



Серия «Математика»

2017. Т. 22. С. 31–49

Онлайн-доступ к журналу:

<http://mathizv.isu.ru>

---

---

ИЗВЕСТИЯ

Иркутского  
государственного  
университета

---

---

УДК 004.93

MSC 68T27

DOI <https://doi.org/10.26516/1997-7670.2017.22.31>

## Прозрачное глубокое обучение на основе вероятностных формальных понятий в задаче обработки естественного языка \*

Е. Е. Витяев, В. В. Мартынович

*Институт математики им. С. Л. Соболева*

**Аннотация.** Несмотря на высокую эффективность методов глубокого обучения (Deep Learning) они остаются «вещью в себе», «черным ящиком» решениям которого невозможно доверять. Это критично для таких областей, как медицина, финансовые вложения, военные применения и другие, где цена ошибки слишком высока. В связи с этим Европейский союз собирается потребовать в 2018 году от компаний, чтобы они давали пользователям объяснения решений, получаемых автоматическими системами.

В данной работе предлагается альтернативный, логико-вероятностный метод глубокого обучения, способный объяснять свои решения. Это метод иерархической кластеризации, основанный на оригинальном логико-вероятностном обобщении формальных понятий (ВФП [13]). Для сравнения с глубоким обучением, основанном на нейронных сетях, была выбрана работа [12], в которой решается задача обработки естественного языка на наборе данных *UMLS*. Для применения логико-вероятностного обобщения формальных понятий определяется алгоритм классификации, основанный на энергии противоречий Energy Learning [10]. Логико-вероятностные формальные понятия определяются через неподвижные точки, как и сами формальные понятия, только в качестве правил используются определенные вероятностные правила. Энергия противоречий позволяет разрешать противоречия, возникающие в неподвижных точках, формирующих вероятностные формальные понятия. Показано, что данный алгоритм кластеризации не уступает по точности методу Deep Learning [12], тем не менее, получаемые им решения объясняются совокупностью вероятностных правил неподвижных точек.

**Ключевые слова:** вероятность, формальные понятия, обработка естественного языка, Deep Learning, Data Mining, семантическая энергия.

---

\* Работа выполнена при поддержке Российского научного фонда (грант РНФ 17-11-01176)

## 1. Введение

Целью данного исследования является демонстрация возможностей семантико-вероятностного подхода к анализу данных. В качестве задачи рассматривается предсказание связей в мультиграфе, представляющем семантические лингвистические данные. Для её решения в работе разрабатывается алгоритм предсказания связей на основе вероятностных формальных понятий [13]. Показывается, что разработанный алгоритм решает задачу не менее успешно, чем метод Deep Learning, приведенный в [5], который представляет собой современный подход к решению задачи предсказания семантических связей (Link Prediction) на данных с семантикой естественного языка.

В анализе формальных понятий [7] классические формальные понятия выступают в качестве классификационных единиц. В разделе 3 показано, как модифицировать формальные понятия для работы с выборками из генеральной совокупности, когда формальные понятия имеют вероятностный характер и их определение через неподвижные точки и соответствующие вероятностные импликации становится логически противоречивым. Для определения вероятностных формальных понятий в условиях логических противоречий мы используем функционал энергии, с целью минимизации количества противоречий.

Для сравнения предложенного метода классификации с методами Deep Learning выбран набор данных *UMLS* [12] из работы [5]. В серии исследований [3–5] проводится анализ этих и других данных в рамках парадигмы глубинного обучения и в частности в рамках схемы многозадачного обучения. Представив исходный мультиграф, задающий *UMLS*, в виде формального контекста, мы в разделе 4 исследуем возможность решения поставленной в [5] задачи предсказания связей (Link prediction) с позиции анализа формальных понятий. Объекты дополняются до обобщенных формальных понятий посредством описанного в разделе 3 алгоритма *StatClosure*, и далее вывод о наличии связи в исходном мультиграфе делается на основе дополненного образа объекта. По сравнению с методами нейронных сетей, предсказание в нашем случае получается явным и прозрачным, и метод не только дает некоторое решение, но и способен дать ответ на вопрос “почему”.

Практические результаты решения задачи предсказания связей приведены в разделе 5. Анализ качества предсказания выполнен с различных точек зрения: с использованием критерия *AUC* (Area Under Curve), а также классического критерия *Accuracy* (точности предсказания). Кроме того, обсуждаются существенное снижение сложности обучения, поскольку метод ВФП автоматически извлекает семантику из предоставленных данных и для её построения, в отличие от [3; 5], не требуется процедуры предварительного обучения на смежных задачах и привлечения дополнительных больших массивов данных.

## 2. Вероятностные формальные понятия

В анализе формальных понятий (далее по тексту — АФП) [7] данные рассматриваются как совокупность наборов объектов и атрибутов, соединенных с помощью отношения принадлежности.

**Определение 1.** *Формальный контекст  $K$  — это тройка  $(G, M, I)$ , где  $G$  и  $M$  — произвольные множества объектов и атрибутов, и  $I \subseteq G \times M$  — отношение принадлежности атрибутов объектам.*

**Определение 2.** *Пусть  $A \subseteq G, B \subseteq M$ . Тогда:*

- 1)  $A^\uparrow = \{m \in M \mid \forall g \in A, (g, m) \in I\};$
- 2)  $B^\downarrow = \{g \in G \mid \forall m \in B, (g, m) \in I\};$
- 3)  $(A, B)$  — формальное понятие, если  $A^\uparrow = B, B^\downarrow = A$ ;

**Определение 3.**  $R = (B, C)$  — импликация  $R \in \text{Imp}(K)$ , если  $B' \subseteq C'$ . При этом,  $B = R^\leftarrow$  называем посылкой, а  $B = R^\rightarrow$  — заключением импликации. Оператор логического вывода на основе множества импликаций  $\mathcal{R}$  действует путем добавления к исходному множеству атрибутов  $L$  других, выводимых из  $L$  атрибутов:

$$\Pi_{\mathcal{R}}(L) = \{m \in M \mid B \subseteq L, (B \rightarrow m) \in \mathcal{R}\}$$

Ключевой является теорема 1 [7], задающая альтернативное определение формальных понятий как неподвижных точек оператора логического вывода. Это позволяет перейти к их вероятностному обобщению.

**Теорема 1.** *Для любого множества  $B \subseteq M$ ,  $B$  будет формальным понятием (в понимании Data Mining — классом)  $\Leftrightarrow \Pi_{\text{Imp}(K)}(B) = B$ .*

Вероятностные формальные понятия мы получим как неподвижные точки соответствующего вероятностного оператора логического вывода. Для его определения построим логико-вероятностную модель, описывающую формальный контекст.

**Определение 4.** *Для конечного контекста  $K = (G, M, I)$  определим сигнатуру  $\sigma_K$ , состоящую только из множества предикатных символов  $M$ . Для сигнатуры  $\sigma_K$  определим модель контекста  $K$  с носителем  $G$  таким образом, чтобы  $K \models m(x) \Leftrightarrow (x, m) \in I$ . Факт истинности формулы  $\phi$  на модели  $K$ , суженной до объекта  $g$ , записываем как  $g \models \phi \Leftrightarrow K_g \models \phi$ .*

**Определение 5.** *В сигнатуре контекста определены следующие логические конструкции:*

- 1)  $\text{Term}(K)$  — множество термов, состоящее из символов переменных;
- 2)  $\text{At}(K)$  — атомами являются выражения  $m(t)$ , где  $m \in \sigma_K$  и  $t \in \text{Term}(K)$ ;
- 3)  $L(K)$  — литеры включают все атомы  $m(t)$  и отрицания  $\neg m(t)$ ;
- 4)  $\Phi(K)$  — определяется индуктивно: всякий атом — формула, и для любых  $\Phi, \Psi \in \Phi_K$  синтаксические конструкции  $\Phi \wedge \Psi, \Phi \vee \Psi, \Phi \rightarrow \Psi, \neg \Phi$  тоже формулы.

Истинность формул на объектах контекста определяется естественным образом. Приведенное обобщение позволяет строить логические высказывания об атрибутных зависимостях с использованием отрицаний. Для обобщения формальных понятий также потребуется некоторый синтез логики и вероятности на произвольном контексте  $K$ .

**Определение 6.** Рассмотрим произвольную вероятностную меру  $\mu$  на множестве  $G$ , определенную в Колмогоровском смысле. Тогда контекстная вероятностная мера на множестве формул:

$$\nu : \Phi_K \rightarrow [0, 1], \quad \nu(\phi) = \mu(\{g \mid g \models \phi\}).$$

**Определение 7.** Пусть  $C, H_i \in L_K$ ,  $C \notin \{H_1, H_2, \dots, H_k\}$ ,  $k \geq 0$ . Тогда:

- 1) **Правило**  $R = (H_1, H_2, \dots, H_k \rightarrow C)$  есть импликация  $(H_1 \wedge H_2 \dots \wedge H_k \rightarrow C)$ ;
- 2) **Посылкой**  $R^{\leftarrow}$  правила  $R$  называется набор литер  $\{H_1, H_2, \dots, H_k\}$ ;
- 3) **Заключением** правила является  $R^{\rightarrow} = C$ ;
- 4) **Длиной правила** мы называем мощность его посылки  $|R^{\leftarrow}|$ ;
- 5) Если  $R_1^{\leftarrow} = R_2^{\leftarrow}$  и  $R_1^{\rightarrow} = R_2^{\rightarrow}$ , тогда  $R_1 = R_2$ .
- 6) **Вероятностью** правила  $R$  является значение

$$\eta(R) = \nu(R^{\rightarrow} | R^{\leftarrow}) = \frac{\nu(R^{\leftarrow} \wedge R^{\rightarrow})}{\nu(R^{\leftarrow})}$$

Если знаменатель  $\nu(R^{\leftarrow})$  нулевой, вероятность правила остается неопределенной.

При определенных условиях выводы, полученные с помощью вероятностных правил из определения 7, будут непротиворечивы. Непротиворечивые неподвижные точки составляют основу метода вероятностных формальных понятий [13].

**Определение 8.** Правило  $R$  назовем максимально специфичным  $R \in \text{MSR}(K)$ , если нет правила  $\tilde{R}$  с более длинной посылкой  $R^{\leftarrow}$  и более высокой вероятностью  $\eta(\tilde{R}) > \eta(R)$ .

Правила из определения 7 позволяют определить вероятностный оператор замыкания. Для этого заменим множество импликаций  $\mathcal{R}$  на какое-либо подмножество максимально специфических вероятностных правил. Поэтому ниже будем предполагать, что  $\mathcal{R}$  – подмножество максимально специфичных правил. По аналогии с теоремой 1, определим вероятностные формальные понятия как неподвижные точки оператора логического вывода, использующего множество правил  $\mathcal{R}$ .

**Определение 9.** Замыканием  $\bar{L}$  множества литер  $L$  будем называть наименьшую неподвижную точку оператора логического вывода, содержащую  $L$ :

$$\bar{L} = \Pi_{\mathcal{R}}(\bar{L}) = \Pi_{\mathcal{R}}^{\infty}(L) = \bigcup_{k \in \mathbb{N}} \Pi_{\mathcal{R}}^k(L)$$

В [13] показано, что процедура замыкания позволяет сохранить свойство непротиворечивости. Поэтому классы, определяемые с помощью последовательного дополнения исходного непротиворечивого множества литер новыми логическими выводами, также будут непротиворечивыми. Такие целостные классы мы называем вероятностными формальными понятиями.

**Теорема 2.** Пусть  $\mathcal{R}$  – множество максимально специфических правил; тогда если  $B$  непротиворечиво, то  $\Pi_{\mathcal{R}}(B)$  – также непротиворечиво.

**Определение 10.** Пусть  $\mathcal{R} \subseteq \text{MSR}(K)$  – множество максимально специфических правил. Тогда  $B$  – вероятностное формальное понятие, если  $\Pi_{\mathcal{R}}(B) = B$ .

На основе определения 9 нетрудно предложить алгоритм замыкания ProbClosure, который строит замыкание  $\bar{L} = \text{ProbClosure}(L)$ .  $\bar{L}$  будет минимальной неподвижной точкой, содержащей множество  $L$ , и, согласно определению 10, вероятностным формальным понятием.

ProbClosure успешно решает большинство теоретических задач, не требуя разрешения противоречий, так как в силу теоремы 2 исключается ситуация, когда в процессе вывода обнаруживается одновременно литера и ее отрицание.

---

**Алгоритм 1** Простейший классификатор *ProbClosure*


---

**Вход:**  $\mathcal{R} \subseteq \text{For}(K)$ ,  $K = (G, M, I)$ ,  $B \subseteq \text{Lit}(K)$

**Выход:**  $\overline{B} \subseteq \text{Lit}(K)$  — вероятностное формальное понятие

1: **Функция**  $\text{PROB\_CLOSURE}(K, \mathcal{R}, B)$

2:      $C \leftarrow B$

3:     **Повторять**

4:          $B \leftarrow C$

5:         **Для всех**  $R \in \mathcal{R}$  **выполнять**

6:             **Если**  $R^{\leftarrow} \subseteq C$  **тогда**

7:                  $C \leftarrow C \cup \{R^{\rightarrow}\}$

8:             **Конец условия**

9:     **Конец цикла**

10:     **До тех пор пока**  $C \setminus B \neq \emptyset$

11:     **Вернуть**  $C$

12: **Конец функции**

---

### 3. Энергия противоречий и разрешение противоречий

В задачах машинного обучения [9] данными являются выборки из некоторой генеральной совокупности, которой в нашем случае является контекст. Тогда выборку и данные можно задать следующим образом:

- 1) Источник данных  $e$  — многомерная случайная величина с заданным распределением;
- 2) Обучающая выборка  $G_{\text{teach}} = \{g_{(1)}, \dots, g_{(n)}\}$  — множество значений этой случайной величины.

Для нас это означает, что моделью наблюдаемого контекста  $K = (G_{\text{teach}}, M, I)$  является выборка из генеральной совокупности  $K^*$ , где каждый  $g \in G_{\text{teach}}$  представлен многомерной бернуллиевской случайной величиной. Однако задача анализа данных должна по-прежнему пониматься в смысле исходного контекста  $K^*$ , образующего генеральную совокупность объектов. В таких условиях непротиворечивость логического вывода с помощью  $\Pi_{\mathcal{R}}$  может быть нарушена, поскольку теперь уже нельзя утверждать о максимальной специфичности правил  $\mathcal{R}$ , извлеченных из  $G_{\text{teach}}$  по отношению к исходному контексту  $K^*$ . Для построения логико-вероятностного замыкания в этом случае требуется видоизменить процедуру замыкания *ProbClosure*.

В работе [5] рассматривается возможность применения методов Deep Learning к задаче анализа данных *UMLS* [12]. Для работы с такими данными необходимо учитывать семантику, заложенную в естественном языке. Серия исследований [5; 10] показала, что в качестве функции,

генерирующей обратную связь, может быть использована так называемая “семантическая функция энергии”. Из физики мы знаем, что оценка различных конфигураций переменных (состояний системы) на основании функционала энергии позволяет найти точки равновесия системы или предсказать динамику её развития.

Функционал энергии может давать четкий критерий оценки для множества практических задач. В нашем случае концепция энергии позволяет отделить наиболее непротиворечивые конфигурации при переходе от контекста  $K^*$  к выборке из генеральной совокупности.

Рассмотрим процесс преобразования набора литер. Пусть исходное множество литер  $B = B_1$  проходит через цепочку преобразований  $B_1, \dots, B_n$  (такие преобразование происходят со стартовым множеством  $B$  в алгоритме *ProbClosure*). Условимся, что для алгоритма преобразования существует некий критерий  $\phi$ , минимизация которого определяет направление поиска в пространстве подмножеств литер  $B \in 2^{\text{Lit}(K)}$ . Такое семейство алгоритмов очень удобно в вычислительном смысле, поскольку позволяет определить их итеративно и исходную задачу свести к задаче минимизации. Последовательности, в которых первый и последний набор совпадают (т. е.  $B_1 = B_n$ ), очевидно, порождают конечное тождественное преобразование, поэтому для критерия  $\phi$  должно выполняться следующее условие:

**Определение 11** (Условие потенциальности).

$$B_1 = B_n \Rightarrow \sum_{i=1 \dots n-1} \phi(B_i, B_{i+1}) = 0$$

Функционал  $\phi$  является аналогом физического потенциала, за исключением того, что в нашем случае он дискретный. Условие в определении 11 — в точности условие независимости потенциала от пути его вычисления и, как известно из физической теории, потенциал позволяет определить функцию энергии. Заметим, что идея введения функционала энергии не нова, в [10] она подробно изучена в контексте механизма обратной связи для глубинных нейронных сетей.

**Теорема 3.** *Критерий  $\phi$  может быть выражен с помощью потенциальной энергии  $\mathcal{E}$ :  $\phi(B, C) = \mathcal{E}(C) - \mathcal{E}(B)$ ; при этом  $\phi(B, C)$  удовлетворяет условию потенциальности, а значение потенциала не зависит от точки начала отсчета энергии.*

Ключевым является выбор некоторого функционала энергии  $\mathcal{E}$ . Для нас важно минимизировать количество противоречий, содержащихся в некоторой конфигурации литер  $B$ .

Зафиксируем некоторое множество правил  $\mathcal{R}$ . Далее будем считать, что все правила  $R$  берутся из этого универсума правил  $\mathcal{R}$ .

**Определение 12.** Пусть  $R$  — правило, а  $B \subset \text{Lit}(K)$ .

$R$  **применимо** (или  $R \in \text{App}(B)$ ) к набору литер  $B$ , если  $R^{\leftarrow} \subset B$ .

$R$  **подтверждается** (или  $R \in \text{Sat}(B)$ ) на наборе литер  $B$ , если  $R \in \text{App}(B)$ , и при этом  $R^{\rightarrow} \in B$ .

$R$  **опровергается** (или  $R \in \text{Fal}(B)$ ) на наборе литер  $B$ , если  $R \in \text{App}(B)$ , и при этом  $\neg R^{\rightarrow} \in B$ .

**Определение 13.** Энергией противоречий мы называем функционал энергии, определенный с помощью веса опровергающихся правил, за вычетом энергии подтверждающихся правил:

$$\mathcal{E}(B) = \sum_{R \in \text{Fal}(B)} \gamma(R) - \sum_{R \in \text{Sat}(B)} \gamma(R)$$

$$\gamma(R): R \rightarrow \mathbb{R}^+$$

При этом  $\phi(B, \emptyset) = \mathcal{E}(B)$ .

Задача состоит в том, чтобы минимизировать энергию противоречий  $\mathcal{E}(B) \rightarrow \min$ , и таким образом найти максимально непротиворечивые комбинации литер. Однако, полное решение задачи минимизации функционала энергии выглядит как полный перебор в пространстве конфигураций литер  $2^{\text{Lit}(K)}$ .

Вспомним, что мы работаем в вероятностном контексте, где приемлемость решения определяется иными способами (например, предсказательной точностью классификатора *Accuracy*). Поэтому поставленная вычислительная проблема может быть решена субоптимальным образом. Предлагается вычислять приближенные решения посредством “жадного” итеративного алгоритма *StatClosure*, который минимизирует потенциал и выполняет поиск локально оптимальных решений. Свойство жадности опирается на то предположение, что для субоптимальности достаточно рассмотреть только потенциал при переходе к ближайшим соседям, т. е. от конфигурации  $B$  к конфигурациям вида  $B \pm l$ , где  $l \in \text{Lit}(K)$ .

Это приводит к алгоритму 2, в котором неподвижные точки вероятностного оператора локально минимизируют функционал энергии противоречий и поэтому в известной степени решают задачу непротиворечивого предсказания. Алгоритм *StatClosure* может быть успешно применен к задачам классификации, в частности рассматриваемой нами задаче типа Link Prediction. Общая схема заключается в следующем:

- 1) Набор данных задается в виде обучающего и контрольного формальных контекстов  $K_t = (G_t, M, I_t)$ ,  $K_c = (G_c, M, I_c)$ ;
- 2) Извлекается множество вероятностных законов [1]  $\mathcal{R}$ , которое проверяется статистическими методами. Полностью процедура извлечения вероятностных законов описана в [1];



**Алгоритм 2** Замыкание статистическим оператором вывода**Вход:**  $\mathcal{R} \subseteq \text{For}(K)$ ,  $K = (G, M, I)$ ,  $B \subseteq \text{Lit}(K)$ **Выход:**  $C \subseteq \text{Lit}(K)$  — статистическое формальное понятие

---

```

1:  Функция STATCLOSURE( $K, \mathcal{R}, B$ )
2:     $B_0 \leftarrow B$ 
3:     $k \leftarrow 0$ 
4:    Повторять
5:       $k \leftarrow k + 1$ 
6:       $B_k \leftarrow B_{k-1}$ 
7:       $\psi \leftarrow 0$ 
8:      Candidates  $\leftarrow \emptyset$ 
9:      Для всех  $L \in \text{Lit}(K) \setminus (B_{k-1} \cup \neg B_{k-1})$  выполнять
10:        Candidates  $\leftarrow \text{Candidates} \cup \{B_{k-1} \cup L\}$ 
11:      Конец цикла
12:      Для всех  $L \in B_{k-1}$  выполнять
13:        Candidates  $\leftarrow \text{Candidates} \cup \{B_{k-1} \setminus L\}$ 
14:      Конец цикла
15:      Для всех  $C \in \text{Candidates}$  выполнять
16:         $\alpha \leftarrow \phi(B_{k-1}, C)$ 
17:        Если  $\alpha < \psi$  тогда
18:           $\psi \leftarrow \alpha$ 
19:           $B_k \leftarrow C$ 
20:        Конец условия
21:      Конец цикла
22:      До тех пор пока  $\psi < 0$ 
23:      Вернуть  $B_k$ 
24: Конец функции

```

---

- 3) Выбирается множество контрольных пар  $(g \in G_c, m \in M)$ , относительно которых ставится задача предсказания  $(g, m) \in I_c$ ;
- 4) Для объекта  $g$  строится его описатель  $L = g' \setminus \{m\}$ ;
- 5) Вычисляется статистическое замыкание описателя  $C = \text{StatClosure}(L)$ ;
- 6) Если  $m \in C$ , то дается положительный ответ, иначе выдается отрицательный ответ о вхождении  $(g, m)$  в контрольный контекст.

Для придания подходу гибкости, в частности для расчета показателя  $AUC$  [11], необходимо параметризовать алгоритм, включив возможность регулирования вероятности ложнопозитивных или ложнонегативных срабатываний. Мы можем увеличивать вес противоречащих

добавлению литер правил, и останавливать добавление литер, которые являются более противоречивыми, чем этого допускает множество закономерностей  $\mathcal{R}$ :

$$\mathcal{E}_w(B) = w \cdot \sum_{\text{Fal}(B)} \gamma(R) - \sum_{\text{Sat}(B)} \gamma(R).$$

Параметр  $w$  из переопределения функционала энергии назовем **весом противоречий**. В разделе 5 с описанием эксперимента более подробно исследована зависимость точности предсказаний для позитивных и негативных образцов от значения параметра веса противоречий.

#### 4. Эксперимент на данных *UMLS*

В работе [5] рассматривается задача предсказания связей (Link Prediction) на семантическом мультиграфе  $X$ . Для решения этой и других связанных с ней задач методом Deep Learning авторами проведена большая работа, включающая серию исследований. В ранних работах [3] предлагается архитектура нейронной сети, обеспечивающая промежуточное представление терминов естественного языка в виде векторов вещественнозначного признакового пространства  $\mathbb{R}^{50}$ . При этом, используется мера схожести:

$$f(e_i^l, r_i, e_i^r) = \|R_{r_i}^{lhs} Ev(e_i^l) - R_{r_i}^{rhs} Ev(e_i^r)\|$$

Тензоры  $R_r$  выражают промежуточное представление отношений между терминами,  $v(\cdot)$  выбирает порядковый номер  $n$  термина-аргумента и задает вектор, дающий при умножении на него строку матрицы с номером  $n$ , а матрица  $E$  построчно задает вложение терминов в признаковое пространство. Обучение состоит в минимизации ошибки посредством градиентного спуска при переходе от истинного ребра мультиграфа  $\in X$  к ложному ребру  $\notin X$ :

$$f(e_i^l, r_i, e_i^r) < f(e_j^l, r_i, e_i^r), \forall j: (e_j^l, r_i, e_i^r) \notin X$$

$$f(e_i^l, r_i, e_i^r) < f(e_i^l, r_i, e_j^r), \forall j: (e_i^l, r_i, e_j^r) \notin X$$

Характерным здесь является сложность и комплексность процесса обучения. Поскольку матрица  $E$  едина для всех отношений, использующихся в семантическом мультиграфе, применяется метод мультизадачного обучения (multi-task learning), в котором одна и та же нейронная сеть применяется к различным задачам. При этом объем исследуемых данных даже на этапе первичной оптимизации (при построении матриц вложения) составляет сотни тысяч семантических связей. Применялись

для обучения достаточно распространенные базы знаний WordNet и Freebase.

Следующим шагом в развитии рассматриваемой задачи являлась работа [4], где уже знакомый нам коллектив авторов предлагает метод обработки связного текста за счет введения дополнительных уровней абстракции, и, соответственно, новых слоев архитектуры глубокой сети:

- 1) Множества синтаксических единиц предложения  $x$  разбиваются на три множества (условно можно их обозначить как подлежащее, сказуемое и дополнение), каждая синтаксическая единица заменяется на её векторное представление. Затем множество векторных представлений комбинируется с помощью специальной функции  $\pi$ :

$$E_{lhs} = \pi(E_{lhs_1}, E_{lhs_2}, \dots)$$

$$E_{rel} = \pi(E_{lhs_1}, E_{lhs_2}, \dots)$$

$$E_{rhs} = \pi(E_{lhs_1}, E_{lhs_2}, \dots)$$

- 2) Пары  $(E_{lhs}, E_{rel})$  и  $(E_{rel}, E_{rhs})$  комбинируются в ещё одно промежуточное представление с помощью специальных функций:

$$E_{lhs(rel)} = g_{left}(E_{lhs}, E_{rel})$$

$$E_{rhs(rel)} = g_{right}(E_{rhs}, E_{rel})$$

Функции  $g_{left}, g_{right}$  — параметризуемы, и конкретный параметр выбирается посредством процедуры обучения;

- 3) Наконец, заключительное преобразование состоит в применении ещё одной параметрической функции  $\mathcal{E}(x) = h(E_{lhs(rel)}, E_{rhs(rel)})$ . Так получается семантическая энергия исходного высказывания.

Обучение происходит аналогичным образом: параметры функций подбираются так, чтобы функционал энергии на обучающей выборке получал возможно минимальные значения. Для решения этой задачи также потребовалось увеличение объема данных: помимо упомянутого WordNet использовались ConceptNet, Wikipedia и производные от них базы знаний, построенные специальным образом [4].

Третий этап исследований [5] включал в себя применение построенной архитектуры к новому спектру задач. Обработка естественно-языковых конструкций оказалась хорошо применимой к базам знаний на основе мультиграфовых структур. Мультиграфовые семантические словари использовались ещё в начале серии исследований [3], когда они были естественно-языковыми. Другие типы терминологических словарей тоже хорошо поддаются анализу с помощью семантической функции энергии. В работе [5] многослойная сеть получила дополнительное

обучение на специализированных словарях и решала задачи различного типа, в том числе интересующую нас задачу Link Prediction.

Как видно, процедуры преобразований в описанной архитектуре достаточно сложны, а процессы, лежащие в их основе, интуитивно неясны. Кроме того, требуется объемная подготовительная работа, многослойное и мультизадачное обучение архитектуры для получения адекватного результата.

В отличие от использованного в [5] метода Deep Learning метод классификации, основанный на основании ВФП, способен извлекать из представленной онтологии структурные зависимости на логическом уровне, что позволяет автоматически отображать контекст высказывания. Метод классификации, основанный на ВФП, может не только эффективно принимать решения, но в отличие от нейронных сетей и других архитектур Data Mining он способен объяснять эти решения с помощью логико-вероятностных закономерностей, участвующих в формировании замыкания в алгоритме *StatClosure*. Для работы метода классификации не требуется замены семантических единиц на вещественнозначные векторы, что неизбежно привело бы к потере части информации и искажению смысла; процесс обучения при вещественнозначной интерпретации выглядит как минимизация потерянного смысла.

Перейдем к сравнительному эксперименту по применению метода классификации, основанного на ВФП к данным из [5]. Сравнение проводилось на основе выборки UMLS (Universal Medical Language System) [12]. Данные были предварительно обработаны и приведены к тензорному представлению авторами статьи [5; 6]. Особенностью этих данных является более общая структура, в которой онтология предметной области (в данном случае биологических и медицинских знаний) выражена в виде размеченного мультиграфа. Различные сущности (субъекты онтологии) связаны между собой множеством взаимоотношений (например, такими как “являться частью”, “использовать” или “производить”).

*UMLS* содержит миниатюрный медицинский словарь в виде семантических соответствий вида  $(l, rel, r)$ , где  $l$  и  $r$  — некоторые термины, а  $rel$  — отношение между ними. Например, в словаре представлены связи:

- (Immunologic\_Factor, indicates, Virus);
- (Molecular\_Biology\_Research\_Technique, is\_a, Research\_Activity).

Всего словарь содержит 135 терминов (вершин мультиграфа), 49 возможных отношений (меток для ребер мультиграфа) и 5932 связей между терминами (ребер в мультиграфе).

Сформулируем задачу для *UMLS* в терминах анализа формальных понятий и изучим точность предлагаемого подхода. На языке формальных понятий задача может быть задана с помощью понятия триадического формального контекста [8]. Помимо множества объектов

и множества атрибутов в триадическом формальном контексте вводится множество “условий”  $Rel$ ; отношение принадлежности атрибута объекту справедливо лишь при указании некоторого условия  $c \in C$ :  $(g, m, c) \in I \Leftrightarrow$  объект  $g$  обладает свойством  $m$  при условии  $c$ .

Пусть множество терминов медицинского словаря  $UMLS$  есть  $Term_U$ , а множество отношений —  $Rel_U$ . Для применения метода извлечения вероятностных формальных понятий требуется преобразовать размеченный мультиграф, или эквивалентный ему триадический контекст, в обыкновенный неориентированный двудольный граф без потери информации. Как было отмечено в [8], процедура преобразования неоднозначна и может зависеть как от семантики разметки графа, так и от семантики задачи. Однако в случае, когда описываются некоторые свойства объектов с помощью сложных, составных прилагательных, разумно будет предположить, что в качестве нового набор признаков  $\tilde{M}$  следует использовать пары, образующие комплексные прилагательные; так, в состав новых атрибутов войдут пары  $(is\_a, Virus)$  и  $(produces, Vitamin)$ . Для этого мы полагаем:

- 1)  $G_U = Term_U$ ;
- 2)  $M_U = Term_U \times Rel_U$ ;
- 3)  $(l, (rel, r)) \in I_U \Leftrightarrow (l, rel, r) \in UMLS$ ;
- 4) Формальный контекст нашей задачи это  $K_U = (G_U, M_U, I_U)$ .

Напомним, что решается задача типа Link Prediction [5]:

**Задача 1.** Определить, верно ли что для заданного ребра  $e = (l, rel, r)$  выполняется  $e \in U\tilde{M}LS$ . Или эквивалентно, верно ли что для заданных термина  $g$  и атрибута  $m = (rel, r)$  выполняется  $gIm$ .

Серия экспериментов состоит в удалении каких-либо связей (или значений в матрице отношения  $I$ ) из мультиграфа  $UMLS$ , а затем использовании алгоритма *StatClosure* в целях восстановления этих связей. Алгоритм 3 дает решение поставленной задачи восстановления связей методом ВФП.

## 5. Результаты

В качестве показателя эффективности работы классификатора авторы [5] используют *AUC* (Area Under Curve, площадь под *ROC*-кривой). Приведем определение *ROC*-кривой:

- FP (False Positive) — образцы, не содержащие анализируемый признак  $m$ , и ошибочно определенные как “содержащие признак  $m$ ”;

**Алгоритм 3** Решение задачи предсказания связи**Вход:**  $\mathcal{R} \subseteq \text{For}(K)$ ,  $K_c = (G_c, M, I_c)$ ,  $g \in G_c$ ,  $m \in M$ **Выход:**  $\top$  или  $\perp$  — верно или нет, что  $g$  обладает атрибутом  $m$ 1: **Функция** PREDICTLINK( $K_c, \mathcal{R}, g, m$ )2:      $L \leftarrow g' \setminus \{m\}$ 3:      $\bar{L} \leftarrow \text{StatClosure}(G_c, \mathcal{R}, L)$ 4:     **Если**  $m \in \bar{L}$  **тогда**5:         **Вернуть**  $\top$ 6:     **Иначе**7:         **Вернуть**  $\perp$ 8:     **Конец условия**9: **Конец функции**

- FN (False Negative) — образцы, не содержащие анализируемый признак  $m$ , и правильно определенные как “не содержащие признак  $m$ ”;
- TN (True Negative) — образцы, содержащие анализируемый признак  $m$ , и ошибочно определенные как “не содержащие признак  $m$ ”;
- TP (True Positive) — образцы, содержащие анализируемый признак  $m$ , и правильно определенные как “содержащие признак  $m$ ”;
- Точка на ROC-кривой определяется как пара  $(\frac{|FP|}{|FP \cup FN|}, \frac{|TP|}{|TP \cup TN|})$ .

ROC-кривая содержит множество таких точек, и в случае конечного множества экспериментов, образует ломаную линию на графике в области  $((0, 0), (1, 1))$ . Площадь под ROC-кривой и будет значением показателя AUC [11].

Обычно алгоритм имеет некоторые параметры, к примеру — порог решающего правила. С помощью этого параметра регулируется баланс между FP и TN образцами. Для алгоритма поиска вероятностных формальных понятий, таким параметром может служить вес противоречий. Повышая вес противоречий, мы понижаем количество FP-объектов (поскольку отвергаются практически все объекты, за исключением наиболее непротиворечивых), но при этом повышаем долю TN-объектов.

Заметим также, что AUC не является безусловно объективным показателем [11], поэтому мы будем также использовать классический показатель *Accuracy* (точность предсказаний). Для этого считаем процент правильно предсказанных случаев среди интересующего нас множества экземпляров  $G_K$ :

$$\text{Accuracy}(G_K) = \frac{|G_K \cap (FN \cup TP)|}{|G_K|}$$

Эксперимент по классификации состоял из двух частей:

- изучения поведения алгоритма при изменении уровня веса противоречий, составления *ROC*-кривой и подсчета *AUC*;
- скользящий контроль типа leave-one-out для тотального контроля точности (ассигасу).

В процессе изменения уровня противоречий алгоритм показывал очень высокую точность предсказаний в рамках задачи Link Prediction. С увеличением веса противоречий удается повысить точность на образцах  $g$  с отсутствующими связями  $(g, (rel, \tilde{g})) \notin I_U$  за счет уменьшения точности на позитивных образцах. Требования к сильной непротиворечивости делают алгоритм более консервативным, что заставляет его признавать наличие ассоциаций типа  $(g, rel, \tilde{g})$  более осторожно.

Таблица 1

Зависимость точности от веса противоречий  $w$ 

Вес противоречий	Негативные	Позитивные	Совокупно
0	93.46%	100%	93.51%
0.2	97.37%	100%	97.33%
0.4	98.13%	100%	98.15%
0.6	98.5%	100%	98.51%
0.8	98.82%	99.46%	98.83%
1.0	98.87%	100%	98.98%
1.5	99.24%	99.52%	99.24%
2.0	99.38%	99.55%	99.39%
3.0	99.69%	99.56%	99.67%
4.0	99.69%	97.56%	99.67%
6.0	99.69%	97.56%	99.67%

Анализ *ROC*-кривой показал, что *StatClosure* способен давать производительность не хуже *SME*-алгоритма, основанного на глубоком обучении [5]. *AUC* для *StatClosure* составляет 0.9997, что даже лучше показателя 0.99 для *SME*-алгоритма глубинного обучения. Это может служить сигналом к дальнейшим экспериментам по прогностической способности вероятностных формальных понятий. Результаты сравнительного анализа приведены в таблице 2.

Перейдем теперь к контролю точности. Объём выборки является небольшим (135 объектов), технически и вычислительно возможен наиболее полный вариант техники Cross-Validation типа JackKnife [2].

Занумеруем объекты исходной выборки  $G = \{g_1, \dots, g_{|G|}\}$ . В leave-one-out, на  $k$ -й итерации для обучения используется выборка  $G_{Teach}^k = G \setminus \{g_k\}$ , а для контроля  $G_{Control}^k = \{g_k\}$ . Сначала следует процедура извлечения правил посредством семантического вероятностного вывода [1] на контексте  $(G_{Teach}^k, \tilde{M}, \tilde{I})$ , которая извлекает все вероятностные

закономерности на данных, а затем процедура предсказания связей *LinkPrediction*.

Таблица 2

Статистика эксперимента для leave-one-out, 135  
итераций

Показатель	Позитивных	Негативных	Всего
Верно	6104	878213	884317
Неверно	648	8060	8708
Всего	6752	886273	893025
Точность	90,4%	99,09%	99,025%

Таким образом, ожидаемая точность алгоритма на контрольных данных составляет порядка 99%. Если ошибка типа *FP* является критичной, требования к весу противоречий может быть снижено с целью более качественной классификации позитивных образцов.

## 6. Заключение

В работе изучена возможность практического применения подхода машинного обучения, основанного на вероятностном обобщении формальных понятий. Концепция семантической энергии способна разрешить противоречия, возникающие в условиях сильной неопределенности.

Четко заданная семантика, как в рассмотренном наборе данных *UMLS*, обеспечивает высокую эффективность алгоритма, даже в сравнении с общепризнанными техниками на основе Deep Learning. Метод классификации, основанный на вероятностных формальных понятиях, дает не только лучшие результаты, но и имеет высокую объясняющую способность, поскольку эти результаты хорошо интерпретируемы.

Также исследована гибкость предложенного алгоритма *StatClosure* за счет введения параметра веса противоречий, который позволяет регулировать баланс между консервативностью и точностью алгоритма и оценить значение *AUC*-показателя метода. *AUC* для *StatClosure* оказывается сравнимым с аналогичным показателем для *SME*-метода на основе глубинного обучения [5].



### Список литературы

1. Витяев Е. Е. Извлечение знаний из данных. Компьютерное познание. Моделирование когнитивных процессов / Е. Е. Витяев. – Новосибирск : НГУ, 2006. – 293 с.
2. Воронцов К. В. Комбинаторный подход к оценке качества обучаемых алгоритмов / К. В. Воронцов // Математические вопросы кибернетики / под ред. О. Б. Лупанов. – М. : Физматлит, 2004. – Т. 13. – С. 5–36.
3. Learning structured embeddings of knowledge bases / A. Bordes, J. Weston, R. Collobert, Y. Bengio // Proceedings of the 25th Conference on Artificial Intelligence (AAAI-11). – USA, San Francisco, 2011.
4. Joint learning of words and meaning representations for open-text semantic parsing / A. Bordes, X. Glorot, J. Weston, Y. Bengio // Proc. of the 15th Intern. Conf. on Artif. Intel. and Stat. (JMLR) – 2012. – Vol. 22. – P. 127–135.
5. A Semantic Matching Energy Function for Learning with Multi-relational Data / A. Bordes [et al.] // Machine Learning: Special Issue on Learning Semantics. – 2013.
6. A Neural Probabilistic Language Model / Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin // Machine Learning Research. – 2003. – P. 1137–1155.
7. Ganter B. Formal Concept Analysis: Methods, and Applications in Computer Science / B. Ganter. – TU Dresden : Springer, 2003
8. Ganter B. Implications in Triadic Formal Contexts / B. Ganter, S. Obiedkov. – TU Dresden : Springer, 2004.
9. Goodfellow I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – MIT Press, 2016.
10. A Tutorial on Energy-Based Learning / Y. LeCun [et al.] // Predicting Structured Outputs / Bakir [et al.] (eds.). – MIT Press, 2006.
11. Lobo J. M. AUC: a misleading measure of the performance of predictive distribution models / J. M. Lobo, A. Jimenez-Valverde, R. Real // Global Ecology and Biogeography. – 2008. – Vol. 17. – P. 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
12. McCray A. T. An upper level ontology for the biomedical domain // Comparative and Functional Genomics. – 2003. – Vol. 4. – P. 80–88. <https://doi.org/10.1002/cfg.255>
13. Vityaev E. E. Probabilistic Formal Concepts with Negation / E. E. Vityaev, V. V. Martynovich // Perspectives of System Informatics / A. Voronkov, I. Virbitskaite (eds.). – LNCS. – 2015. – Vol. 8974. – P. 385–399.

**Витяев Евгений Евгеньевич**, доктор физико-математических наук, профессор, Институт математики СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Коптюга, 4, тел.: (+7383) 363 46 58 (e-mail: [vityaev@math.nsc.ru](mailto:vityaev@math.nsc.ru))

**Мартынович Виталий Валерьевич**, Институт математики СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Коптюга, 4, тел.: +7 963 948 71 57 (e-mail: [vilco@ya.ru](mailto:vilco@ya.ru))

---

**V. V. Martynovich, E. E. Vityaev**  
**Probabilistic Formal Concepts in a Natural Language Processing Task**

**Abstract.** Despite the high effectiveness of the Deep Learning methods, they remain a "thing in themselves", a "black box" decisions of which can not be trusted. This is critical for such areas as medicine, financial investments, military applications and others, where the price of the error is too high. In this regard, the European Union is going to demand in 2018 from companies that they give users an explanation of the solutions obtained by automatic systems.

In this paper, we offer an alternative, logical-probabilistic method of Deep Learning that can explain its decisions. This is a method of hierarchical clustering, based on the original logical-probabilistic generalization of formal concepts [13]. For comparison with deep learning based on neural networks, the work [12] was chosen, in which the task of processing natural language on a set of data *UMLS* is solved. To apply the logical-probabilistic generalization of formal concepts, a classification algorithm based on the energy of the contradictions Energy Learning [10] is defined. Logical-probabilistic formal concepts are defined through fixed points, as well as the formal concepts themselves, only certain rules of probability are used as rules. The energy of contradictions allows us to resolve the contradictions arising at fixed points that form probabilistic formal concepts. It is shown that this clustering algorithm is not inferior in accuracy to the method of Deep Learning [12]; nevertheless, the solutions obtained by it are explained by the set of probabilistic fixed-point rules.

**Keywords:** probability, formal concepts, natural language processing, Deep Learning, Data Mining, semantic energy.

## References

1. Vityaev E.E. *Izvlechenie znaniy iz dannykh. Komp'yuternoe poznanie. Modelirovanie kognitivnykh protsessov* [Knowledge discovery. Computational cognition. Cognitive process models]. Novosibirsk, Novosibirsk State University Press, 2006. 293 p. (in Russian)
2. Vorontsov K.V. Combinatorial approach to learning algorithms estimation. *Matematicheskie voprosy kibernetiki*, eds. Lupanov O.B., Moscow, FizMatLit, 2004, vol. 13, pp. 5-36. (in Russian).
3. Bordes A., Weston J., Collobert R., Bengio Y. Learning structured embeddings of knowledge bases. *Proceedings of the 25th Conference on Artificial Intelligence*, AAAI-11, USA, San Francisco, 2011.
4. Bordes A., Glorot X., Weston J., Bengio Y. Joint learning of words and meaning representations for open-text semantic parsing. *Proc. of the 15th Intern. Conf. on Artif. Intel. and Stat.*, JMLR, 2012, vol. 22, pp. 127–135.
5. Bordes A. et al. A Semantic Matching Energy Function for Learning with Multi-relational Data. *Machine Learning: Special Issue on Learning Semantics*, 2013.
6. Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. A Neural Probabilistic Language Model. *Machine Learning Research*, 2003, pp. 1137-1155.
7. Ganter B. *Formal Concept Analysis: Methods, and Applications in Computer Science*. TU Dresden, Springer, 2003.
8. Ganter B., Obiedkov S. *Implications in Triadic Formal Contexts*. TU Dresden, Springer, 2004.
9. Goodfellow I., Bengio Y. and Courville A. *Deep Learning*. MIT Press, 2016.
10. LeCun Y. et al. A Tutorial on Energy-Based Learning. *Predicting Structured Outputs*, Bakir et al. (eds.), MIT Press, 2006.

11. Lobo J.M., Jimenez-Valverde A., Real R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 2008, vol. 17, pp. 145-151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
12. A. T. McCray. An upper level ontology for the biomedical domain. *Comparative and Functional Genomics*, 2003, vol. 4, P. 80-88. <https://doi.org/10.1002/cfg.255>
13. Vityaev E.E., Martynovich V.V. Probabilistic Formal Concepts with Negation. *Perspectives of System Informatics*, A. Voronkov, I. Virbitskaite (eds.), LNCS, 2015, vol. 8974, pp. 385-399.

**Martynovich Vitaly Valerievich**, Institute of Mathematics SB RAS, 4, Akademik Koptiyag st., Novosibirsk, 630090, Russian Federation, tel.: +7 963 948 71 57, (e-mail: [vilco@ya.ru](mailto:vilco@ya.ru))

**Vityaev Evgeny Evgenievich**, Doctor of Sciences (Physics and Mathematics), Professor, Institute of Mathematics SB RAS, 4, Akademik Koptiyag st., Novosibirsk, 630090, Russian Federation, tel.: (+7383) 363 46 58 (e-mail: [vityaev@math.nsc.ru](mailto:vityaev@math.nsc.ru))