

# OSTIS-2015

## (Open Semantic Technologies for Intelligent Systems)

УДК 004.89

### ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ДОВЕРИЯ ДЛЯ ВЫДЕЛЕНИЯ СЕМАНТИЧЕСКИ ЗНАЧИМЫХ ПРИЗНАКОВ

Головко В.А., Крощенко А.А.

*Брестский государственный технический университет,  
г. Брест, Республика Беларусь*

*gva@bstu.by*

*kroschenko@gmail.com*

В работе приводятся основные принципы построения и обучения нейронной сети глубокого доверия для выделения семантически значимых признаков на основе выборки CIFAR-10. Для предобучения нейронной сети глубокого доверия применяется разработанный подход, базирующийся на минимизации ошибки реконструкции видимых и скрытых образов для ограниченной машины Больцмана (RBM).

**Ключевые слова:** Семантическая сеть, нейронная сеть глубокого доверия, ограниченная машина Больцмана, ошибка реконструкции, метод обратного распространения ошибки

#### Введение

Семантические и нейронные сети являются моделями представления знаний. При этом если семантическая сеть хранит знания об объектах и процессах окружающего мира в виде структуры с связями и отношениями, то нейронная сеть оперирует их численными свойствами. Возникает важный вопрос: каким образом возможно использовать аппарат нейронных сетей в построении семантических моделей? Ответить на данный вопрос помогут новые идеи и подходы, появившиеся в теории нейронных сетей в последнее десятилетие, связанные с нейронными сетями глубокого доверия.

Начиная с 2006 г., благодаря работам [Hinton et al., 2006a], [Hinton, 2002], [Hinton et al., 2006b], [Hinton, 2010] Хинтона (J. Hinton), начался обратный отсчет в развитии многослойных перцептронов под новым названием - нейронные сети глубокого доверия (deep belief neural networks). Такие сети в общем случае представляют собой дальнейшее развитие многослойных перцептронов и интегрируют различные парадигмы обучения нейронных сетей. Благодаря своей многослойной архитектуре, они позволяют обрабатывать и анализировать большой объем данных, а также моделировать когнитивные процессы в различных областях. В настоящее время большинство высокотехнологичных компаний в США (Microsoft, Google и т.д.) используют нейронные сети глубокого доверия для проектирования различных

интеллектуальных систем. По версии ученых Массачусетского технологического института (США), нейронные сети глубокого доверия (deep belief neural networks), входят в список 10 наиболее прорывных высоких технологий, способных в недалеком будущем в значительной степени преобразить повседневную жизнь большинства людей на нашей планете. В настоящее время нейронные сети глубокого доверия считаются революционным шагом в области интеллектуальной обработки данных.

В данной статье рассмотрен одна из основных моделей обучения нейронных сетей глубокого доверия, базирующееся на ограниченной машине Больцмана (restricted Boltzmann machine (RBM)). Предложен новый метод для обучения ограниченной машины Больцмана и показано, что правило обучения ограниченной машины Больцмана является частным случаем предложенного метода обучения, который базируется на минимизации суммарной квадратичной ошибки восстановления информации. Доказана эквивалентность максимизации вероятности распределения данных в ограниченной машине Больцмана и минимизация суммарной квадратичной ошибки восстановления информации в слоях (RBM).

Рассматривается применение нейронных сетей глубокого доверия для решения задачи выделения семантически значимых признаков на примере выборки CIFAR-10.

## 1. Нейронные сети глубокого доверия

Нейронная сеть глубокого доверия содержит множество скрытых слоев (рисунок 1) и осуществляет глубокое иерархическое преобразование входного пространства образов.

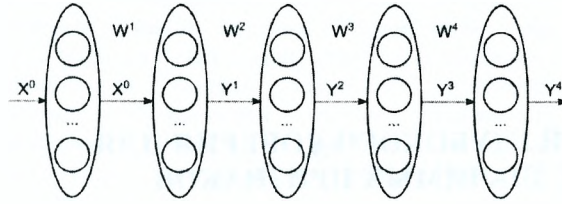


Рисунок 1 - Нейронная сеть глубокого доверия

Выходное значение  $j$ -го нейрона  $k$ -го слоя определяется следующим образом:

$$y_j^k = F(S_j^k), \quad (1)$$

$$S_j^k = \sum_{i=1} w_{ij}^k y_i^{k-1} + T_j^k, \quad (2)$$

где  $F$  – функция активации нейронного элемента,  $S_j^k$  – взвешенная сумма  $j$ -го нейрона  $k$ -слоя,  $w_{ij}^k$  – весовой коэффициент между  $i$ -ым нейроном  $(k-1)$ -го слоя и  $j$ -м нейроном  $k$ -го слоя,  $T_j^k$  – пороговое значение  $j$ -го нейрона  $k$ -го слоя.

Для первого (распределительного) слоя

$$y_i^0 = x_i. \quad (3)$$

В матричном виде выходной вектор  $k$ -го слоя

$$Y^k = F(S^k) = F(W^k Y^{k-1} + T^k), \quad (4)$$

где  $W$  – матрица весовых коэффициентов,  $Y^{k-1}$  – выходной вектор  $(k-1)$ -го слоя,  $T^k$  – вектор пороговых значений нейронов  $k$ -го слоя. Если нейронная сеть глубокого доверия используется для классификации образов, то выходные значения сети часто определяются на основе функции активации **softmax**:

$$y_j^F = \text{softmax}(S_j) = \frac{e^{S_j}}{\sum_i e^{S_i}} \quad (5)$$

Процесс обучения нейронных сетей глубокого доверия в общем случае состоит из двух этапов: предобучение нейронной сети методом послойного обучения, начиная с первого слоя (pre-training). Данное обучение осуществляется без учителя. Настройка синаптических связей всей сети (fine-tuning) при помощи алгоритма обратного распространения ошибки или алгоритма «бодрствования и сна» (wake-sleep algorithm).

Важным этапом обучения нейронных сетей глубокого доверия является предобучение слоев нейронной сети. Один из основных подходов к

предобучению – метод, базирующийся на представлении каждого слоя нейронной сети в виде ограниченной машины Больцмана (RBM).

### 1.1. Ограниченная машина Больцмана

Ограниченная машина Больцмана состоит из двух слоев стохастических бинарных нейронных элементов, которые соединены между собой двунаправленными симметричными связями (рисунок 2). Входной слой нейронных элементов называется видимым (слой  $X$ ), а второй слой называется скрытым (слой  $Y$ ). Нейронную сеть глубокого доверия можно представить как совокупность ограниченных машин Больцмана. Ограниченная машина Больцмана может генерировать (представить) любое дискретное распределение, если используется достаточное количество нейронов скрытого слоя [Bengio, 2009].

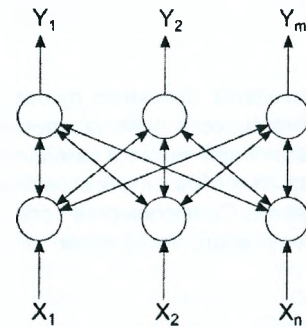


Рисунок 2 - Ограниченная машина Больцмана

Данная сеть является стохастической нейронной сетью, в которой состояния видимых и скрытых нейронов меняются в соответствии с вероятностной версией сигмоидной функции активации:

$$p(y_j | x) = \frac{1}{1 + e^{-S_j}}, S_j = \sum_i w_{ij} x_i + T_j \quad (6)$$

$$p(x_i | y) = \frac{1}{1 + e^{-S_i}}, S_i = \sum_j w_{ij} y_j + T_i \quad (7)$$

Состояния видимых и скрытых нейронных элементов принимаются независимыми:

$$P(x | y) = \prod_{i=1}^n P(x_i | y)$$

$$P(y | x) = \prod_{j=1}^m P(y_j | x)$$

Таким образом, состояния всех нейронных элементов ограниченной машины Больцмана определяются через распределение вероятностей. В RBM нейроны скрытого слоя являются детекторами признаков, которые сохраняют закономерности входных данных. Основная задача обучения состоит в воспроизведении распределения входных данных на основе состояний нейронов скрытого слоя как можно точнее. Это эквивалентно максимизации функции правдоподобия путем модификации синаптических связей нейронной сети. Рассмотрим

это подробнее. Вероятность нахождения видимого и скрытого нейрона в состоянии  $(x, y)$  определяется на основе распределения Гиббса:

$$P(x, y) = \frac{e^{-E(x, y)}}{Z}$$

где  $E(x, y)$  – энергия системы в состоянии  $(x, y)$ ,  $Z$  – параметр, который определяет условие нормализации вероятностей, то есть, чтобы сумма вероятностей равнялась единице. Данный параметр определяется следующим образом:

$$Z = \sum_{x, y} e^{-E(x, y)}$$

Вероятность нахождения видимых нейронов в определенном состоянии равняется сумме вероятностей конфигураций  $P(x, y)$  по состояниям скрытых нейронов:

$$P(x) = \sum_y P(x, y) = \sum_y \frac{e^{-E(x, y)}}{Z} = \frac{\sum_y e^{-E(x, y)}}{\sum_{x, y} e^{-E(x, y)}}$$

Для нахождения правила модификации синаптических связей необходимо максимизировать вероятность воспроизведения состояний видимых нейронов  $P(x)$  ограниченной машиной Больцмана. Для того, чтобы определить максимум функции правдоподобия распределения данных  $P(x)$  будем использовать метод градиентного спуска в пространстве весовых коэффициентов и пороговых значений сети, где в качестве градиента применим функцию логарифмического правдоподобия:

$$\ln P(x) = \ln \sum_y e^{-E(x, y)} - \ln \sum_{x, y} e^{-E(x, y)}$$

Тогда градиент равен

$$\frac{\partial \ln P(x)}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} \ln \sum_y e^{-E(x, y)} - \frac{\partial}{\partial w_{ij}} \ln \sum_{x, y} e^{-E(x, y)}$$

Преобразуя последнее выражение, получим

$$\begin{aligned} \frac{\partial \ln P(x)}{\partial w_{ij}} = & -\frac{1}{\sum_y e^{-E(x, y)}} \sum_y e^{-E(x, y)} \frac{\partial E(x, y)}{\partial w_{ij}} + \\ & + \frac{1}{\sum_{x, y} e^{-E(x, y)}} \sum_{x, y} e^{-E(x, y)} \frac{\partial E(x, y)}{\partial w_{ij}} \end{aligned}$$

Так как

$$P(x, y) = P(y | x)P(x)$$

то

$$P(y | x) = \frac{P(x, y)}{P(x)} = \frac{(1/Z)e^{-E(x, y)}}{(1/Z)\sum_y e^{-E(x, y)}} = \frac{e^{-E(x, y)}}{\sum_y e^{-E(x, y)}}$$

В результате можно получить следующее выражение:

$$\begin{aligned} \frac{\partial \ln P(x)}{\partial w_{ij}} = & -\sum_y P(y | x) \frac{\partial E(x, y)}{\partial w_{ij}} + \\ & + \sum_{x, y} P(x, y) \frac{\partial E(x, y)}{\partial w_{ij}} \end{aligned}$$

В данном выражении первое слагаемое определяет позитивную фазу работы машины Больцмана, когда сеть работает на основе образов из обучающей выборки. Второе слагаемое характеризует негативную фазу функционирования, когда сеть работает в свободном режиме независимо от окружающей среды. Рассмотрим энергию сети RBM. С точки зрения энергии сети задача обучения состоит в том, чтобы на основе входных данных найти конфигурацию выходных переменных с минимальной энергией. В результате на обучающем множестве сеть будет иметь меньшую энергию по сравнению с другими состояниями. Функция энергии бинарного состояния  $(x, y)$  определяется аналогично сети Хопфилда:

$$E(x, y) = -\sum_i x_i T_i - \sum_j y_j T_j - \sum_{i, j} x_i y_j w_{ij} \quad (8)$$

В этом случае

$$\frac{\partial E(x, y)}{\partial w_{ij}} = -x_i y_j$$

и

$$\frac{\partial \ln P(x)}{\partial w_{ij}} = \sum_y P(y | x) x_i y_j - \sum_{x, y} P(x, y) x_i y_j$$

Так как математическое ожидание равняется:

$$E(x) = \sum_i x_i P_i$$

то

$$\frac{\partial \ln P(x)}{\partial w_{ij}} = E[x_i y_j]_{data} - E[x_i y_j]_{model}$$

Аналогичным образом можно получить градиенты для пороговых значений:

$$\begin{aligned} \frac{\partial \ln P(x)}{\partial T_i} &= E[x_i]_{data} - E[x_i]_{model} \\ \frac{\partial \ln P(x)}{\partial T_j} &= E[y_j]_{data} - E[y_j]_{model} \end{aligned}$$

Как следует из последних выражений, первое слагаемое характеризует работу сети на основе данных из обучающей выборки, а второе слагаемое – на основе данных модели (данные генерируемые сетью), то есть в свободном режиме независимо от окружающей среды. Так как вычисление математического ожидания на основе RBM сети



является очень сложным, Хинтон предложил использовать аппроксимацию данных слагаемых, которую он назвал контрастным расхождением (contrastive divergence (CD)) [Hinton et al., 2006a]. Такая аппроксимация основывается на дискретизаторе Гиббса (Gibbs sampling). В этом случае первые слагаемые в выражениях для градиента характеризуют распределение данных в момент времени  $t = 0$ , а вторые слагаемые характеризуют реконструированные или генерируемые моделью состояния в момент времени  $t = k$ . Исходя из этого, CD-k процедура может быть представлена следующим образом:

$$x(0) \rightarrow y(0) \rightarrow x(1) \rightarrow y(1) \rightarrow \dots \rightarrow x(k) \rightarrow y(k)$$

В результате можно получить следующие правила для обучения RBM сети. В случае применения CD-1,  $k = 1$  и учитывая, что в соответствии с методом градиентного спуска

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \frac{\partial \ln P(x)}{\partial w_{ij}(t)}$$

Можно получить, что

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(x_i(0)y_j(0) - x_i(1)y_j(1))$$

$$T_i(t+1) = T_i(t) + \alpha(x_i(0) - x_i(1))$$

$$T_j(t+1) = T_j(t) + \alpha(y_j(0) - y_j(1)).$$

Аналогичным образом, для алгоритма CD-k

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(x_i(0)y_j(0) - x_i(k)y_j(k))$$

$$T_i(t+1) = T_i(t) + \alpha(x_i(0) - x_i(k))$$

$$T_j(t+1) = T_j(t) + \alpha(y_j(0) - y_j(k)).$$

Из последних выражений видно, что правила обучения ограниченной машины Больцмана минимизируют разницу между оригинальными данными и данными генерируемыми моделью. Генерируемые моделью данные получаются при помощи сэмплирования Гиббса.

## 1.2. Обучение нейронной сети глубокого доверия на основе RBM

Обучение нейронной сети глубокого доверия происходит на основе «жадного» алгоритма послойного обучения (greedy layer-wise algorithm). В соответствии с ним вначале обучается первый слой сети как RBM машина. Для этого входные данные поступают на видимый слой нейронных элементов и используя CD-k процедуру вычисляются состояния скрытых  $p(y|x)$  и видимых нейронов  $p(x|y)$ . В процессе выполнения данной процедуры (не более 100 эпох) изменяются весовые коэффициенты и пороговые значения RBM сети, которые затем фиксируются. Затем берется второй слой нейронной сети и конструируется RBM машина. Входными данными для нее являются данные с предыдущего слоя. Происходит обучение и процесс продолжается для всех слоев нейронной сети [Hinton, 2009]. В результате такого обучения без учителя можно получить подходящую

начальную инициализацию настраиваемых параметров сети глубокого доверия. На заключительном этапе осуществляется точная настройка параметров всей сети при помощи алгоритма обратного распространения ошибки или алгоритма «бодрствования и сна» (wake-sleep algorithm).

## 1.3. Альтернативный взгляд на ограниченную машину

В данном разделе рассматривается альтернативный взгляд на ограниченную машину Больцмана как автоассоциативную нейронную сеть, которая может функционировать с любыми данными, как бинарными, так и числовыми. Предлагается новый метод для получения правила обучения ограниченной машины Больцмана [Golovko et al., 2014]. Он базируется на минимизации ошибки реконструкции видимых и скрытых образов, которую можно получить, используя итерации сэмплирования Гиббса. По сравнению с традиционным подходом, основанным на энергии методе (energy-based method), который базируется на линейном представлении нейронных элементов, предложенный метод позволяет учитывать нелинейную природу нейронных элементов. Рассмотрим ограниченную машину Больцмана, которую будем представлять в виде трех слоев нейронных элементов [Golovko et al., 2012]: видимый, скрытый и видимый (рисунок 3).

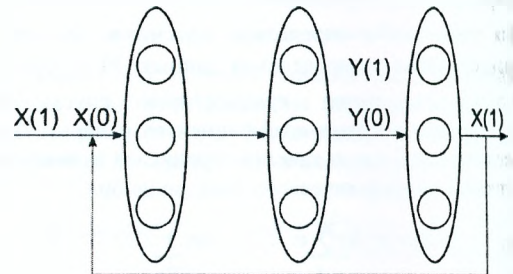


Рисунок 3 - Представление RBM в виде рециркуляционной сети

Сэмплирование Гиббса заключается в следующей процедуре. Пусть  $x(0)$  входной вектор, который поступает на видимый слой в момент времени 0. Тогда выходные значения нейронов скрытого слоя:

$$y_j(0) = F(S_j(0)), \quad (9)$$

$$S_j(0) = \sum_i w_{ij} x_i(0) + T_j. \quad (10)$$

Инверсный (последний) слой реконструирует входной вектор на основе данных со скрытого слоя. В результате получается восстановленный вектор  $x(1)$  в момент времени 1:

$$x_i(1) = F(S_i(1)), \quad (11)$$

$$S_i(1) = \sum_j w_{ji} y_j(0) + T_i. \quad (12)$$

Затем вектор  $x(1)$  поступает на видимый слой, и вычисляются выходные значения нейронов скрытого слоя:

$$y_j(1) = F(S_j(1)), \quad (13)$$

$$S_j(1) = \sum_i w_{ij} x_i(1) + T_j. \quad (14)$$

Продолжая данный процесс, можно получить на шаге  $k$

$$y_j(k) = F(S_j(k)), S_j(k) = \sum_i w_{ij} x_i(k) + T_j.$$

$$x_i(k) = F(S_i(k)), S_i(k) = \sum_j w_{ij} y_j(k-1) + T_i.$$

Целью обучения ограниченной машины Больцмана является минимизация суммарной квадратичной ошибки реконструкции данных на скрытом и восстанавливающем слое, которая в случае CD- $k$  определяется следующим образом:

$$E_s = \frac{1}{2} \sum_{l=1}^L \sum_{j=1}^m \sum_{p=1}^k (y'_j(p) - y_j^l(p-1))^2 + \\ + \frac{1}{2} \sum_{l=1}^L \sum_{i=1}^n \sum_{p=1}^k (x'_i(p) - x_i^l(p-1))^2$$

В случае CD-1 суммарная квадратичная ошибка

$$E_s = \frac{1}{2} \sum_{l=1}^L \sum_{j=1}^m (y'_j(1) - y_j^l(0))^2 + \\ + \frac{1}{2} \sum_{l=1}^L \sum_{i=1}^n (x'_i(1) - x_i^l(0))^2$$

где  $L$  количество входных образов. Как следует из приведенных выше выражений ошибка состоит из двух частей: ошибки восстановления информации на скрытом и выходном слое.

**Теорема 1.** Максимизация функции правдоподобия распределения данных  $P(x)$  в пространстве синаптических связей ограниченной машины Больцмана эквивалентна минимизации суммарной квадратичной ошибки сети в том же пространстве при использовании линейных нейронов.

**Следствие 1.** Линейная ограниченная машина Больцмана с точки зрения обучения эквивалентна автоассоциативной нейронной сети при использовании в ней при обучении сэмплирования Гиббса.

**Следствие 2.** Для нелинейной ограниченной машины Больцмана правило модификации синаптических связей в случае CD- $k$  будет следующим:

$$w_{ij}(t+1) = w_{ij}(t) -$$

$$- \alpha \left( \sum_{p=1}^k (y_j(p) - y_j(p-1)) x_i(p) F'(S_j(p)) + (x_i(p) - x_i(p-1)) y_j(p-1) F'(S_i(p)) \right) \\ T_j(t+1) = T_j(t) - \alpha \left( \sum_{p=1}^k (y_j(p) - y_j(p-1)) F'(S_j(p)) \right) \\ T_i(t+1) = T_i(t) - \alpha \left( \sum_{p=1}^k (x_i(p) - x_i(p-1)) F'(S_i(p)) \right)$$

**Следствие 3.** Для нелинейной ограниченной машины Больцмана правило модификации синаптических связей в случае CD-1 будет следующим:

$$w_{ij}(t+1) = w_{ij}(t) -$$

$$\alpha ((y_j(1) - y_j(0)) F'(S_j(1)) x_i(1) + (x_i(1) - x_i(0)) F'(S_i(1)) y_j(0)),$$

$$T_i(t+1) = T_i(t) - \alpha (x_i(1) - x_i(0)) F'(S_i(1)),$$

$$T_j(t+1) = T_j(t) - \alpha (y_j(1) - y_j(0)) F'(S_j(1))$$

Если используется групповое обучение (batch learning), то в этом случае метод градиентного спуска записывается следующим образом:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\partial E_s}{\partial w_{ij}(t)} \quad (15)$$

$$T_i(t+1) = T_i(t) - \alpha \frac{\partial E_s}{\partial T_i(t)} \quad (16)$$

$$T_j(t+1) = T_j(t) - \alpha \frac{\partial E_s}{\partial T_j(t)} \quad (17)$$

В данном разделе получены правила обучения для ограниченной машины Больцмана, которые базируются на минимизации квадратичной ошибки восстановления информации в скрытом и видимом слоях. Предложенный метод позволяет учитывать нелинейную природу нейронных элементов. Показано, что классические выражения для обучения ограниченной машины являются частным случаем предложенного метода. Доказана теорема об эквивалентности максимизации функции правдоподобия распределения данных  $P(x)$  в пространстве синаптических связей и минимизации суммарной квадратичной ошибки сети в том же пространстве для линейной ограниченной машины Больцмана.

## 2. Применение нейронных сетей глубокого доверия

### 2.1. Постановка задачи и условия проведения эксперимента

Для анализа эффективности обучения разработанных алгоритмов была выбрана известная выборка CIFAR-10. Эта выборка состоит из 60000 изображений размерностью 32X32 различных объектов, принадлежащих 10 основным классам: 1.

самолет, 2. автомобиль, 3. птица, 4. кот, 5. олень, 6. собака, 7. лягушка, 8. лошадь, 9. корабль и 10. грузовик.

Для обучения распознавания объектов данных классов использовалась нейронная сеть глубокого доверия архитектуры 3072 – 1500 – 750 – 10. Из полной выборки было взято 20000 изображений для формирования выборки обучения. Основные параметры обучения: скорость обучения – 0,03 (на этапе предобучения и fine-tuning), размер mini-batch – 200, количество эпох предобучения – 10, fine-tuning – 50.

После выполнения обучения данной нейронной сети глубокого доверия на выходном слое были оценены расстояния между векторами весовых коэффициентов.

## 2.2. Постановка задачи и условия проведения эксперимента

Данные о расстояниях между весовыми коэффициентами на последнем (выходном) слое нейронной сети были упорядочены, прогнурованы и сведены в следующую таблицу псевдорасстояний:

Таблица 1 – Псевдорасстояния для классов CIFAR

|   |    |    |    |    |    |    |    |    |    |
|---|----|----|----|----|----|----|----|----|----|
| 0 | 37 | 16 | 20 | 2  | 17 | 21 | 7  | 1  | 34 |
|   | 0  | 42 | 36 | 41 | 38 | 45 | 44 | 26 | 15 |
|   |    | 0  | 33 | 11 | 24 | 18 | 23 | 13 | 30 |
|   |    |    | 0  | 8  | 3  | 10 | 32 | 29 | 35 |
|   |    |    |    | 0  | 4  | 5  | 12 | 6  | 40 |
|   |    |    |    |    | 0  | 14 | 22 | 28 | 39 |
|   |    |    |    |    |    | 0  | 27 | 25 | 43 |
|   |    |    |    |    |    |    | 0  | 19 | 31 |
|   |    |    |    |    |    |    |    | 0  | 9  |
|   |    |    |    |    |    |    |    |    | 0  |

Исследуя полученную таблицу, можно заметить, что близкими семантическими признаками обладают такие классы как, например, самолет и корабль, самолет и олень, кот и собака, олень и собака, олень и лягушка. При этом наименьшими общими признаками обладают классы автомобиль и лошадь, автомобиль и лягушка.

## Заключение

С развитием областей искусственного интеллекта наблюдается появление различных возможностей для интеграции идей и их совместного использования в рамках общих теорий. К примеру, вполне возможно использование нейронных сетей, а в частности, нейронных сетей глубокого доверия (как обладающий потенциалом в обработке больших объемов данных) в задачах выделения семантически значимых признаков и их

анализа. Как правило, при этом получаемые данные способствуют определению скрытых, зачастую плохо выявляемых при ручном анализе, общих признаков.

В данной статье предлагается новый подход к предобучению нейронных сетей глубокого доверия, а также приводится пример его использования для предобучения сети классификации образов. Параметры полученной сети анализируются с целью определения семантически значимых признаков.

## Библиографический список

- [Hinton et al., 2006a] Hinton, G. A fast learning algorithm for deep belief nets / G. Hinton, S. Osindero, Y. Teh // *Neural Computation*. – 2006. – Vol. 18. – P. 1527–1554.
- [Hinton, 2002] Hinton, G. Training products of experts by minimizing contrastive divergence // *Neural Computation*. – 2002. – Vol. 14. – P. 1771–1800.
- [Hinton et al., 2006b] Hinton, G. Reducing the dimensionality of data with neural networks / G. Hinton, R. Salakhutdinov // *Science*, 313 (5786). – 2006. – P. 504–507.
- [Hinton, 2010] Hinton, G. A practical guide to training restricted Boltzmann machines // *Tech. Rep. 2010-000*. – Toronto: Machine Learning Group, University of Toronto, 2010.
- [Bengio, 2009] Bengio, Y. Learning deep architectures for AI // *Foundations and Trends in Machine Learning*. – 2009. – Vol. 2(1). – P. 1–127.
- [Hinton, 2009] Hinton, G. Greedy layer-wise algorithm // *Journal of Machine Learning Research* 1. – 2009. – P. 1–40.
- [Golovko et al., 2014] Golovko, V. A Learning Technique for Deep Belief Neural Networks / V. Golovko, A. Kroshchanka, U. Rubanau, S. Jankowski // in book *Neural Networks and Artificial Intelligence*. – Springer, 2014. – Vol. 440. Communication in Computer and Information Science. – P. 136–146.
- [Golovko et al., 2012] Golovko, V. Neural network model for transient ischemic attacks diagnostics / V. Golovko, H. Vaitsekhovich, E. Apanel, A. Mastyskin // *Optical Memory and Neural Networks (Information Optics)*. – 2012. – 21(3). – P. 166–176.

## APPLYING DEEP BELIEF NEURAL NETWORKS TO EXTRACTION VALUABLE SEMANTIC FEATURES

Golovko V.A. \*, Kroshchanka A.A. \*

\*Brest State Technical University

gva@bstu.by

kroschenko@gmail.com

The main principles of construction and learning deep belief neural networks for extraction valuable semantic features are proposed. The proposed approach is based on minimization of reconstruction mean square error, which we can obtain using a simple iterations of Gibbs sampling.