

УДК 004.934.1'1

**Е.Е. ФЕДОРОВ**

*Донецкий государственный институт искусственного интеллекта, Украина*

## МЕТОДИКА НЕЙРОСЕТЕВОГО АНАЛИЗА СЛОВ

Для разработки естественно-языкового интерфейса автоматизированной системы управления (АСУ) в статье предлагается методика нейросетевого анализа слов технического языка, на котором осуществляется взаимодействие между человеком и АСУ.

**методика нейросетевого анализа слов, АСУ, естественно-языковой интерфейс, модель технического языка, структура предприятия, особенности функционирования предприятия**

**Постановка проблемы.** Современные тенденции, отраженные в методологии создания естественно-языковых (ЕЯ) интерфейсов, в мировой практике имеют два четко выраженных направления: первое – универсальное, второе – специализированное. Первое направление нашло свое отображение в концепции создания ЭВМ 5-го поколения. Второе направление используется при создании субъектно- и объектно-ориентированных систем с четко очерченными границами предметной области.

В данной работе за основу взято второе направление. Предметной областью является производственная сфера. В соответствии с ней выбран профессиональный язык человека, управляющего технологией, который позволяет описать структуру и особенности функционирования предприятия (объекта управления). Структура предприятия содержит технологические агрегаты, предметы труда, продукцию. В технологических агрегатах производятся технологические операции над предметами труда и продукцией. Для технологических агрегатов задаются режимы работы. Производство заданного вида продукции определяется технологическими картами, которые содержат в себе задания на все уровни управления предприятием.

**Анализ исследований.** Традиционно для управления предприятием создавались автоматизированные системы управления (АСУ), связанные с ЭВМ.

Эти системы основываются на математических моделях технологических процессов [1-3].

Анализ последних достижений и публикаций, посвященных этой проблеме, позволяет сделать вывод, что эти модели не используют в качестве источника информации о структуре и особенностях функционирования объекта управления чувственное восприятие и экспертные оценки технологического процесса человеком, которые описываются посредством технического языка и отражают общие закономерности протекания процесса, как для конкретного агрегата и конкретных условий, так и для множества агрегатов данного типа

Нерешенным является вопрос, связанный с созданием модели технического языка, на котором осуществляется взаимодействие между человеком и АСУ.

**Постановка задачи.** Целью настоящей работы является создание методики нейросетевого анализа слов, используемой при создании модели технического языка.

**Решение задачи.** В данной статье предлагается методика нейросетевого анализа слов, для которой разрабатываются следующие правила:

- формализации и численного исследования элементов анализа слов (частей речи, порождающих и порожденных слов, морфонологических вариантов основ порождающих слов, словообразовательных

аффиксов);

- закрепления взаимосвязей между элементами анализа слов;
- анализа слов.

## 1. Правила формализации и численного исследования частей речи

Слова языка, описывающие объекты производственной системы и действия над ними, характеризуются определенными частями речи.

Множество пар частей речи формально определено в виде

$$\overline{\mathbf{H1}} = \bigcup_j \overline{H1_j}, \overline{H1_j} = (h_{j1}, h_{j2}), j \in \overline{1, \eta(\overline{\mathbf{H1}})}, \quad (1)$$

где  $h_{j1}$  – часть речи порождающего слова  $j$ -й пары;

$h_{j2}$  – часть речи порожденного слова  $j$ -й пары.

Каждой части речи  $h_{jt}$  ставится в соответствие ранг  $r(h_{jt})$ . Каждой паре  $\overline{H1_j}$ , используя ранги  $r(h_{jt})$ , ставится в соответствие информационная мера  $M(\overline{H1_j})$ .

Множество пар частей речи упорядочивается и преобразуется к вектору (2)

$$\overline{\mathbf{H1}} = (\overline{H1_1}, \dots, \overline{H1_j}, \dots, \overline{H1_M});$$

$$M(\overline{H1_{j-1}}) < M(\overline{H1_j}) < M(\overline{H1_{j+1}}). \quad (2)$$

Вектору (2) соответствует вектор (3), содержащий информационные меры

$$M(\overline{\mathbf{H1}}) = (M(\overline{H1_1}), \dots, M(\overline{H1_j}), \dots, M(\overline{H1_M}));$$

$$M(\overline{H1_{j-1}}) < M(\overline{H1_j}) < M(\overline{H1_{j+1}}). \quad (3)$$

## 2. Правила формализации и численного исследования порождающих слов

Слова языка, используются для описания объектов производственной системы и действий над ними.

Множество порождающих слов в буквенном представлении формально определено в виде (4)

$$\overline{\mathbf{W}^1} = \bigcup_r \overline{W_r^1};$$

$$\overline{W_r^1} = (w_{r1}^1, \dots, w_{rt}^1, \dots, w_{rN}^1), \quad r \in \overline{1, \eta(\overline{\mathbf{W}^1})}, \quad (4)$$

где  $r$  – номер слова  $\overline{W_r^1}$ ;

$t$  – номер буквы  $w_{rt}^1$  в слове  $\overline{W_r^1}$ ;

$\eta(\overline{\mathbf{W}^1})$  – количество слов.

Аналогично множество слов в фонемном представлении формально определено в виде (5)

$$\overline{\mathbf{W}^2} = \bigcup_r \overline{W_r^2};$$

$$\overline{W_r^2} = (w_{r1}^2, \dots, w_{rt}^2, \dots, w_{rN}^2), \quad r \in \overline{1, \eta(\overline{\mathbf{W}^2})}, \quad (5)$$

где  $r$  – номер слова  $\overline{W_r^2}$ ;

$t$  – номер фонемы  $w_{rt}^2$  в слове  $\overline{W_r^2}$ .

Каждой букве  $w_{rt}^1$  ставится в соответствие ранг  $r(w_{rt}^1)$ , а каждой фонеме  $w_{rt}^2$  – ранг  $r(w_{rt}^2)$ . Каждому слову  $\overline{W_r^1}$ , используя ранги  $r(w_{rt}^1)$ , ставится в соответствие информационная мера  $M(\overline{W_r^1})$ , а слову  $\overline{W_r^2}$ , используя ранги  $r(w_{rt}^2)$ , – информационная мера  $M(\overline{W_r^2})$ .

Множество слов в буквенном представлении упорядочивается и преобразуется к вектору (6)

$$\overline{\mathbf{W}^1} = (\overline{W_1^1}, \dots, \overline{W_r^1}, \dots, \overline{W_M^1});$$

$$M(\overline{W_{r-1}^1}) < M(\overline{W_r^1}) < M(\overline{W_{r+1}^1}). \quad (6)$$

Аналогично множество слов в фонемном представлении преобразуется к вектору (7)

$$\overline{\mathbf{W}^2} = (\overline{W_1^2}, \dots, \overline{W_r^2}, \dots, \overline{W_M^2});$$

$$M(\overline{W_{r-1}^2}) < M(\overline{W_r^2}) < M(\overline{W_{r+1}^2}). \quad (7)$$

Вектору (6) соответствует вектор (8), содержащий информационные меры:

$$M(\overline{\mathbf{W}^1}) = (M(\overline{W_1^1}), \dots, M(\overline{W_r^1}), \dots, M(\overline{W_M^1}));$$

$$M(\overline{W_{r-1}^1}) < M(\overline{W_r^1}) < M(\overline{W_{r+1}^1}). \quad (8)$$

Аналогічно вектору (7) відповідає вектор (9), що містить інформаційні міри:

$$M(\overline{W}^2) = (M(\overline{W}_1^2), \dots, M(\overline{W}_r^2), \dots, M(\overline{W}_M^2));$$

$$M(\overline{W}_{r-1}^2) < M(\overline{W}_r^2) < M(\overline{W}_{r+1}^2). \quad (9)$$

### 3. Правила формалізації та численного дослідження основ породжуючих слів

Основа є головною частиною слів, використовуваних при описанні об'єктів виробничої системи та дій над ними. Основи виділяються з слів або беруть участь в їх аналізі відповідно до частин мови та наборів морфологічних ознак.

Множина основ в буквенному представленні формально визначена в вигляді (10)

$$\overline{L}^1 = \bigcup_q \overline{L}_q^1;$$

$$\overline{L}_q^1 = (l_{q1}^1, \dots, l_{qt}^1, \dots, l_{qM}^1), \quad q \in 1, \eta(\overline{L}^1), \quad (10)$$

де  $q$  – номер основи  $\overline{L}_q^1$ ;

$t$  – номер букви  $l_{qt}^1$  в основі  $\overline{L}_q^1$ ;

$\eta(\overline{L}^1)$  – кількість основ.

Аналогічно множина основ в фонемному представленні формально визначена в вигляді (11)

$$\overline{L}^2 = \bigcup_q \overline{L}_q^2;$$

$$\overline{L}_q^2 = (l_{q1}^2, \dots, l_{qt}^2, \dots, l_{qM}^2), \quad q \in 1, \eta(\overline{L}^2), \quad (11)$$

де  $q$  – номер основи  $\overline{L}_q^2$ ;

$t$  – номер фонем  $l_{qt}^2$  в основі  $\overline{L}_q^2$ .

Кожній букві  $l_{qt}^1$  ставиться в відповідності з рангом  $r(l_{qt}^1)$ , а кожній фонемі  $l_{qt}^2$  – ранг  $r(l_{qt}^2)$ . Кожній основі  $\overline{L}_q^1$ , використовуючи ранги  $r(l_{qt}^1)$ , ставиться в відповідності інформаційна міра  $M(\overline{L}_q^1)$ , а основі

$\overline{L}_q^2$ , використовуючи ранги  $r(l_{qt}^2)$ , – інформаційна міра  $M(\overline{L}_q^2)$ .

Множина основ в буквенному представленні упорядковується та перетворюється в вектор

$$\overline{L}^1 = (\overline{L}_1^1, \dots, \overline{L}_q^1, \dots, \overline{L}_M^1);$$

$$M(\overline{L}_{q-1}^1) < M(\overline{L}_q^1) < M(\overline{L}_{q+1}^1). \quad (12)$$

Аналогічно множина основ в фонемному представленні перетворюється в вектор (13)

$$\overline{L}^2 = (\overline{L}_1^2, \dots, \overline{L}_q^2, \dots, \overline{L}_M^2);$$

$$M(\overline{L}_{q-1}^2) < M(\overline{L}_q^2) < M(\overline{L}_{q+1}^2). \quad (13)$$

Вектору (12) відповідає вектор (14), що містить інформаційні міри:

$$M(\overline{L}^1) = (M(\overline{L}_1^1), \dots, M(\overline{L}_q^1), \dots, M(\overline{L}_M^1));$$

$$M(\overline{L}_{q-1}^1) < M(\overline{L}_q^1) < M(\overline{L}_{q+1}^1). \quad (14)$$

Аналогічно вектору (13) відповідає вектор (15), що містить інформаційні міри:

$$M(\overline{L}^2) = (M(\overline{L}_1^2), \dots, M(\overline{L}_q^2), \dots, M(\overline{L}_M^2));$$

$$M(\overline{L}_{q-1}^2) < M(\overline{L}_q^2) < M(\overline{L}_{q+1}^2). \quad (15)$$

### 4. Правила формалізації та численного дослідження морфологічних варіантів основ породжуючих слів

Виділені з слів основи можуть піддаватися морфологічним перетворенням – усечення, нарощення та чередування букв/фонем.

Множина морфологічних варіантів основ в буквенному представленні формально визначена в вигляді (16)

$$\overline{\tilde{L}}^1 = \bigcup_q \overline{\tilde{L}}_q^1;$$

$$\overline{\tilde{L}}_q^1 = (l_{q1}^1, \dots, l_{qt}^1, \dots, l_{qM}^1), \quad q \in 1, \eta(\overline{\tilde{L}}^1), \quad (16)$$

де  $q$  – номер основи  $\overline{\tilde{L}}_q^1$ ;

$t$  – номер букви  $l_{qt}^1$  в основі  $\overline{\tilde{L}}_q^1$ ;

$\eta(\bar{L}^1)$  – количество основ.

Аналогично множество основ в фонемном представлении формально определено в виде (17)

$$\bar{L}^2 = \bigcup_q \bar{L}_q^2;$$

$$\bar{L}_q^2 = (l_{q1}^2, \dots, l_{qt}^2, \dots, l_{qM}^2), \quad q \in \overline{1, \eta(\bar{L}^1)}, \quad (17)$$

где  $q$  – номер основы  $\bar{L}_q^2$ ;

$t$  – номер фонемы  $l_{qt}^2$  в основе  $\bar{L}_q^2$ .

Каждой букве  $l_{qt}^1$  ставится в соответствие ранг  $r(l_{qt}^1)$ , а каждой фонеме  $l_{qt}^2$  – ранг  $r(l_{qt}^2)$ . Каждой основе  $\bar{L}_q^1$ , используя ранги  $r(l_{qt}^1)$ , ставится в соответствие информационная мера  $M(\bar{L}_q^1)$ , а основе  $\bar{L}_q^2$ , используя ранги  $r(l_{qt}^2)$ , – информационная мера  $M(\bar{L}_q^2)$ .

Множество основ в буквенном представлении упорядочивается и преобразуется к вектору (18)

$$\bar{L}^1 = (\bar{L}_1^1, \dots, \bar{L}_q^1, \dots, \bar{L}_M^1);$$

$$M(\bar{L}_{q-1}^1) < M(\bar{L}_q^1) < M(\bar{L}_{q+1}^1). \quad (18)$$

Аналогично множество основ в фонемном представлении преобразуется к вектору (19)

$$\bar{L}^2 = (\bar{L}_1^2, \dots, \bar{L}_q^2, \dots, \bar{L}_M^2);$$

$$M(\bar{L}_{q-1}^2) < M(\bar{L}_q^2) < M(\bar{L}_{q+1}^2). \quad (19)$$

Вектору (18) соответствует вектор (20), содержащий информационные меры:

$$M(\bar{L}^1) = (M(\bar{L}_1^1), \dots, M(\bar{L}_q^1), \dots, M(\bar{L}_M^1));$$

$$M(\bar{L}_{q-1}^1) < M(\bar{L}_q^1) < M(\bar{L}_{q+1}^1). \quad (20)$$

Аналогично вектору (19) соответствует вектор (21), содержащий информационные меры:

$$M(\bar{L}^2) = (M(\bar{L}_1^2), \dots, M(\bar{L}_q^2), \dots, M(\bar{L}_M^2));$$

$$M(\bar{L}_{q-1}^2) < M(\bar{L}_q^2) < M(\bar{L}_{q+1}^2). \quad (21)$$

## 5. Правила формализации и численного исследования словообразовательных аффиксов

Переход между порождающим и порожденным словом, используемым при описании объектов производственной системы и действий над ними, осуществляется посредством частей слов (словообразовательные аффиксы). Словообразовательные аффиксы выделяются из слов или участвуют в их анализе в соответствии с частями речи и наборами морфологических признаков.

Множества словообразовательных аффиксов в буквенном представлении – префиксы, интерфиксы, суффиксы, флексии и постфиксы формально определены как  $\bar{B}^{11}, \bar{B}^{12}, \bar{B}^{13}, \bar{B}^{14}, \bar{B}^{15}$  в виде

$$\bar{B}^{1k} = \bigcup_m \bar{B}_m^{1k}; \quad \bar{B}_m^{1k} = (\sigma_{m1}^{1k}, \dots, \sigma_{mt}^{1k}, \dots, \sigma_{mM}^{1k});$$

$$k \in \overline{1,5}, m \in \overline{1, \eta(\bar{B}^{1k})}, \quad (22)$$

где  $k$  – номер типа аффикса  $\bar{B}_m^{1k}$ ;

$m$  – номер аффикса  $\bar{B}_m^{1k}$ ;

$t$  – номер буквы  $\sigma_{mt}^{1k}$  в аффиксе  $\bar{B}_m^{1k}$ ;

$\eta(\bar{B}^{1k})$  – количество аффиксов  $\bar{B}_m^{1k}$ .

Аналогично множества словообразовательных аффиксов в фонемном представлении – префиксы  $\bar{B}^{21}$ , интерфиксы  $\bar{B}^{22}$ , суффиксы  $\bar{B}^{23}$ , флексии  $\bar{B}^{24}$ , постфиксы  $\bar{B}^{25}$  формально определены в виде (23)

$$\bar{B}^{2k} = \bigcup_m \bar{B}_m^{2k}; \quad \bar{B}_m^{2k} = (\sigma_{m1}^{2k}, \dots, \sigma_{mt}^{2k}, \dots, \sigma_{mM}^{2k}),$$

$$k \in \overline{1,5}, m \in \overline{1, \eta(\bar{B}^{2k})}, \quad (23)$$

где  $k$  – номер типа аффикса  $\bar{B}_m^{2k}$ ;

$m$  – номер аффикса  $\bar{B}_m^{2k}$ ;

$t$  – номер фонемы  $\sigma_{mt}^{2k}$  в аффиксе  $\bar{B}_m^{2k}$ .

Каждой букве  $\sigma_{mt}^1$  ставится в соответствие ранг  $r(\sigma_{mt}^1)$ , а каждой фонеме  $\sigma_{mt}^2$  – ранг  $r(\sigma_{mt}^2)$ . Каждому аффиксу  $\bar{B}_m^{1k}$ , используя ранги  $r(\sigma_{mt}^1)$ , ставится в соответствие информационная мера  $M(\bar{B}_m^{1k})$ , а аффиксу  $\bar{B}_m^{2k}$ , используя ранги  $r(\sigma_{mt}^2)$ , – информационная мера  $M(\bar{B}_m^{2k})$ .

Множество аффиксов в буквенном представлении упорядочивается и преобразуется к вектору (24)

$$\bar{\mathbf{B}}^{1k} = (\bar{B}_1^{1k}, \dots, \bar{B}_m^{1k}, \dots, \bar{B}_M^{1k});$$

$$M(\bar{B}_{m-1}^{1k}) < M(\bar{B}_m^{1k}) < M(\bar{B}_{m+1}^{1k}). \quad (24)$$

Аналогично множество аффиксов в фонемном представлении преобразуется к вектору (25)

$$\bar{\mathbf{B}}^{2k} = (\bar{B}_1^{2k}, \dots, \bar{B}_m^{2k}, \dots, \bar{B}_M^{2k});$$

$$M(\bar{B}_{m-1}^{2k}) < M(\bar{B}_m^{2k}) < M(\bar{B}_{m+1}^{2k}). \quad (25)$$

Вектору (24) соответствует вектор (26), содержащий информационные меры:

$$M(\bar{\mathbf{B}}^{1k}) = (M(\bar{B}_1^{1k}), \dots, M(\bar{B}_m^{1k}), \dots, M(\bar{B}_M^{1k}));$$

$$M(\bar{B}_{m-1}^{1k}) < M(\bar{B}_m^{1k}) < M(\bar{B}_{m+1}^{1k}). \quad (26)$$

Аналогично вектору (25) соответствует вектор (27), содержащий информационные меры:

$$M(\bar{\mathbf{B}}^{2k}) = (M(\bar{B}_1^{2k}), \dots, M(\bar{B}_m^{2k}), \dots, M(\bar{B}_M^{2k}));$$

$$M(\bar{B}_{m-1}^{2k}) < M(\bar{B}_m^{2k}) < M(\bar{B}_{m+1}^{2k}). \quad (27)$$

Множество наборов аффиксов в буквенном представлении, используемых при словообразовании, формально определено в виде

$$\bar{\mathbf{B}}^1 = \bigcup_j \bar{B}_j^1; \quad \bar{B}_j^1 = (\sigma_{j1}^1, \sigma_{j2}^1, \sigma_{j3}^1, \sigma_{j4}^1, \sigma_{j5}^1),$$

$$j \in \overline{1, \eta(\bar{\mathbf{B}}^1)}, \quad (28)$$

где  $\sigma_{jt}^1$  – аффикс  $j$ -го вектора  $t$ -го типа;

$\eta(\bar{\mathbf{B}}^1)$  – количество векторов.

Аналогично множество наборов аффиксов в фонемном представлении, используемых при словообразовании, формально определено в виде

$$\bar{\mathbf{B}}^2 = \bigcup_j \bar{B}_j^2; \quad \bar{B}_j^2 = (\sigma_{j1}^2, \sigma_{j2}^2, \sigma_{j3}^2, \sigma_{j4}^2, \sigma_{j5}^2),$$

$$j \in \overline{1, \eta(\bar{\mathbf{B}}^2)}, \quad (29)$$

где  $\sigma_{jt}^2$  – аффикс  $j$ -го вектора  $t$ -го типа.

Каждому набору аффиксов  $\bar{B}_j^s$ ,  $s \in \overline{1, 2}$ , используя информационные меры  $M(\sigma_{jt}^s)$ , ставится в соответствие информационная мера  $M(\bar{B}_j^s)$ .

Множество наборов аффиксов в символьном представлении упорядочивается и преобразуется к вектору (30)

$$\bar{\mathbf{B}}^s = (\bar{B}_1^s, \dots, \bar{B}_j^s, \dots, \bar{B}_M^s);$$

$$M(\bar{B}_{j-1}^s) < M(\bar{B}_j^s) < M(\bar{B}_{j+1}^s). \quad (30)$$

Вектору (30) соответствует вектор (31), содержащий информационные меры:

$$M(\bar{\mathbf{B}}^s) = (M(\bar{B}_1^s), \dots, M(\bar{B}_j^s), \dots, M(\bar{B}_M^s));$$

$$M(\bar{B}_{j-1}^s) < M(\bar{B}_j^s) < M(\bar{B}_{j+1}^s). \quad (31)$$

## 6. Правила формализации и численного исследования порожденных слов

Порожденные слова используются при описании объектов производственной системы и действий над ними. Они анализируются из основ порождающих слов и словообразовательных аффиксов в соответствии с частями речи и наборами морфологических признаков.

Множество порожденных слов в буквенном представлении формально определено в виде (32)

$$\bar{\mathbf{S}}^1 = \bigcup_l \bar{S}_l^1;$$

$$\bar{S}_l^1 = (s_{l1}^1, \dots, s_{li}^1, \dots, s_{lN}^1), \quad l \in \overline{1, \eta(\bar{\mathbf{S}}^1)}, \quad (32)$$

где  $l$  – номер порожденного слова  $\bar{S}_l^1$ ;

$t$  – номер буквы  $s_{lt}^1$  в слове  $\bar{S}_l^1$ ;

$\eta(\bar{S}^1)$  – количество слов.

Аналогично множество порожденных слов в фонемном представлении формально определено в виде (33)

$$\bar{S}^2 = \bigcup_l \bar{S}_l^2 ;$$

$$\bar{S}_l^2 = (s_{l1}^2, \dots, s_{lt}^2, \dots, s_{lN}^2), \quad l \in \overline{1, \eta(\bar{S}^1)}, \quad (33)$$

где  $l$  – номер порожденного слова  $\bar{S}_l^2$ ;

$t$  – номер фонемы  $s_{lt}^2$  в слове  $\bar{S}_l^2$ .

Каждой букве  $s_{lt}^1$  ставится в соответствие ранг  $r(s_{lt}^1)$ , а каждой фонеме  $s_{lt}^2$  – ранг  $r(s_{lt}^2)$ .

Каждому слову  $\bar{S}_l^1$ , используя ранги  $r(s_{lt}^1)$ , ставится в соответствие информационная мера  $M(\bar{S}_l^1)$ .

Каждому слову  $\bar{S}_l^2$ , используя ранги  $r(s_{lt}^2)$ , ставится в соответствие информационная мера  $M(\bar{S}_l^2)$ .

Множество порожденных слов в буквенном представлении упорядочивается и преобразуется к вектору (34)

$$\bar{S}^1 = (\bar{S}_1^1, \dots, \bar{S}_l^1, \dots, \bar{S}_M^1);$$

$$M(\bar{S}_{l-1}^1) < M(\bar{S}_l^1) < M(\bar{S}_{l+1}^1). \quad (34)$$

Аналогично множество порожденных слов в фонемном представлении преобразуется к вектору (35)

$$\bar{S}^2 = (\bar{S}_1^2, \dots, \bar{S}_l^2, \dots, \bar{S}_M^2);$$

$$M(\bar{S}_{l-1}^2) < M(\bar{S}_l^2) < M(\bar{S}_{l+1}^2). \quad (35)$$

Вектору (34) соответствует вектор (36), содержащий информационные меры:

$$M(\bar{S}^1) = (M(\bar{S}_1^1), \dots, M(\bar{S}_l^1), \dots, M(\bar{S}_M^1));$$

$$M(\bar{S}_{l-1}^1) < M(\bar{S}_l^1) < M(\bar{S}_{l+1}^1). \quad (36)$$

Аналогично вектору (35) соответствует вектор (37), содержащий информационные меры:

$$M(\bar{S}^2) = (M(\bar{S}_1^2), \dots, M(\bar{S}_l^2), \dots, M(\bar{S}_M^2));$$

$$M(\bar{S}_{l-1}^2) < M(\bar{S}_l^2) < M(\bar{S}_{l+1}^2). \quad (37)$$

## 7. Правила закрепления взаимосвязей между элементами анализа слов

В пунктах 1) – 6), согласно разработанным правилам, были формализованы, ранжированы и упорядочены элементы анализа слов, которые используются при описании объектов производственной системы и действий над ними. Для проведения анализа слов разрабатываются правила, устанавливающие логические взаимосвязи между этими элементами.

Введем следующие матрицы бинарных отношений между элементами анализа слов:

$$\Gamma(\overline{H1}_i, \overline{B1}_j^s); \Gamma(\overline{H1}_i, \overline{W}_r^s); \Gamma(\overline{H1}_i, \overline{L}_q^s);$$

$$\Gamma(\overline{H1}_i, \overline{L}_q^s); \Gamma(\overline{H1}_i, \overline{B}_m^{sk}); \Gamma(\overline{H1}_i, \overline{S}_l^s); \Gamma(\overline{W}_r^s, \overline{B1}_j^s);$$

$$\Gamma(\overline{W}_r^s, \overline{L}_q^s); \Gamma(\overline{W}_r^s, \overline{L}_q^s); \Gamma(\overline{W}_r^s, \overline{B}_m^{sk}); \Gamma(\overline{W}_r^s, \overline{S}_l^s);$$

$$\Gamma(\overline{B1}_j^s, \overline{L}_q^s); \Gamma(\overline{B1}_j^s, \overline{L}_q^s); \Gamma(\overline{B1}_j^s, \overline{B}_m^{sk}); \Gamma(\overline{B1}_j^s, \overline{S}_l^s),$$

устанавливающие взаимосвязь между порождающим словом  $\overline{W}_r^s$ , парой частей речи порождающего и порожденного слова  $\overline{H1}_i$ , вектором аффиксов  $\overline{B1}_j^s$ , основой порождающего слова  $\overline{L}_q^s$  и ее морфологическим вариантом  $\overline{L}_q^s$ , словообразовательным аффиксом  $\overline{B}_m^{sk}$ , порожденным словом  $\overline{S}_l^s$ ,  $s \in \overline{1,2}$ ,  $k \in \overline{1,5}$ .

## 8. Правила анализа слов

В пунктах 1) – 6), согласно разработанным правилам, были формализованы, ранжированы и упорядочены элементы анализа слов, которые исполь-

зуються при описанні об'єктів производственной системи и действий над ними.

В пункте 7 были разработаны правила, устанавливающие взаимосвязи между этими элементами. На основании результатов пунктов 1 – 7 производится анализ слов.

Правила анализа слов разработаны в соответствии с морфологическими правилами русского языка [4 – 6] и предусматривают конструирование слова из основы и набора словообразовательных аффиксов.

Анализ слова происходит в соответствии с парой частей речи порождающего и порожденного слова и набором словообразовательных аффиксов, при этом порожденные слова, основы и аффиксы на письме представляются последовательностью букв или фонем.

С алгебраической точки зрения анализ слов производится следующим образом.

Введем матрицы информационных мер:

– порожденных слов  $M(\overline{S2}^s)$  :

$$\begin{aligned}
 &M(\overline{S2}^s) = \| M(\overline{S2}_{rj}^s) \|; \\
 &M(\overline{W}_{r-1}^s) < M(\overline{W}_r^s) < M(\overline{W}_{r+1}^s); \\
 &M(\overline{B1}_{j-1}^s) < M(\overline{B1}_j^s) < M(\overline{B1}_{j+1}^s); \quad (38) \\
 &\forall \overline{W}_r^s \wedge \overline{B1}_j^s \wedge \overline{S1}_l^s ; \\
 &\exists M(\overline{S2}_{rj}^s) = \begin{cases} M(\overline{S1}_l^s), & \begin{aligned} &\Gamma(\overline{H1}_i, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{W}_r^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{S1}_l^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{S1}_l^s) = 1 \wedge \\ &\Gamma(\overline{B1}_j^s, \overline{S1}_l^s) = 1 \end{aligned} \\ 0, & \text{иначе} \end{cases}
 \end{aligned}$$

– основ порождающих слов  $M(\overline{\Lambda 2}^s)$  :

$$M(\overline{\Lambda 2}^s) = \| M(\overline{\Lambda 2}_{rj}^s) \|;$$

$$M(\overline{W}_{r-1}^s) < M(\overline{W}_r^s) < M(\overline{W}_{r+1}^s);$$

$$M(\overline{B1}_{j-1}^s) < M(\overline{B1}_j^s) < M(\overline{B1}_{j+1}^s); \quad (39)$$

$$\forall \overline{W}_r^s \wedge \overline{B1}_j^s \wedge \overline{L}_q^s ;$$

$$\exists M(\overline{L2}_{rj}^s) = \begin{cases} M(\overline{L}_q^s), & \begin{aligned} &\Gamma(\overline{H1}_i, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{W}_r^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{L}_q^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{L}_q^s) = 1 \wedge \\ &\Gamma(\overline{B1}_j^s, \overline{L}_q^s) = 1 \end{aligned} \\ 0, & \text{иначе} \end{cases}$$

– морфонологических вариантов основ порождающих слов  $M(\overline{\tilde{L}2}^s)$  :

$$M(\overline{\tilde{L}2}^s) = \| M(\overline{\tilde{L}2}_{rj}^s) \|;$$

$$M(\overline{W}_{r-1}^s) < M(\overline{W}_r^s) < M(\overline{W}_{r+1}^s);$$

$$M(\overline{B1}_{j-1}^s) < M(\overline{B1}_j^s) < M(\overline{B1}_{j+1}^s); \quad (40)$$

$$\forall \overline{W}_r^s \wedge \overline{B1}_j^s \wedge \overline{\tilde{L}}_q^s ;$$

$$\exists M(\overline{\tilde{L}2}_{rj}^s) = \begin{cases} M(\overline{\tilde{L}}_q^s), & \begin{aligned} &\Gamma(\overline{H1}_i, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{W}_r^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{\tilde{L}}_q^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{\tilde{L}}_q^s) = 1 \wedge \\ &\Gamma(\overline{B1}_j^s, \overline{\tilde{L}}_q^s) = 1 \end{aligned} \\ 0, & \text{иначе} \end{cases}$$

– словообразовательных аффиксов  $M(\overline{B2}^{sk})$  :

$$M(\overline{B2}^{sk}) = \| M(\overline{B2}_{rj}^{sk}) \|;$$

$$M(\overline{W}_{r-1}^s) < M(\overline{W}_r^s) < M(\overline{W}_{r+1}^s);$$

$$M(\overline{B1}_{j-1}^s) < M(\overline{B1}_j^s) < M(\overline{B1}_{j+1}^s); \quad (41)$$

$$\forall \overline{W}_r^s \wedge \overline{B1}_j^s \wedge \overline{B}_m^{sk} ;$$

$$\exists M(\overline{B2}_{rj}^{sk}) = \begin{cases} M(\overline{B}_m^{sk}), & \begin{aligned} &\Gamma(\overline{H1}_i, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{W}_r^s) = 1 \wedge \\ &\Gamma(\overline{H1}_i, \overline{B}_m^{sk}) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{B1}_j^s) = 1 \wedge \\ &\Gamma(\overline{B1}_j^s, \overline{B}_m^{sk}) = 1 \wedge \\ &\Gamma(\overline{W}_r^s, \overline{B}_m^{sk}) = 1 \end{aligned} \\ 0, & \text{иначе} \end{cases}$$

Используя матрицы (38) – (41), осуществим нейросетевой анализ слов. На рис.1 представлена трехслойная неполносвязная с прямыми связями нейронная сеть анализа слов.

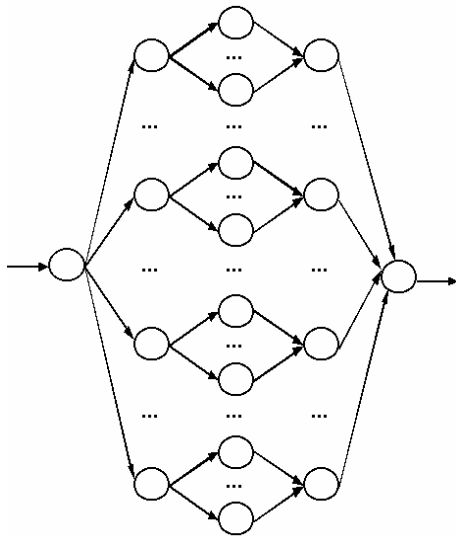


Рис. 1. Нейронная сеть анализа слов

Нулевой (входной) слой содержит один нейрон и осуществляет анализ входного вектора  $X$ , состоящего из следующих элементов – парой частей речи порождающего и порожденного слова ( $X_1$ ), вектора словообразовательных аффиксов в буквенном ( $X_2$ ), и фонемном ( $X_3$ ) представлении, анализируемого слова в буквенном ( $X_4$ ) и фонемном ( $X_5$ ) представлении. Результатом анализа является установление соответствия:

$$X_1 = M(\overline{H1}_i); X_2 = M(\overline{B1}_j^1); X_3 = M(\overline{B1}_j^2);$$

$$X_4 = M(\overline{S1}_i^1); X_5 = M(\overline{S1}_i^2).$$

Первый слой состоит из  $\eta(\overline{W1})$  групп, содержащих в себе  $\eta(\overline{B1})$  нейронов. Каждый из нейронов, согласно паре частей речи  $X_1$  и вектора словообразовательных аффиксов  $X_2 / X_3$ , выполняет выделение морфонологического варианта основы  $M(\overline{L2}_{rj}^1) / M(\overline{L2}_{rj}^2)$  из входного слова  $X_4 / X_5$ , посредством усечения префикса  $M(\overline{B2}_{rj}^{11}) / M(\overline{B2}_{rj}^{21})$ , интерфикса  $M(\overline{B2}_{rj}^{12}) / M(\overline{B2}_{rj}^{22})$ , суффикса  $M(\overline{B2}_{rj}^{13}) / M(\overline{B2}_{rj}^{23})$ , флексии  $M(\overline{B2}_{rj}^{14}) / M(\overline{B2}_{rj}^{24})$  и постфикса  $M(\overline{B2}_{rj}^{15}) / M(\overline{B2}_{rj}^{25})$  в виде:

$$M(\overline{L2}_{rj}^1) = (M(\overline{B2}_{rj}^{11}) \bullet M(\overline{W2}_{rj}^1)) \circ$$

$$\circ (M(\overline{B2}_{rj}^{13}) \diamond M(\overline{B2}_{rj}^{14}) \diamond M(\overline{B2}_{rj}^{15})); \quad (41)$$

$$M(\overline{L2}_{rj}^2) = (M(\overline{B2}_{rj}^{21}) \bullet M(\overline{W2}_{rj}^2)) \circ$$

$$\circ (M(\overline{B2}_{rj}^{23}) \diamond M(\overline{B2}_{rj}^{24}) \diamond M(\overline{B2}_{rj}^{25})), \quad (42)$$

где  $\diamond$  – операция конкатенации;

$\circ$  – операция усечения справа;

$\bullet$  – операция усечения слева.

Второй слой состоит из  $\eta(\overline{W1})$  групп, содержащих в себе  $\eta(\overline{B1})$  нейронов. Каждый из нейронов, согласно входному слову  $X_4 / X_5$ , заменяет выделенный на первом слое морфонологический вариант основы  $M(\overline{L2}_{rj}^1) / M(\overline{L2}_{rj}^2)$  исходной основой  $M(\overline{L2}_{rj}^1) / M(\overline{L2}_{rj}^2)$ , в виде

$$M(\overline{L2}_{rj}^1) = F(M(\overline{L2}_{rj}^1)), \quad (43)$$

где функция  $F$  отображает  $M(\overline{L2}_{rj}^1)$  в  $M(\overline{L2}_{rj}^1)$ ;

$$M(\overline{L2}_{rj}^2) = F(M(\overline{L2}_{rj}^2)), \quad (44)$$

где функция  $F$  отображает  $M(\overline{L2}_{rj}^2)$  в  $M(\overline{L2}_{rj}^2)$

Третий слой состоит из  $\eta(\overline{W1})$  групп, содержащих в себе  $\eta(\overline{B1})$  нейронов. Каждый из нейронов,



согласно паре частей речи  $X_1$ , выполняет конкатенацию полученной на втором слое основы

$$M(\overline{L2_{rj}}^1) / M(\overline{L2_{rj}}^2)$$

с флексией  $M(\overline{B2_{rj}}^{14}) / M(\overline{B2_{rj}}^{24})$  и постфиксом  $M(\overline{B2_{rj}}^{15}) / M(\overline{B2_{rj}}^{25})$  в виде:

$$M(\overline{W2_{rj}}^1) = M(\overline{L2_{rj}}^1) \diamond M(\overline{B2_{rj}}^{14}) \diamond M(\overline{B2_{rj}}^{15}); \quad (45)$$

$$M(\overline{W2_{rj}}^2) = M(\overline{L2_{rj}}^2) \diamond M(\overline{B2_{rj}}^{24}) \diamond M(\overline{B2_{rj}}^{25}). \quad (46)$$

Четвертый (выходной) слой содержит один нейрон и осуществляет анализ выходного вектора  $Y$ , состоящего из следующих элементов – слова в буквенном ( $Y_1$ ) и фонемном ( $Y_2$ ) представлении.

### Выводы

**Новизна.** В данной работе разработаны правила формализации и численного исследования частей речи, наборов морфологических признаков, букв и фонем, порождающих слов, основ порождающих слов и их морфонологических вариантов, словообразовательных аффиксов, порожденных слов, закрепления взаимосвязей между элементами анализа слов, анализа слов. На основании этих правил предложена нейросетевая методика анализа слов, входящих в описание объектов и действий команд производственной системы.

**Практическое значение.** Основные положения работы могут быть реализованы в интеллектуальной системе в виде алгоритмов, обеспечивающих общение с пользователем на естественном языке.

### Литература

1. Бутковский А.Г., Полтавский Л.Н. Фinitное управление системами с распределенными параметрами // Автоматика и телемеханика. – 1969. – № 4. – С. 23-33.
2. Бутковский А.Г., Полтавский Л.Н. Фinitное управление линейными системами с средоточенными параметрами // Автоматика и телемеханика. – 1967. – № 9. – С. 44-58.
3. Интеллектуальное управление динамическими системами / С.Н. Васильев, А.К. Жерлов, Е.А.Федосов, Б.Е. Федун. – М.: Физико-математическая литература, 2000. – 352 с.
4. Криводубский О.А., Федоров Е.Е. Формальное представление русского языка и речи // Искусственный интеллект. – 2003. – № 4. – С. 402-410.
5. Современный русский язык: Учеб. для филол. спец. высших учебных заведений / В.А. Белошапкова, Е.А.Брызгунова, Е.А.Земская и др.; Под ред. В.А. Белошапковой. – М.: Азбуковник, 1997. – 928 с.
6. Русская грамматика: В 2т. – М.: Наука, 1980. – Т.1: Фонетика. Словообразование. Морфология. – 784 с.

Поступила в редакцию 24.05.2006

**Рецензент:** д-р техн. наук, проф. Н.И. Чичикало, Донецкий национальный технический университет, Донецк.