



УДК 004.89

СЕМАНТИЧЕСКОЕ КОДИРОВАНИЕ НА ОСНОВЕ ГЛУБОКИХ АВТОАССОЦИАТИВНЫХ НЕЙРОННЫХ СЕТЕЙ

Головко В.А., Крощенко А.А., Хацкевич М.В.

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

gva@gmail.com

kroschenko@gmail.com

В работе предлагается подход к реализации семантического кодирования (хеширования) с помощью глубоких автоассоциативных нейронных сетей. Построенная на ее основе поисковая система отличается простотой и скоростью осуществления поиска релевантных изображений.

Ключевые слова: нейронные сети глубокого доверия; глубокие автоассоциативные сети; семантическое хеширование; анализ и распознавание изображений; ограниченная машина Больцмана

Введение

Вопросы семантического кодирования имеют первостепенное значение при реализации поиска по образцу. При осуществлении поиска среди изображений особое значение приобретает проблема скорости выполнения такой операции.

Теория нейронных сетей предоставляет для решения данной задачи модели, называемые автоассоциативными нейронными сетями. Применение данных моделей позволяет сжать исходные изображения и выделить наиболее информативную часть на среднем слое («бутылочном горлышке»). Обученная сеть позволяет соотнести с каждым изображением код, позволяющий оценить близость данного изображения с уже оцененными. Применение подобных сетей для бинарного кодирования изображений имеет ряд преимуществ – прежде всего быстрый поиск, на который не влияет размер базы данных [Krizhevsky et al., 2011].

Благодаря работам [Hinton et al., 2006a], [Hinton, 2002], [Hinton et al., 2006b], [Hinton, 2010], стало возможным осуществлять обучение сетей, характеризующихся большим количеством слоев. Эти сети получили название нейронных сетей глубокого доверия (Deep Belief Networks – DBN). Глубокие автоассоциативные нейронные сети представляют собой тип сетей, включающий ключевые особенности и нейронных сетей глубокого доверия, и автоассоциативных сетей. Модели этого класса обладают чертами, наличие

которых сделало возможным использовать их в решении задач, характеризующихся большими (гигабайтными) объемами данных.

В данной работе рассматривается подход к реализации семантического кодирования на основе глубокой автоассоциативной нейронной сети. В качестве обучающей выборки выступает база изображений CIFAR-10.

1. Нейронные сети глубокого доверия

1.1. Общая структура. Обучение

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

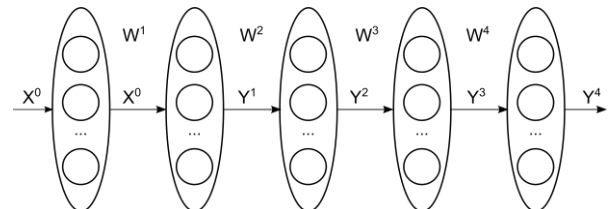


Рисунок 1 - Нейронная сеть глубокого доверия

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

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

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

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**:

$$y_j^F = \text{softmax}(S_j) = \frac{e^{s_j}}{\sum_l e^{s_l}} \quad (5)$$

Процесс обучения нейронных сетей глубокого доверия в общем случае состоит из двух этапов:

1. предобучение нейронной сети методом послыоного обучения, начиная с первого слоя (pre-training). Данное обучение осуществляется без учителя.
2. Настройка синаптических связей всей сети (fine-tuning) при помощи алгоритма обратного распространения ошибки или алгоритма «бодрствования и сна» [Hinton et al., 1995] (wake-sleep algorithm).

Ключевым этапом обучения нейронных сетей глубокого доверия, определяющим их преимущества перед поверхностными моделями, является предобучение. Один из основных подходов к предобучению – метод, базирующийся на представлении каждого слоя нейронной сети в виде ограниченной машины Больцмана (RBM), а всей сети – в виде стека таких машин.

1.2. 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].

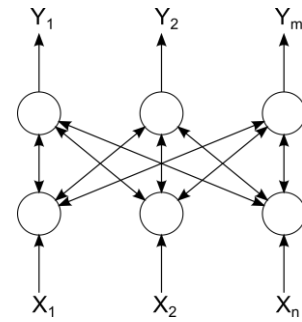


Рисунок 2 - Ограниченная машина Больцмана

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

$$p(y_j = 1 | x) = \sigma\left(\sum_i^n w_{ij} x_i + T_j\right) \quad (6)$$

$$p(x_i = 1 | y) = \sigma\left(\sum_j^m w_{ij} y_j + T_i\right) \quad (7)$$

где $\sigma(x) = \frac{1}{1 + e^{-x}}$, w_{ij} – весовые коэффициенты, T_i, T_j – пороговые элементы.

RBM такой структуры получила название бинарной RBM.

Помимо представленной модели, ориентированной на использование с бинарными данными, существует модель линейно-бинарной RBM, в которой на видимом слое применяется линейная функция активации. Подобная RBM рекомендуется к использованию при работе с не бинарными (вещественными) данными [Hinton, 2010].

Для обучения линейно-бинарной RBM используются следующие правила:

$$p(y_j = 1 | x) = \sigma\left(\sum_i^n w_{ij} x_i + T_j\right) \quad (8)$$

$$x_i = N\left(\sum_j^m w_{ij} y_j + T_i, 1\right) \quad (9)$$

где $N(m, s)$ – нормально распределенная случайная величина со средним значением m и стандартным отклонением s .

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 [Головко, 2015a].

Получаемые в процессе вывода выражения для обучения RBM являются достаточно сложными в вычислительном плане, поэтому Хинтон предложил использовать аппроксимацию, 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:

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) + \alpha(x_i(0)y_j(0) - x_i(k)y_j(k)) \\ T_i(t+1) &= T_i(t) + \alpha(x_i(0) - x_i(k)) \\ T_j(t+1) &= T_j(t) + \alpha(y_j(0) - y_j(k)). \end{aligned} \quad (10)$$

где $x_i(0)$, $y_j(0)$ характеризует распределение данных в момент времени $t=0$, $x_i(k)$, $y_j(k)$ - в момент времени $t=k$, k - параметр, определяющий количество стадий проходов «вперед-назад» для осуществления аппроксимации соответствующих градиентов.

В случае применения CD-1, $k = 1$ получим следующие правила обучения:

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) + \alpha(x_i(0)y_j(0) - x_i(1)y_j(1)) \\ T_i(t+1) &= T_i(t) + \alpha(x_i(0) - x_i(1)) \\ T_j(t+1) &= T_j(t) + \alpha(y_j(0) - y_j(1)). \end{aligned} \quad (11)$$

Из выражений (8), (9) видно, что правила обучения ограниченной машины Больцмана минимизируют разницу между оригинальными данными и данными генерируемыми моделью. Генерируемые моделью данные получаются при помощи процедуры сэмплирования Гиббса.

Для обучения RBM могут быть использованы и другие методы, например, [Golovko et al., 2014], [Головко, 2015b], [Головко и др., 2015c], [Golovko et al., 2015d].

1.3. Ограниченная машина Больцмана

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 новая 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).

2. Глубокие автоассоциативные нейронные сети

Автоассоциативная сеть – это многослойная нейронная сеть прямого распространения сигнала, обученная для того, чтобы выдавать входные данные на выходе сети. При обучении автоассоциативной сети «учителем» для нее является сама входная информация. На первой половине сети происходит так называемое «прямое распространение», т.е. осуществляется сжатие входных данных. Далее на второй половине сети, которая является зеркальным отражением первой, происходит «обратное распространение», т.е. восстановление входного образа. Обычно сеть имеет скрытый слой меньшей размерности, который выделяет наиболее значимые признаки во входной информации. Пример такой сети представлен на рисунке 3.

Автоассоциативные сети с таким слоем оказываются полезны при решении задач визуализации и обработки данных высокой размерности, так как позволяют существенно сократить объем данных. В этой работе исследуются глубокие автоассоциативные сети.

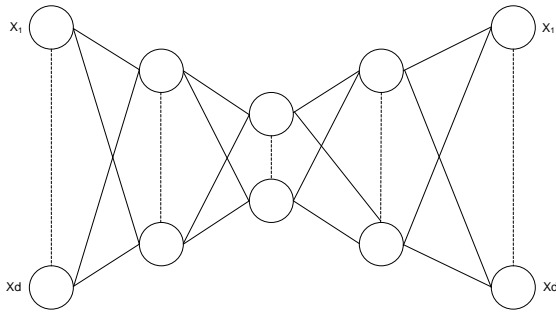


Рисунок 3 - Пример автоассоциативной сети

2.1. Обучение

В обучении глубоких автоассоциативных сетей также можно выделить два этапа:

1. Предобучение. На этом этапе выполняются действия, аналогичные производимым при предобучении обычных глубоких сетей с той лишь разницей, что предобучаются лишь те слои, которые производят кодирование информации.

2. Обучение (fine-tuning). На этом этапе предобученный кодировщик «разворачивается», формируя полную автоассоциативную сеть (рисунок 4). При этом в качестве матриц весовых коэффициентов для декодирующих слоев сети берутся транспонированные матрицы для соответствующих кодирующих слоев, а в качестве пороговых элементов выступают векторы порогов соответствующих видимых слоев RBM. После этого производится обучение получившейся автоассоциативной сети методом обратного распространения ошибки.

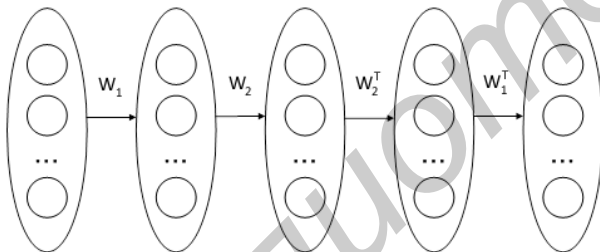


Рисунок 4 – Структура глубокой автоассоциативной сети

3. Применение глубоких автоассоциативных сетей для решения задачи семантического кодирования

3.1. Постановка задачи

В качестве обучающей выборки для решения задачи построения бинарных семантических кодов изображений нами была использована база CIFAR-10. Эта выборка включает в себя 50.000 изображений технических средств и живых существ, принадлежащий 10 различным классам (рис. 5). Каждое изображение имеет размер 32X32 пикселя.

Помимо этого, CIFAR-10 включает тестирующую выборку, состоящую из 10.000 изображений. Выборка предстает собой массив

байтовых значений. На сегодняшний момент задача классификации изображений данной выборки можно считать в сущности решенной [LeCun, 2014].

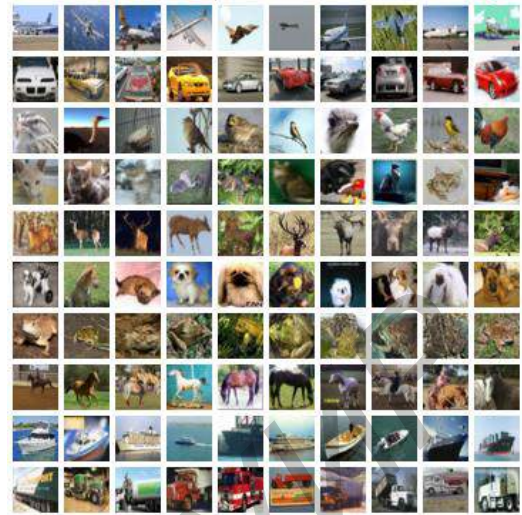


Рисунок 5 – Фрагмент базы изображений CIFAR-10

Семантическое кодирование (хеширование) относится к более сложному и не менее важному типу задач, позволяющему формировать бинарный код ограниченной длины, емко и однозначно описывающий изображение в редуцированном признаковом пространстве.

3.2. Структура сети и основные параметры обучения

Для решения задачи нами использовалась 10-слойная ассоциативная нейронная сеть с архитектурой 3072-4096-2048-1024-512-256-512-1024-2048-4096-3072. На среднем слое такой сети формируется вектор вещественных значений из отрезка $[0, 1]$, элементы которого затем округляются. Таким образом формируется бинарный код изображения. Легко видеть, что таким образом может быть закодировано 2^{256} изображений, что более чем достаточно для базы CIFAR-10.

Обучение проводилось в два этапа. На первом этапе предобучались соответствующие RBM, формирующие кодирующие слои сети. Исходные данные перед обучением были стандартизованы: из каждого компонента вектора изображения вычиталось его среднее значение по всем изображениям выборки и затем полученное значение делилось на стандартное отклонение по всем компонентам всех изображений. Данное преобразование определило тип первой обучаемой машины Больцмана – линейно-бинарная RBM. Остальные машины обучались как бинарная RBM.

Каждая RBM обучалась на протяжении 100 эпох мини-батчами по 100 элементов. Для линейно-бинарной RBM использовалась скорость обучения 0,001, для бинарной – 0,01. Помимо этого для ускорения процесса обучения использовался моментный параметр, равный 0,9.

После проведения предобучения выполнялось обучение развернутой автоассоциативной сети методом обратного распространения ошибки. Параметры обучения: скорость – $1e-6$, количество эпох обучения – 150, моментный параметр – 0.9.

После выполнения обучения, результат оценивался вычислением расстояния Хэмминга между бинарными кодами для тестового изображения и изображениями из базы (формула 10) После этого полученный ряд значений сортировался по возрастанию для выделения наиболее релевантных результатов.

$$H(v_1, v_2) = \sum_{i=1}^n |v_1^i - v_2^i| \quad (12)$$

Вычисления, производимые в процессе обучения глубокой автоассоциативной нейронной сети производились на видеокарте GTX 750 Ti и заняли около 90 минут.

3.3. Результаты обучения

Мы протестировали нашу модель двумя способами. Первый вариант предусматривал визуальное сравнение оригинального и восстановленного изображения. Некоторые из выполненных тестов представлены на рисунке 6.



Рисунок 6 – Оригинальные и восстановленные нейронной сетью изображения

Второй вариант тестирования заключался в подаче на обученную автоассоциативную сеть изображения, получения его бинарного кода с промежуточного слоя сети и вычисления расстояния Хэмминга для всех остальных изображений. Отсортировав получившуюся последовательность по возрастанию, можно изучить наиболее релевантные результаты поиска (рис. 7-8).

Исходя из представленных данных, можно отметить, что с увеличением расстояния Хэмминга количество изображений того же класса, что и

целевое изображение постепенно уменьшается.

	43	61	61
	64	64	65
	65	68	69

Рисунок 7 – Поискный шаблон и результаты поиска с расстояниями Хэмминга (тест №1)

	56	62	63
	66	70	72
	74	74	75

Рисунок 8 – Поискный шаблон и результаты поиска с расстояниями Хэмминга (тест №2)

Заключение

В последнее время в мире наблюдается активизация исследований в области нейронных сетей глубокого доверия. Эта безусловно перспективная область занимает достойное место в теории нейронных сетей. Из года в год увеличивается количество публикаций по этой тематике. Появляются целые книги, посвященные глубокому обучению (например, [Masters, 2015]). При этом особое значение приобретают скорость обучения таких сетей. Гетерогенные вычисления

уже давно стали стандартом де-факто при обучении таких типов сетей. При реализации практической части этой работы ключевое значение имело использование видеокарты для выполнения вычислительной работы. Это позволило существенно ускорить процесс обучения и провести большое количество экспериментов за приемлемое время.

В данной работе предлагается подход к семантическому кодированию изображений. На основе данного подхода может быть реализована полноценная поисковая система изображений. На текущем этапе реализованная простейшая система поиска позволяет находить релевантные результаты по поисковому запросу, сформированному в виде целевого изображения. Хочется отметить, что уровень pertinентности системы пока невысок. Его повышение является дальнейшим направлением работы.

Библиографический список

- [Krizhevsky et al., 2011] Krizhevsky, A. Using Very Deep Autoencoders for Content-Based Image Retrieval / A. Krizhevsky, G. Hinton // European Symposium on Artificial Neural Networks, Bruges, Belgium
- [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.
- [Hinton et al., 1995] Hinton, G. The "wake-sleep" algorithm for unsupervised neural networks / G. Hinton, P. Dayan, B. Frey, R. Neal // Science. – 1995. – Vol. 268. – P. 1158–1161.
- [Bengio, 2009] Bengio, Y. Learning deep architectures for AI // Foundations and Trends in Machine Learning. – 2009. – Vol. 2(1). – P. 1–127
- [Головко, 2015a] Головко В.А. Применение нейронных сетей глубокого доверия для выделения семантически значимых признаков / В.А. Головко, А.А. Крощенко // Материалы V Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (Open Semantic Technologies for Intelligent Systems). – Минск : БГУИР, 2015. – с. 481–486.
- [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.
- [Головко, 2015b] Головко, В.А. От многослойных перцептронов к нейронным сетям глубокого доверия: парадигмы обучения и применение / В.А.Головко // XVII Всероссийская научно-техническая конференция «Нейроинформатика-2015»: Лекции по Нейроинформатике. – М.: НИЯУ МИФИ, 2015. – С. 47–84.
- [Головко и др., 2015c] Головко, В.А.Метод обучения нейронной сети глубокого доверия и применение для визуализации данных / В. Головко, А. Крощенко// Комп'ютерно інтегровані технології: освіта, наука, виробництво.– 2015. – № 19. – С. 6–12.
- [Golovko et al., 2015d] Golovko, Vladimir. A New Technique for Restricted Boltzmann Machine Learning / Aliaksandr Kroshchanka, Volodymyr Turchenko, Stanislaw Jankowski, Douglas Treadwell // Proceedings of the 8th IEEE International Conference

IDAACS-2015, Warsaw 24-26 September 2015. – Warsaw, 2015 – P.182–186.

[Hinton, 2009] Hinton, G. Greedy layer-wise algorithm // Journal of Machine Learning Research 1. – 2009. – P. 1–40.

[LeCun, 2014] Convolutional Nets and CIFAR-10: An Interview with Yann LeCun [Электронный ресурс]. – Режим доступа: <http://blog.kaggle.com/2014/12/22/convolutional-nets-and-cifar-10-an-interview-with-yan-lecun/>. – Дата доступа 01.11.2015.

[Masters, 2015] Masters, T. Deep Belief Nets in C++ and CUDA C: Volume 1: Restricted Boltzmann Machines and Supervised Feedforward Networks. – 1 edition. – CreateSpace Independent Publishing Platform, 2015. – 244 p.

SEMANTIC CODING BASED ON DEEP AUTOASSOCIATIVE NEURAL NETWORK

Golovko V.A., Kroshchanka A.A.,
Hackevich M.V.

Brest State Technical University

gva@bstu.by

kroschenko@gmail.com

An approach to implementation of semantic coding (hashing) using deep autoassociative neural network is proposed. The system based on the deep neural network is simple and fast concerning search relevant images.