

# РЕАЛИЗАЦИЯ ОБУЧЕНИЯ И ВЫВОДА МОДЕЛИ ОГРАНИЧЕННОЙ МАШИНЫ БОЛЬЦМАНА С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ NVIDIA CUDA

*Хомутов Никита Юрьевич*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: inikitah@gmail.com*

Глубинная Машина Больцмана (DBM) представляет собой Марковское случайное поле с бинарными переменными, граф которого представляет многослойную сеть. Совместное распределение вероятностей, задаваемое данной моделью, выражается распределением Больцмана:

$$P(\mathbf{x}|\Theta) \propto e^{-E(\mathbf{x},\Theta)},$$

$$E(\mathbf{v}, \mathbf{h}_1, \dots, \mathbf{h}_L, \Theta) = \mathbf{v}b_0^T + \mathbf{v}W_1\mathbf{h}_1^T + \dots + \mathbf{h}_{L-1}W_L\mathbf{h}_L + \mathbf{h}_Lb_L^T.$$

Здесь вектор  $\mathbf{v}$  — видимые переменные, а  $\mathbf{h}_1, \dots, \mathbf{h}_L$  — скрытые переменные соответствующих слоёв.  $\Theta = \{W_1, \dots, W_L, b_1, \dots, b_L\}$  — настраиваемые параметры модели,  $W_i$  — матрица бинарных потенциалов связи слоёв  $i$  и  $i - 1$ ,  $b_i$  — вектор унарных потенциалов для слоя  $i$ .  $E$  — функционал энергии.

Являясь одновременно генеративной и дискриминативной моделью, машина Больцмана может быть применена к задачам: распознавания образов; сегментации изображений; подавления шумов и восстановления сигналов и изображений; сжатия данных с потерями

Задача обучения состоит в максимизации правдоподобия модели на данных обучающей выборки:  $P(\Theta|\text{Data}) \rightarrow \max_{\Theta}$ . Задача вывода — в оценке  $P(\mathbf{v}|\Theta) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}|\Theta)$ , либо генерации выборки из данного распределения. Задачи обучения и вывода являются вычислительно сложными, но могут быть эффективно решены с использованием однослойных (ограниченных, RBM) машин Больцмана [3, 4].

В рамках работы были реализованы схемы обучения и вывода ограниченной машины Больцмана для CPU и GPU. Эффективные реализации её обучения методом сопряжённых градиентов [1, 2], а также её вывода используют матричные операции, требующие значительных вычислительных затрат. Были проведены эксперименты поведения модели на базе данных рукописных изображений MNIST. Время обучения на данной базе алгоритмом CD-5 [2] на CPU составило 717 сек, а на GPU — 409 сек.

## Иллюстрации

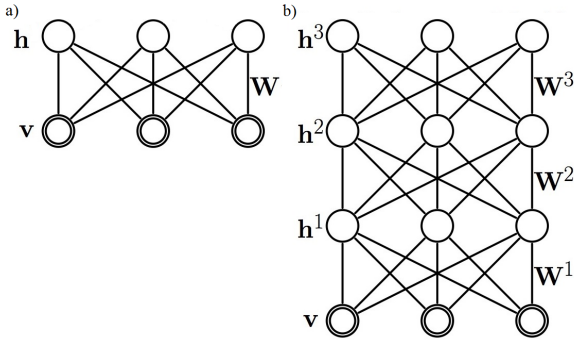


Рис. 1: Машины Больцмана: а) ограниченная; б) глубинная

## Литература

1. Bo Chen, Marlin B., de Freitas N., Swersky K. A Tutorial on Stochastic Approximation Algorithms for Training Restricted Boltzmann Machines and Deep Belief Nets. Information Theory and Applications Workshop, 2010. С. 80–89, Сан-Диего, штат Калифорния, США.
2. Carreira–Perpinan M., Hinton G. On Contrastive Divergence Learning. 10th Int. Workshop on Artificial Intelligence and Statistics (AISTATS 2005), С. 59–66, 2005, Барбадос.
3. Hinton G., Salakhutdinov R. An Efficient Learning Procedure for Deep Boltzmann Machines. MIT Technical Report MIT-CSAIL-TR-2010-037, 2010.
4. Hinton G., Salakhutdinov R. A Better Way to Pretrain Deep Boltzmann Machines. Neural Information Processing Systems Foundation Conference 2012 (NIPS 2012). С 2447–2455. озеро Тахо, штат Невада, США.