового объекта и его взвешенной нормы

В статье производится формализация следующих ЕЯО ($s=1$ – если буква, $s=2$ – если фонема):

- частей речи $H_i, i \in \overline{1, \eta(\mathbf{H})}$;
- букв $A_i, i \in \overline{1, \eta(\mathbf{A})}$ и фонем $\Phi_i, i \in \overline{1, \eta(\Phi)}$;
- порождающих слов $\overline{Q}_e^s, e \in \overline{1, \eta(\overline{Q}^s)}$;
- основ слов $\overline{L}_q^s, q \in \overline{1, \eta(\overline{L}^s)}$;
- форм основ слов $\overline{\tilde{L}}_q^s, q \in \overline{1, \eta(\overline{\tilde{L}}^s)}$, полученных при линейных и нелинейных морфологических преобразованиях;
- словообразовательных аффиксов (префиксов \overline{B}_p^{s1} , интерфиксов \overline{B}_p^{s2} , суффиксов \overline{B}_p^{s3} , флексий \overline{B}_p^{s4} , постфиксов $\overline{B}_p^{s5}, p \in \overline{1, \eta(\overline{B}^{sk})}, k \in \overline{1, 5}$) и их наборов $\overline{B1}_p^s, p \in \overline{1, \eta(\overline{B1}^s)}$;
- неосновных частей слов $\overline{V}_u^s, u \in \overline{1, \eta(\overline{V}^s)}$;

– $\overline{D_w / \overline{D_w}^{s1} \overline{D_w}^{s2}}$, $w \in 1, \eta(\overline{\mathbf{D}}^{sk})$, $k \in \overline{1, 2}$ и чередующихся $\overline{D_v / \overline{D_v}^{s1} \overline{D_v}^{s2}}$, $v \in 1, \eta(\overline{\mathbf{D}}^{sk})$, $k \in \overline{1, 2}$ буквенных и фонемных последовательностей и их наборов $\overline{D1_w^s}$, $w \in 1, \eta(\overline{\mathbf{D1}}^s)$, $\overline{D1_v^s}$, $v \in 1, \eta(\overline{\mathbf{D1}}^s)$, участвующих в линейных и нелинейных морфонологических преобразованиях основ слов;

– порожденных слов $\overline{C_r^s}$, $r \in 1, \eta(\overline{\mathbf{C}}^s)$.

Частям речи H_i , буквам A_i и фонемам Φ_i присваиваются количественные оценки в виде рангов $r(H_i)$, $r(A_i)$, $r(\Phi_i)$ соответственно.

На основе рангов букв $r(A_i)$ и рангов фонем $r(\Phi_i)$ формируются вектора признаков ЕЯО $\mu(\overline{Q_e^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{V_u^s})$, $\mu(\overline{B_p^{sk}})$, $\mu(\overline{D_w^{sk}})$, $\mu(\overline{D_v^{sk}})$, $\mu(\overline{C_r^s})$ в виде

$$\mu(\overline{\xi_d^s}) = (r(\xi_{d1}^s), \dots, r(\xi_{dT}^s)), \quad (1)$$

где ξ_{dt}^s – буква или фонема,

T – максимальная длина ЕЯО.

Каждому вектору признаков ЕЯО $\mu(\overline{Q_e^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{L_q^s})$, $\mu(\overline{V_u^s})$, $\mu(\overline{B_p^{sk}})$, $\mu(\overline{D_w^{sk}})$, $\mu(\overline{D_v^{sk}})$, $\mu(\overline{C_r^s})$ ставится в соответствие его взвешенная норма $M(\overline{Q_e^s})$, $M(\overline{L_q^s})$, $M(\overline{L_q^s})$, $M(\overline{L_q^s})$, $M(\overline{V_u^s})$, $M(\overline{B_p^{sk}})$, $M(\overline{D_w^{sk}})$, $M(\overline{D_v^{sk}})$, $M(\overline{C_r^s})$ в виде

$$M(\overline{\xi_d^s}) = \left\| \mu(\overline{\xi_d^s}) \right\| = \sum_{t=1}^T \left(|r(\xi_{dt}^s)| * 2^{(T-t)n} \right), \quad (2)$$

где n – количество бит, приходящихся на одну букву или фонему ξ_{dt}^s .

На основе взвешенных норм ЕЯО $M(\overline{B_p^{sk}})$, $M(\overline{D_w^{sk}})$, $M(\overline{D_v^{sk}})$ формируются вектора признаков наборов ЕЯО $\mu(\overline{B1_p^s})$, $\mu(\overline{D1_w^s})$, $\mu(\overline{D1_v^s})$ в виде

$$\mu(\overline{\xi1_d^s}) = (M(\overline{\xi1_{d1}^s}), \dots, M(\overline{\xi1_{dT}^s})), \quad (3)$$

где $\overline{\xi1_{dt}^s}$ – ЕЯО, состоящий из набора букв или фонем.

Каждому вектору признаков набора ЕЯО $\mu(\overline{B1_p^s})$, $\mu(\overline{D1_w^s})$, $\mu(\overline{D1_v^s})$ ставится в соответствие его взвешенная норма $M(\overline{B1_p^s})$, $M(\overline{D1_w^s})$, $M(\overline{D1_v^s})$ в виде

$$M(\overline{\xi1_d^s}) = \left\| \mu(\overline{\xi1_d^s}) \right\| = \sum_{t=1}^T \left(|M(\overline{\xi1_{dt}^s})| * 2^{(T-t)n} \right) \quad (4)$$

2. Закрепление связей между естественно-языковыми объектами

Для закрепления взаимосвязей между ЕЯО вводятся:

– матрица бинарных отношений $\Gamma(\overline{\mathbf{Q}}^s, \overline{\mathbf{C}}^s)$

$$\Gamma(\overline{\mathbf{Q}}^s, \overline{\mathbf{C}}^s) = \left\| \Gamma(\overline{Q_e^s}, \overline{C_r^s}) \right\|, \quad (5)$$

$$M(\overline{Q_{e-1}^s}) < M(\overline{Q_e^s}) < M(\overline{Q_{e+1}^s}),$$

$$M(\overline{C_{r-1}^s}) < M(\overline{C_r^s}) < M(\overline{C_{r+1}^s}),$$

$$\forall \overline{Q_e^s} \wedge \overline{C_r^s} \exists \Gamma(\overline{Q_e^s}, \overline{C_r^s}) = \begin{cases} 1, & \overline{Q_e^s} \rightarrow \overline{C_r^s}; \\ 0, & \text{иначе} \end{cases}$$

3. Операции над естественно-языковыми объектами

Над ЕЯО $\overline{x} = (x_1, \dots, x_m)$, $\overline{y} = (y_1, \dots, y_m)$, $\overline{z} = (z_1, \dots, z_m)$, которые заданы количественными оценками $M(\overline{x})$, $M(\overline{y})$, $M(\overline{z})$, определены операции конкатенации (6), усечение слева (7), усечение справа (8), где n – количество бит на один символ, $|\overline{x}|$ – длина \overline{x} , $\overline{x} \subset \overline{z}$ – вхождение \overline{x} слева, $\overline{y} \subset \overline{z}$ – вхождение \overline{y} справа.

$$M(\bar{z}) = M(\bar{x}) \diamond M(\bar{y}) = \begin{cases} M(\bar{x}) * 2^{|\bar{y}|n} + M(\bar{y}), & |\bar{x}| + |\bar{y}| \leq m, \\ 0, & \text{иначе} \end{cases}, \quad (6)$$

$$M(\bar{y}) = M(\bar{z}) \bullet M(\bar{x}) = \begin{cases} M(\bar{z}) - M(\bar{x}) * 2^{(|\bar{z}| - |\bar{x}|)n}, & \bar{x} \subset \bar{z}, \\ M(\bar{z}), & \text{иначе} \end{cases}, \quad (7)$$

$$M(\bar{x}) = M(\bar{z}) \circ M(\bar{y}) = \begin{cases} (M(\bar{z}) - M(\bar{y})) / 2^{|\bar{y}|n}, & \bar{y} \subset \bar{z}, \\ M(\bar{z}), & \text{иначе} \end{cases}, \quad (8)$$

4. Правила нелинейных морфонологических преобразований

Используя алгебру конечных предикатов, определены следующие правила нелинейных морфонологических преобразований слов в буквенном и фонемном представлении:

– если в основе $\bar{x} = (x_1, \dots, x_m)$ появляется гласная между согласными, идущими в конце основы (выполняется условие $|\bar{d}_{v1}^s| = 0 \wedge |\bar{d}_{v2}^s| > 0$), используется правило (9) и результат помещается в переменную $\bar{y} = (y_1, \dots, y_m)$

– если в основе $\bar{x} = (x_1, \dots, x_m)$ исчезает гласная между согласными, идущими

в конце основы (выполняется условие $|\bar{d}_{v1}^s| > 0 \wedge |\bar{d}_{v2}^s| = 0$), используется правило (10) и результат помещается в переменную $\bar{y} = (y_1, \dots, y_m)$

– если в основе $\bar{x} = (x_1, \dots, x_m)$ заменяются последовательности букв/фонем, идущие в конце основы (выполняется условие $|\bar{d}_{v1}^s| > 0 \wedge |\bar{d}_{v2}^s| > 0$), используется правило (11) и результат помещается в переменную $\bar{y} = (y_1, \dots, y_m)$

$$\{r(T_i^1)\} \subset \{r(A_i)\}, \{r(T_i^2)\} \subset \{r(\Phi_i)\}$$

– множество согласных;

$$P_1(\bar{x}, \bar{d}_{v1}^s, \bar{d}_{v2}^s, \bar{y}) = (|\bar{d}_{v1}^s| = 0) \wedge (|\bar{d}_{v2}^s| > 0) \wedge (|\bar{x}| = \text{len}) \wedge (r(x_c) \in \{r(T_i^s)\} \wedge r(x_{c+1}) \in \{r(T_i^s)\}) \wedge \left((c+1 = \text{len}) \vee \left(\bigwedge_{k=c+1}^{\text{len}-1} \neg(r(x_k) \in \{r(T_i^s)\} \wedge r(x_{k+1}) \in \{r(T_i^s)\}) \right) \right) \rightarrow \left(\bigwedge_{p=1}^j r(y_p) = r(x_p) \right) \wedge r(y_{c+1}) = r(\tilde{d}_{v21}^s) \wedge \left(\bigwedge_{b=c+2}^{\text{len}} r(y_b) = r(x_{b-1}) \right), \quad c \in \overline{1, |\bar{x}| - 1}, \quad (9)$$

$$P_2(\bar{x}, \bar{d}_{v1}^s, \bar{d}_{v2}^s, \bar{y}) = (|\bar{d}_{v1}^s| > 0) \wedge (|\bar{d}_{v2}^s| = 0) \wedge (|\bar{x}| = \text{len}) \wedge (r(x_{c-1}) \in \{r(T_i^s)\} \wedge r(x_c) = r(\tilde{d}_{v11}^s) \wedge r(x_{c-1}) \in \{r(T_i^s)\}) \wedge \left((c+1 = \text{len}) \vee \left(\bigwedge_{k=c+1}^{\text{len}-1} \neg(r(x_{k-1}) \in \{r(T_i^s)\} \wedge r(x_k) = r(\tilde{d}_{v11}^s) \wedge r(x_{k+1}) \in \{r(T_i^s)\}) \right) \right) \rightarrow \left(\bigwedge_{p=1}^j r(y_p) = r(x_p) \right) \wedge \left(\bigwedge_{b=c+1}^{\text{len}-1} r(y_b) = r(x_{b+1}) \right), \quad c \in \overline{2, |\bar{x}| - 1}; \quad (10)$$

$$P_3(\bar{x}, \bar{d}_{v1}^s, \bar{d}_{v2}^s, \bar{y}) = (|\bar{d}_{v1}^s| > 0) \wedge (|\bar{d}_{v2}^s| > 0) \wedge (|\bar{d}_{v1}^s| = \text{len}1) \wedge (|\bar{d}_{v2}^s| = \text{len}2) \wedge (|\bar{x}| = \text{len}3) \wedge \left(\bigwedge_{m=1}^{\text{len}1} r(x_{c+m}) = r(\tilde{d}_{v1m}^s) \right) \wedge \left((c + \text{len}1 = \text{len}3) \vee \left(\bigwedge_{k=c+1}^{\text{len}3} \neg \left(\bigwedge_{n=1}^{\text{len}1} (r(x_{k+n}) = r(\tilde{d}_{v1n}^s)) \right) \right) \right) \rightarrow$$

$$\left(c = 0 \vee \left(\bigwedge_{p=1}^j r(y_p) = r(x_p) \right) \right) \wedge \left(\bigwedge_{z=1}^{len2} r(y_{(c+1)+z}) = r(\tilde{d}_{v2z}^s) \right) \wedge \left(\bigwedge_{b=1}^{len3-(c+len1)} r(y_{(c+len2+1)+b}) = r(x_{(c+len1+1)+b}) \right), c \in \overline{0, |x|-1}. \quad (11)$$

5. Разработка структуры синтеза слов

Структура синтеза слов (рис. 1), используемых в технологической карте (блок 11) предусматривает создание слова (блок 10) из порождающего слова (блок 2) и набора словообразовательных аффиксов (блок 9). Из порождающих слов выделяется основа (блок 4), путем усечения неосновной части (блок 3). Основа преобразуется к своим формам (блоки 6, 8) посредством наборов наращиваемых/усекаемых (блок 5) и чередующихся (блок 7) буквенно/фонемных последовательностей. Конструирование слова происходит в соответствии с частью речи (блок 1).

6. Метод построения модели синтеза слова

Метод построения модели включает в себя:

- создание структуры модели;
- формирование опорных образцов для обучения модели;
- разработку процедуры синтеза слова по модели;
- создание критериев оценки эффективности модели.

$$M(\bar{y}) = f(M(\bar{x})) = f_6(f_5(f_4(f_3(f_2(f_1(M(\bar{x}))))))) \quad (12)$$

6.2. Формирование опорных образцов

Для синтеза слов создаются опорные образцы $\bar{V}_u^s, \bar{D}l_m^s, \bar{D}l_v^s, \bar{B}l_j^s, \bar{C}_r^s$ с взвешенными нормами $M(\bar{V}_u^s), M(\bar{D}l_m^s), M(\bar{D}l_v^s), M(\bar{B}l_j^s), M(\bar{C}_r^s)$. Обучение модели состоит в добавлении в нее сформированных опорных образцов.

6.1. Создание структуры модели

Для синтеза слова предложена модель шестислойной семиотико-алгебраической нейронной сети (рис. 2). Модель нейросети синтеза слов, формируется по следующему принципу. Порождающему слову соответствует нейрон входного слоя, $N^{(0)} = 1$. Опорным образцам неосновных частей слов соответствуют нейроны первого слоя. Опорным образцам, состоящим из наращиваемых/усекаемых буквенных/фонемных последовательностей для линейных морфологических преобразований, соответствуют нейроны второго слоя. Опорным образцам, состоящим из чередующихся буквенных/фонемных последовательностей для нелинейных морфологических преобразований, соответствуют нейроны третьего слоя. Опорным образцам, состоящим из добавляемой основы и словообразовательных аффиксов, соответствуют нейроны четвертого слоя. Опорным образцам порожденных слов соответствуют нейроны пятого слоя.

Модель нейросети синтеза слов представлена в виде (12)

6.3. Процедура синтеза слов по модели

После формирования структуры модели сети и добавления в нее опорных образцов осуществляется синтез слов для формирования ассоциативных связей между ЕЯО. Процедура синтеза для модели, на вход которой подается образец порождающего слова \bar{X} (соответствует \bar{Q}_e^s), состоит из инициализации, прямого и обратного хода.

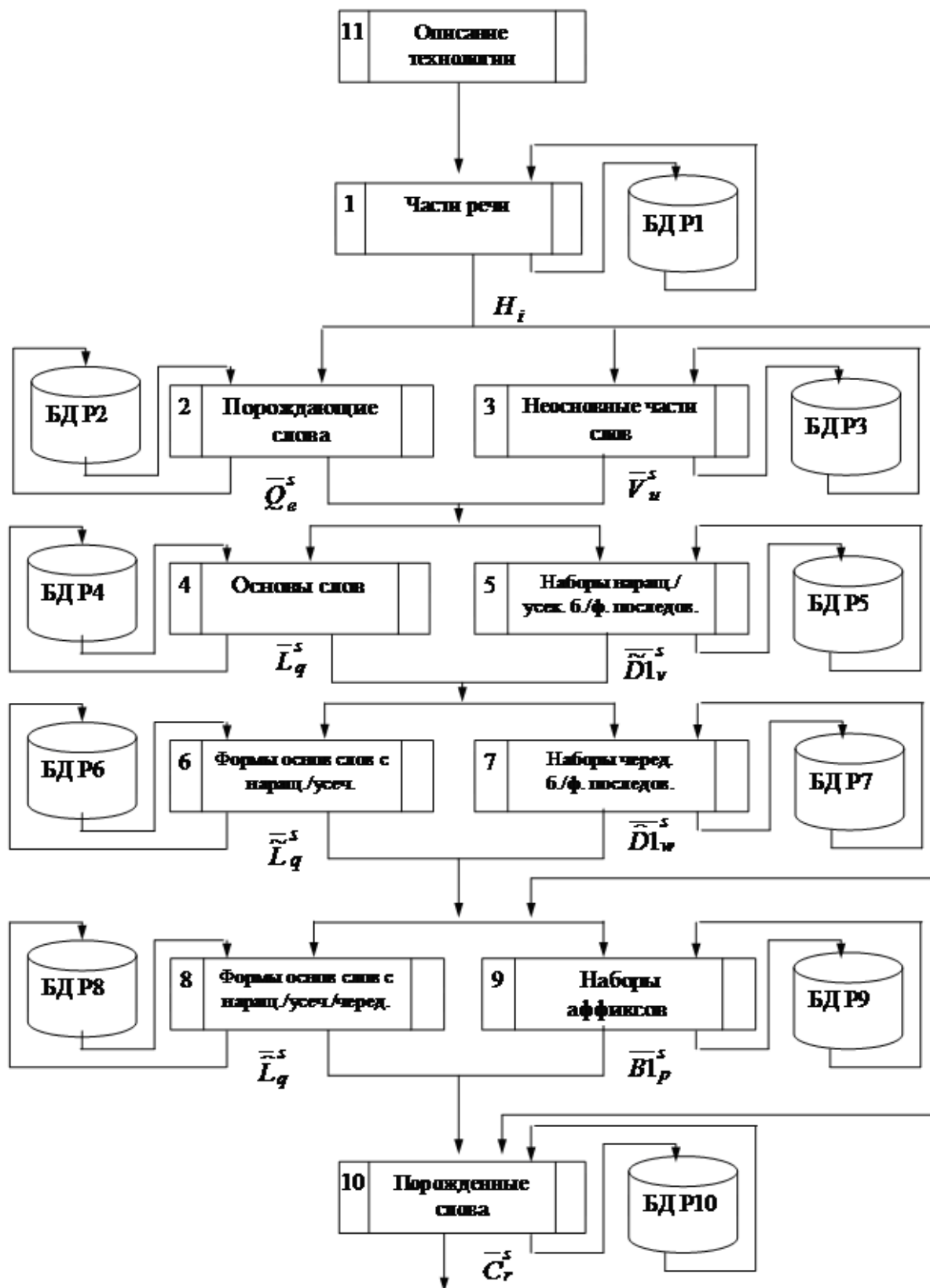


Рис. 1. Структурная детализация синтеза слов

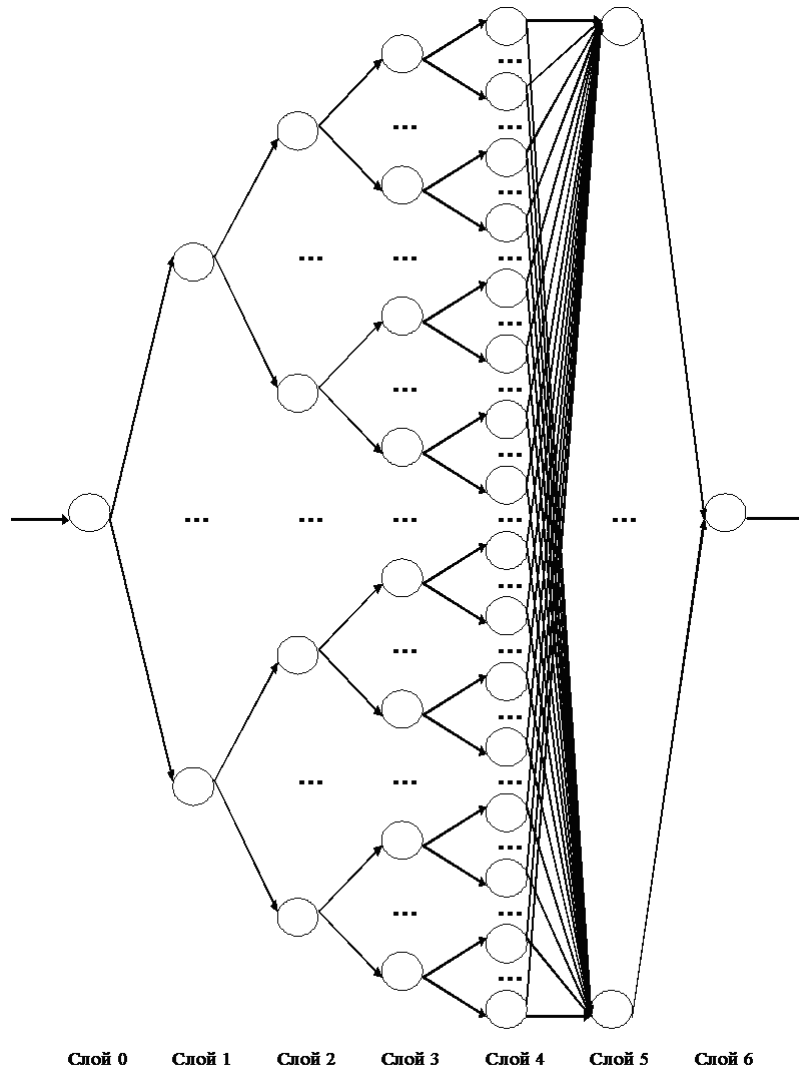


Рис. 2. Структура модели семиотико-алгебраической нейросети синтеза слов

1. Инициализация

Производится инициализация весов связей

$$w_{ij}^{(t)} = 1, t \in \overline{1,6}. \quad (13)$$

2. Прямой ход

Первый слой состоит из $\eta(\overline{\mathbf{V}}^s)$ нейронов, каждый из которых соответствует взвешенной норме неосновной части слова $M(\overline{V}_j^s)$. В этом слое вычисляется взвешенная норма основ порождающих слов, используя функцию активации (14).

Второй слой состоит из $\eta(\overline{\mathbf{V}}^s)$ групп, каждая из которых содержит $\eta(\overline{\mathbf{D1}}^s)$ нейронов, соответствующих век-

торам наращиваемых/усекаемых буквенных/фонемных последовательностей

$$\mu(\overline{\mathbf{D1}}_m^s) = (M(\overline{d}_{m1}^s), M(\overline{d}_{m2}^s), M(\overline{d}_{m3}^s)) \quad \text{с}$$

взвешенной нормой $M(\overline{\mathbf{D1}}_m^s)$. В этом слое формируется взвешенная норма форм основ порождающих слов (с наращением/усечением букв/фонем), используя функцию активации (15).

Третий слой состоит из $\eta(\overline{\mathbf{V}}^s) \cdot \eta(\overline{\mathbf{D1}}^s)$ групп, каждая из которых содержит $\eta(\overline{\mathbf{D1}}^s)$ нейронов, соответствующих векторам чередующихся буквенных/фонемных последовательностей $\mu(\overline{\mathbf{D1}}_v^s) = (M(\overline{d}_{v1}^s), M(\overline{d}_{v2}^s))$ с взвешенной

нормой $M(\overline{D1}_v^s)$. В этом слое формируется взвешенная норма форм основ порождающих слов (с наращением/усечением/чередованием букв/фонем), используя функцию активации (16).

Четвертый слой состоит из $\eta(\overline{V}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{D1}^s)$ групп, каждая из которых содержит $\eta(\overline{B1}^s)$ нейронов, соответствующих векторам добавляемой основы и аффиксов $\mu(\overline{B1}_p^s) = (M(\overline{l}_p^s), M(\overline{b}_{p1}^s), M(\overline{b}_{p2}^s), M(\overline{b}_{p3}^s), M(\overline{b}_{p4}^s), M(\overline{b}_{p5}^s))$ с взвешенной нормой $M(\overline{B1}_p^s)$. В этом слое формируется взвешенная норма порожденного слова, используя функцию активации (17).

Пятый слой состоит из $\eta(\overline{V}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{B1}^s)$ групп, каждая из которых содержит $\eta(\overline{C}^s)$ нейронов, каждый из которых соответствует взвешенной норме порожденного слова $M(\overline{C}_r^s)$. В этом слое полученные слова сопоставляются с допустимыми, используя функцию активации.

Выходной слой содержит один нейрон. В этом слое анализируются полученные результаты, и выдается порожденное слово \overline{y}_1 , используя функцию активации (18)

$$M(x_j^{-(1)}) = f_1(M(x_1^{-(0)}), M(\overline{V}_j^s)) = w_{1j}^{(1)} \cdot (M(x_1^{-(0)}) \circ M(\overline{V}_j^s)), \quad (14)$$

$$N^{(1)} = \eta(\overline{V}^s), \quad j \in \overline{1, N^{(1)}}.$$

$$\begin{aligned} M(x_j^{-(2)}) &= f_2(M(x_i^{-(1)}), M(\overline{d}_{m1}^s), M(\overline{d}_{m2}^s), M(\overline{d}_{m3}^s)) = \\ &= w_{ij}^{(2)} \cdot (((M(x_i^{-(1)}) \bullet M(\overline{d}_{m1}^s)) \circ M(\overline{d}_{m3}^s)) \diamond M(\overline{d}_{m2}^s)), \end{aligned} \quad (15)$$

$$N^{(3)} = \eta(\overline{V}^s) \cdot \eta(\overline{D1}^s), \quad m = j \bmod \eta(\overline{D1}^s), \quad i \in \overline{1, N^{(1)}}, \quad j \in \overline{1, N^{(3)}}.$$

$$M(x_j^{-(3)}) = f_3(M(x_i^{-(2)}), M(\overline{d}_{v1}^s), M(\overline{d}_{v2}^s)) = w_{ij}^{(3)} g(M(x_i^{-(2)}), M(\overline{d}_{v1}^s), M(\overline{d}_{v2}^s)),$$

$$N^{(3)} = \eta(\overline{V}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{D1}^s), \quad v = j \bmod \eta(\overline{D1}^s), \quad i \in \overline{1, N^{(2)}}, \quad j \in \overline{1, N^{(3)}}.$$

$$g(M(x_i^{-(2)}), M(\overline{d}_{v1}^s), M(\overline{d}_{v2}^s)) = \begin{cases} M(\overline{\zeta}_1^s), & (\overline{\zeta}_1^s = \overline{\xi}^s) \wedge P_1(x_i^{-(2)}, \overline{d}_{v1}^s, \overline{d}_{v2}^s, \overline{\xi}^s) \\ M(\overline{\zeta}_2^s), & (\overline{\zeta}_2^s = \overline{\xi}^s) \wedge P_2(x_i^{-(2)}, \overline{d}_{v1}^s, \overline{d}_{v2}^s, \overline{\xi}^s), \\ M(\overline{\zeta}_3^s), & (\overline{\zeta}_3^s = \overline{\xi}^s) \wedge P_3(x_i^{-(2)}, \overline{d}_{v1}^s, \overline{d}_{v2}^s, \overline{\xi}^s) \end{cases}$$

$$M(\overline{\zeta}_k^s) = \left\| \mu(\overline{\zeta}_k^s) \right\| = \sum_{t=1}^N \left(|r(\zeta_{kt}^s)| * 2^{(N-t)n} \right), \quad k \in \overline{1, 3}$$

$$\begin{aligned} M(x_j^{-(4)}) &= f_2(M(x_1^{-(3)}), M(\overline{l}_p^s), M(\overline{b}_{p1}^s), M(\overline{b}_{p2}^s), M(\overline{b}_{p3}^s), M(\overline{b}_{p4}^s), M(\overline{b}_{p5}^s)) = \\ &= w_{ij}^{(4)} \cdot (M(\overline{b}_{p1}^s) \diamond M(\overline{l}_p^s) \diamond M(\overline{b}_{p2}^s) \diamond M(x_1^{-(3)}) \diamond M(\overline{b}_{p3}^s) \diamond M(\overline{b}_{p4}^s) \diamond M(\overline{b}_{p5}^s)), \end{aligned} \quad (16)$$

$$N^{(4)} = \eta(\overline{V}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{B1}^s), \quad p = j \bmod \eta(\overline{B1}^s), \quad i \in \overline{1, N^{(3)}}, \quad j \in \overline{1, N^{(4)}}.$$

$$M(x_j^{-(5)}) = f_5(M(x_1^{-(4)}), \dots, M(x_{N^{(4)}}^{-(4)}), M(\overline{C}_r^s)) = \sum_{i=1}^{N^{(4)}} w_{ij}^{(5)} \cdot \varphi(M(x_i^{-(4)}), M(\overline{C}_r^s)), \quad (17)$$

$$N^{(5)} = \eta(\overline{V}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{D1}^s) \cdot \eta(\overline{B1}^s) \cdot \eta(\overline{C}^s), \quad r = j \bmod \eta(\overline{C}^s), \quad i \in \overline{1, N^{(4)}}, \quad j \in \overline{1, N^{(5)}}.$$

$$\varphi(M(\bar{x}_i^{-(4)}), M(\bar{C}_r^s)) = \begin{cases} M(\bar{C}_r^s), & M(\bar{x}_i^{-(4)}) = M(\bar{C}_r^s) \\ 0, & M(\bar{x}_i^{-(4)}) \neq M(\bar{C}_r^s) \end{cases}$$

$$M(\bar{y}_1) = M(\bar{x}_1^{-(6)}) = f_6(M(\bar{x}_1^{-(5)}), \dots, M(\bar{x}_{N^{(5)}}^{-(5)})) = \sum_{i=1}^{N^{(5)}} w_{i1}^{(6)} \cdot M(\bar{x}_i^{-(5)}). \quad (18)$$

3. Обратный ход

Весовые коэффициенты модели вычисляются в виде (19)-(24)

$$w_{i1}^{(6)} = \begin{cases} 1, & M(\bar{y}_1) = f_6(M(\bar{x}_1^{-(5)}), \dots, M(\bar{x}_{N^{(5)}}^{-(5)})), \\ 0, & \text{иначе} \end{cases}, \quad (19)$$

$$w_{ij}^{(5)} = \begin{cases} 1, & M(\bar{x}_j^{-(5)}) = f_5(M(\bar{x}_1^{-(4)}), \dots, M(\bar{x}_{N^{(4)}}^{-(4)}), M(\bar{C}_r^s)) \wedge w_{j1}^{(6)} = 1, \\ 0, & \text{иначе} \end{cases}, \quad (20)$$

$$w_{ij}^{(4)} = \begin{cases} 1, & M(\bar{x}_j^{-(4)}) = f_4(M(\bar{x}_1^{-(3)}), M(\bar{l}_p^s), M(\bar{b}_{p1}^s), M(\bar{b}_{p2}^s), M(\bar{b}_{p3}^s), M(\bar{b}_{p4}^s), M(\bar{b}_{p5}^s)) \\ & \wedge \sum_k w_{jk}^{(5)} = 1 \\ 0, & \text{иначе} \end{cases}, \quad (21)$$

$$w_{ij}^{(3)} = \begin{cases} 1, & M(\bar{x}_j^{-(3)}) = f_3(M(\bar{x}_i^{-(2)}), M(\bar{d}_{v1}^s), M(\bar{d}_{v2}^s)) \wedge \sum_k w_{jk}^{(4)} = 1, \\ 0, & \text{иначе} \end{cases}, \quad (22)$$

$$w_{ij}^{(2)} = \begin{cases} 1, & M(\bar{x}_j^{-(2)}) = f_2(M(\bar{x}_i^{-(1)}), M(\bar{d}_{m1}^s), M(\bar{d}_{m2}^s)) \wedge \sum_k w_{jk}^{(3)} = 1, \\ 0, & \text{иначе} \end{cases}, \quad (23)$$

$$w_{1j}^{(1)} = \begin{cases} 1, & M(\bar{x}_j^{-(1)}) = f_1(M(\bar{x}_i^{-(0)}), M(\bar{V}_j^s)) \wedge \sum_k w_{jk}^{(2)} = 1 \\ 0, & \text{иначе} \end{cases} \quad (24)$$

В результате процедуры синтеза формируются ассоциативные связи «порождающее слово – порожденное слово» вида

$$\bar{Q}_e^s \rightarrow \bar{L}_q^s \rightarrow \bar{L}_v^s \rightarrow \bar{L}_w^s \rightarrow \bar{C}_r^s \quad (25)$$

6.4. Критерии оценки эффективности модели синтеза слов

В работе выбраны следующие критерии оценки синтеза слов

1. Критерий скорости синтеза (26), который для данного случая означает выбор такого количества опорных образцов $N^{(1)}, N^{(2)}, N^{(3)}, N^{(4)}, N^{(5)}$, которое достав-

ляет минимум времени синтеза по модели.

2. Критерий вероятности синтеза (27), который для данного случая означает выбор таких опорных образцов, которые доставляют максимум вероятности правильного синтеза (отношения количества правильно синтезированных слов к их общему количеству).

3. Критерий адекватности модели (28), который для данного случая означает выбор таких опорных образцов, которые доставляют минимум среднеквадратичной ошибки (разности выхода по модели и тестового выхода).

$$F = T(N^{(1)}, N^{(2)}, N^{(3)}, N^{(4)}, N^{(5)}) \rightarrow \min_{N^{(1)}, N^{(2)}, N^{(3)}, N^{(4)}, N^{(5)}} \quad (26)$$

$$F = \frac{1}{P} \sum_{p=1}^P I(f2(M(\bar{x}_p)), M(\bar{y}_p)) \rightarrow \max_{M(\bar{B}1_j^s), M(\bar{D}1_v^s), M(\bar{D}1_m^s), M(\bar{V}_u^s), M(\bar{Q}_e^s)}, \quad (27)$$

$$I(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases}$$

$$F = \frac{1}{P} \sum_{p=1}^P (f2(M(\bar{x}_p)) - M(\bar{y}_p))^2 \rightarrow \min_{M(\bar{B}1_j^s), M(\bar{D}1_v^s), M(\bar{D}1_m^s), M(\bar{V}_u^s), M(\bar{Q}_e^s)}. \quad (28)$$

7. Формирования логико-формальных правил оценивания результата синтеза слов

Правила для синтеза слов представлены в виде (29) и (30), где q – номер слова.

$$\Gamma(\bar{Q}_e^s, \bar{C}_r^s) = 1 \wedge M(\bar{x}) = M(\bar{Q}_e^s) \wedge M(\bar{y}) = M(\bar{C}_r^s) \rightarrow q = r \quad (29)$$

$$\Gamma(\bar{Q}_e^s, \bar{C}_r^s) = 1 \wedge M(\bar{x}) = M(\bar{Q}_e^s) \wedge M(\bar{y}) \neq M(\bar{C}_r^s) \rightarrow q = 0 \quad (30)$$

Выводы

Новизна. Основным результатом данной статьи является метод синтеза слов для нейросетевого лингвистического процессора, который включает в себя: определения вектора признаков ЕЯО и его взвешенной нормы; формализация взаимосвязей между ЕЯО; формализация операций над ЕЯО; формализация морфонологических преобразований; структура синтеза слов; метод построения модели синтеза слов; логико-формальные правила оценивания результата синтеза слов. Впервые формализованы правила линейных и нелинейных морфонологических преобразований слов на основе алгебры конечных предикатов и даны количественные оценки ЕЯО на основе взвешенной нормы.

Практическое значение. Основные положения работы реализованы на языке Matlab 7.1 в виде алгоритмов, обеспечивающих общение с человеком на естественном языке, и могут быть использованы в различных компьютерных системах.

Литература

1. Шуклин Д.Е. Разработка системы обрабатывающей текст естественного языка на основе семантической нейронной сети / Д.Е. Шуклин // Искусственный интеллект. Донецк. Изд-во Института

проблем искусственного интеллекта НАН Украины, 2002. - №4. С.312-320

2. Бондаренко М.Ф. Основы теории синтеза над швидкодіючих структур мовних систем штучного інтелекту: моногр. / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – К.: ІЗМН, 1997. – 264 с.

3. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Технические средства / Ю.П. Шабанов-Кушнарченко. – Х.: Вища шк. Изд-во при Харьк. ун-те, 1986. – 136 с.

4. Современный русский язык / В.А. Белошапкова, Е.А. Брызгунова, Е.А. Земская и др.; под ред. В.А. Белошапковой. – М.: Азбуковник, 1997. – 928 с.

5. Русская грамматика: [в 2 т.] / под ред. Н.Ю. Шведовой. – М.: Наука, 1982. – Т. 1: Фонетика. Фонология. Ударение. Интонация. Введение в морфемику. Словообразование. Морфология. – 1982. – 784 с.

6. Бондарев В.Н. Искусственный интеллект / В.Н. Бондарев, Ф.Г. Аде. – Севастополь: Изд-во СевНТУ, 2002. – 615 с.

7. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Изд. дом «Вильямс», 2001. – 288 с.

8. Pollack J. Recursive distributed representations / J. Pollack // Artificial Intelligence. – № 46. – P.77-105.