

Минимизация энергии в моделях Изинга с помощью нейронных сетей

Выпускная квалификационная работа

выполнил: *Андрей Охотин*

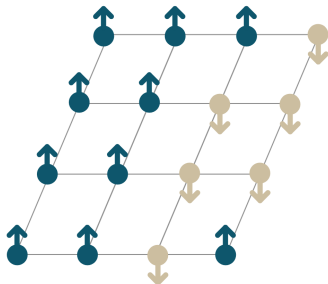
научный руководитель: *Дмитрий Кропотов*

научные консультанты: *Максим Кодрян, Дмитрий Ветров*

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

Модель Изинга

Графическая иллюстрация классической модели Изинга: квадратная решетка, в каждом узле которой находится число, называемое *спином* и равное $+1$ или -1 («поле вверх» или «поле вниз»).



Задача минимизации энергии в модели Изинга

Постановка:

$$E(x) = \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{\{i,j\} \in \mathcal{E}} \theta_{ij}(x_i, x_j) + \theta_0 \rightarrow \min_{x \in \text{Binary}}$$

– задача дискретной оптимизации

Пример

1	1	1	1	0	1
0	1	0	0	0	1
0	0	1	1	1	1
1	1	0	1	0	0
0	0	1	1	0	1
1	1	0	1	1	1

Пример оптимального решения \bar{x} .

Проблемы существующих солверов

1. каждую задачу решают заново, поэтому для одновременного решения множества задач неэффективны
2. время решения сильно зависит от размера задачи
3. долго работают

Мотивация

Преимущества использования нейронных сетей:

1. разнообразие методов построения и обучения
2. удобство масштабирования
3. простота использования

Ограничение на модели Изинга

Выполнение ограничений следующего вида на парные потенциалы:

$$(1) \quad \forall i \in \mathcal{V}: \theta_i(0) \geq 0, \quad \theta_i(1) \geq 0;$$

$$(2) \quad \forall \{i, j\} \in \mathcal{E}: \theta_{ij}(0, 0) = \theta_{ij}(1, 1) = 0;$$

$$(3) \quad \forall \{i, j\} \in \mathcal{E}: \theta_{ij}(0, 1) \geq 0, \quad \theta_{ij}(1, 0) \geq 0$$

является *достаточным* условием применимости алгоритмов разрезов графов для минимизации парно-сепарабельной энергии. С такими ограничениями задача минимизации энергии является разрешимой за полиномиальное время. Описанные условия необходимы для возможности оценить качество решения нейронной сети.

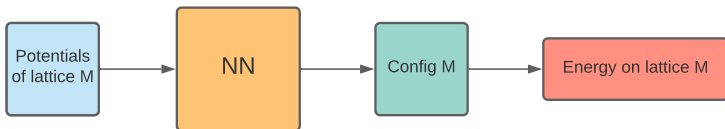
Применение задачи минимизации энергии в модели Изинга

В графических моделях, решающих задачи:

1. сегментации с семенами
2. семантической сегментации
3. анатомической разметки

Идея подхода

Подавать в нейронную сеть потенциалы модели Изинга (условие задачи минимизации) и получать на выходе конфигурацию с минимальной энергией (решения задачи минимизации).



Энергия конфигурации считается на решетке, чьи потенциалы были поданы в нейронную сеть.

Идея подхода

Формальное описание подхода с использованием нейронных сетей:

$$\mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} E [f_\phi(\theta^U, \theta^V, \theta^H), \theta^U, \theta^V, \theta^H] \rightarrow \min_\phi, \quad (1)$$

где $f_\phi : \{\theta^U, \theta^V, \theta^H\} \rightarrow p^\phi$ – нейронная сеть

$$\mathbb{P}[x_{i,j} = 1] = p_{i,j}, \quad i, j = \overline{1, n},$$

Конкретная конфигурация получается из вероятностей округлением.

Релаксация функционала энергии

$$\begin{aligned}
 & \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \mathbb{E}_{f_\phi(\theta^U, \theta^V, \theta^H)} E \left[f_\phi(\theta^U, \theta^V, \theta^H), \theta^U, \theta^V, \theta^H \right] = \\
 & = \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \mathbb{E}_{p^\phi} E \left[p^\phi, \theta^U, \theta^V, \theta^H \right] = \\
 & = \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \left[\mathbb{E}_{p^\phi} \text{Unary}(x, \theta^U) + \mathbb{E}_{p^\phi} \text{Paired}(x, \theta^V, \theta^H) \right] =
 \end{aligned}$$

Релаксация функционала энергии

$$\begin{aligned}
&= \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \left[\mathbb{E}_{p^\phi} \left(\sum_{i,j=1}^{n,n} [\theta_{i,j}^U(0)(1 - x_{i,j}^\phi) + \theta_{i,j}^U(1)x_{i,j}^\phi] \right) + \right. \\
&\quad + \mathbb{E}_{p^\phi} \left(\sum_{i,j=1}^{n-1,n} [\theta_{i,j}^V(0,1)(1 - x_{i,j}^\phi)x_{i+1,j}^\phi + \theta_{i,j}^V(1,0)x_{i,j}^\phi(1 - x_{i+1,j}^\phi)] \right) + \\
&\quad \left. + \mathbb{E}_{p^\phi} \left(\sum_{i,j=1}^{n,n-1} [\theta_{i,j}^H(0,1)(1 - x_{i,j}^\phi)x_{i,j+1}^\phi + \theta_{i,j}^H(1,0)x_{i,j}^\phi(1 - x_{i,j+1}^\phi)] \right) \right] = \\
&= \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \left[\sum_{i,j=1}^{n,n} [\theta_{i,j}^U(0)(1 - p_{i,j}^\phi) + \theta_{i,j}^U(1)p_{i,j}^\phi] + \right. \\
&\quad + \sum_{i,j=1}^{n-1,n} [\theta_{i,j}^V(0,1)(1 - p_{i,j}^\phi)p_{i+1,j}^\phi + \theta_{i,j}^V(1,0)p_{i,j}^\phi(1 - p_{i+1,j}^\phi)] + \\
&\quad \left. + \sum_{i,j=1}^{n,n-1} [\theta_{i,j}^H(0,1)(1 - p_{i,j}^\phi)p_{i,j+1}^\phi + \theta_{i,j}^H(1,0)p_{i,j}^\phi(1 - p_{i,j+1}^\phi)] \right] = \\
&= \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} E \left(p^\phi, \theta^U, \theta^V, \theta^H \right) = \mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} \mathbb{E}_{f_\phi} E \left[f_\phi(\theta^U, \theta^V, \theta^H), \theta^U, \theta^V, \theta^H \right]
\end{aligned}$$

В каких моделях Изинга минимизировать энергию?

Распределение моделей Изинга, в котором будет минимизироваться энергия:

$$p(\theta^U, \theta^V, \theta^H | \psi)$$

Параметры ψ определяют свойства моделей Изинга, в которых будет минимизироваться энергия (подробнее далее).

Постановки

Обучение без учителя

$$\mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} E(p^\phi, \theta^U, \theta^V, \theta^H) \rightarrow \min_\phi$$

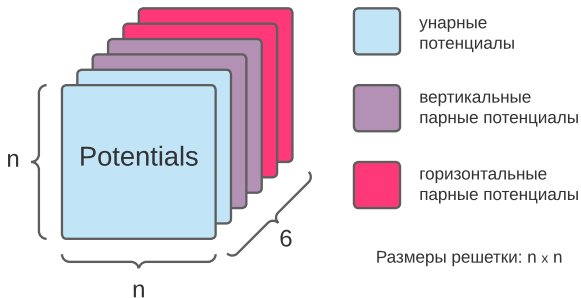
Обучение с учителем

$$x_{optimal}(\theta^U, \theta^V, \theta^H) = \operatorname{argmin}_{x \in \text{Binary}^{n \times n}} E(x, \theta^U, \theta^V, \theta^H)$$

$$CE(x, p^\phi) = -\frac{1}{n^2} \sum_{i,j=1}^n [x_{i,j} \log p_{i,j} + (1 - x_{i,j}) \log(1 - p_{i,j})]$$

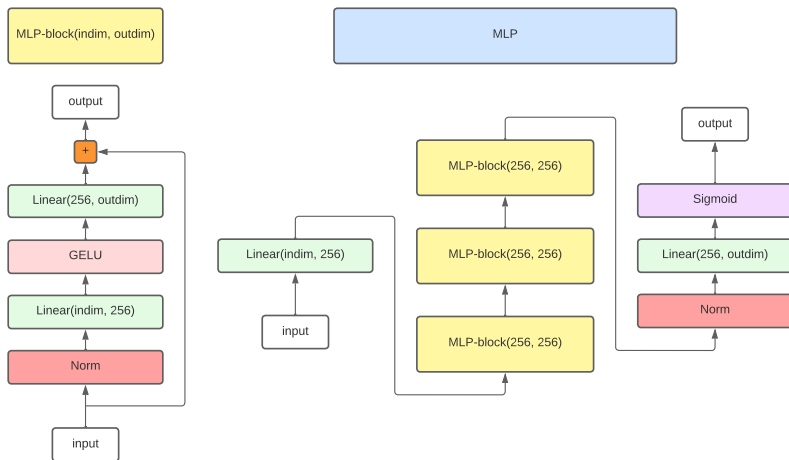
$$\mathbb{E}_{p(\theta^U, \theta^V, \theta^H | \psi)} CE(x_{optimal}(\theta^U, \theta^V, \theta^H), p^\phi) \rightarrow \min_\phi$$

Данные



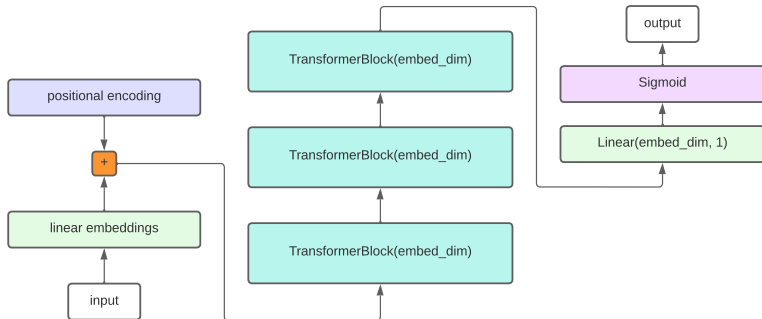
Тензор из потенциалов, подающийся в нейронную сеть.

MLP



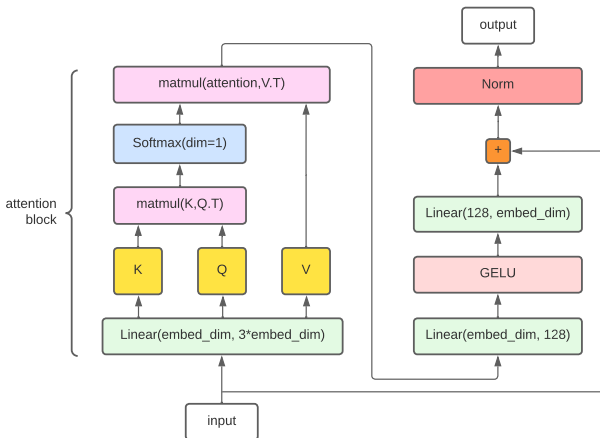
Архитектура Multilayer Perceptron, MLP.

Transformer



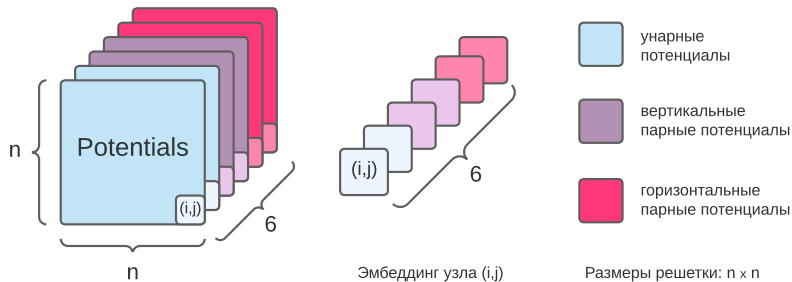
Архитектура Transformer.

Transformer



Архитектура TransformerBlock.

Transformer



Эмбединги узлов.

Embeddings

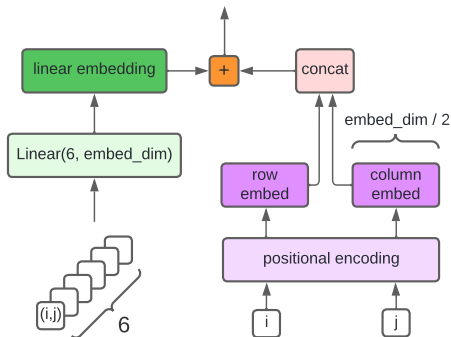


Схема создания эмбеддингов для нейронной сети вида Transformer.

Характер задачи минимизации энергии

Функционал энергии в модели Изинга распадается на линейную и квадратичную компоненты:

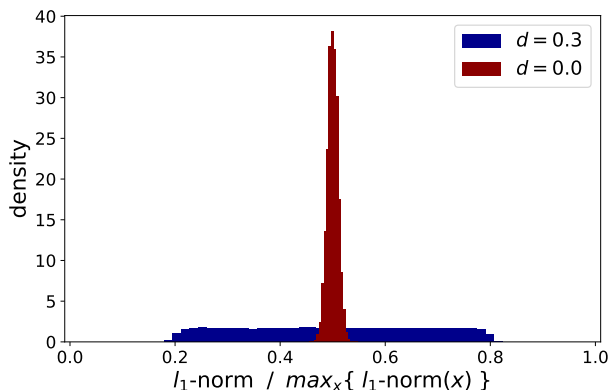
$$\mathbb{E}_{x^{\mathbb{R}}} E(x^{\mathbb{R}}, \theta^U, \theta^V, \theta^H) = \text{Unary}(x^{\mathbb{R}}, \theta^U) + \text{Paired}(x^{\mathbb{R}}, \theta^V, \theta^H)$$

$x^{\mathbb{R}}$ – случайные конфигурации из некоторого распределения

Насколько задача минимизации энергии является линейной можно оценить следующей величиной:

$$\text{balance}(\theta^U, \theta^V, \theta^H) = \mathbb{E}_{x^{\mathbb{R}}} \left[\frac{\text{Unary}(x^{\mathbb{R}}, \theta^U)}{\text{Unary}(x^{\mathbb{R}}, \theta^U) + \text{Paired}(x^{\mathbb{R}}, \theta^V, \theta^H)} \right]$$

Распределение случайных конфигураций



$$x_{i,j}^{\mathbb{R}} = \text{Bernoulli}(b), \quad i, j = \overline{1, n}$$

$$b \sim \text{Uniform}\left(\frac{1}{2} - d, \frac{1}{2} + d\right), \quad d = 0.3$$

Принципы построения распределения моделей Изинга

Теперь мы можем перейти от рассмотрения параметров в распределении

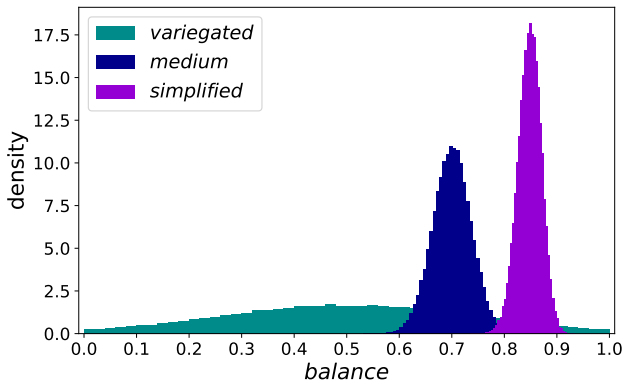
$$p(\theta^U, \theta^V, \theta^H | \psi),$$

к распределению

$$p(\text{balance}(\theta^U, \theta^V, \theta^H) | \psi),$$

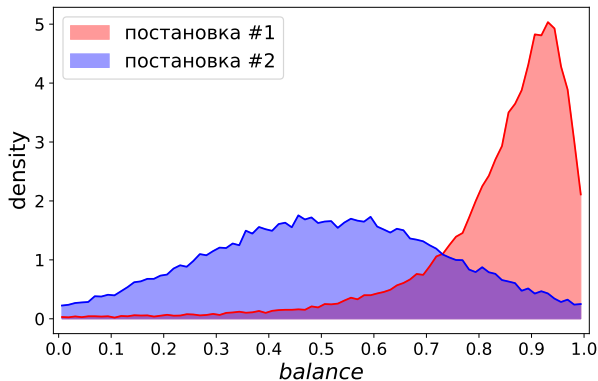
в котором сможем оценить вариативность задачи минимизации энергии в шкале $\text{balance} \in [0, 1]$: от полностью квадратичной ($\text{balance} = 0$) задачи до полностью линейной ($\text{balance} = 1$).

Выборки с различными свойствами



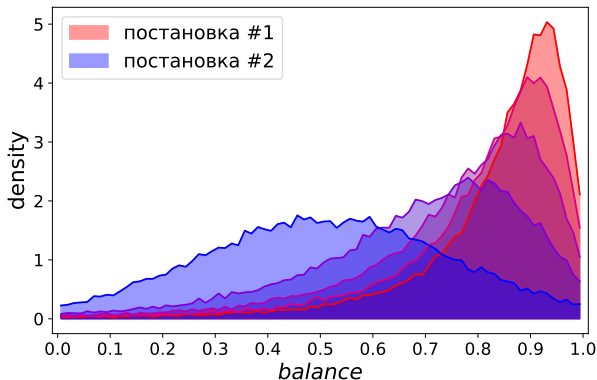
Сравнение распределений $p(\text{balance}(\theta^U, \theta^V, \theta^H) \mid \psi)$ для наборов данных *simplified*, *medium* и *variegated*.

Fine-tuning



Постановки задач для нейронной сети в случае использования fine-tuning.

Relaxed fine-tuning



Постановки задач для нейронной сети в случае использования relaxed fine-tuning.

Лучшие результаты

Architecture train loss	Dataset Types			Metrics
	simplified	medium	variegated	
MLP BCEloss	9.777 ± 3.038 91.034 ± 6.291	23.011 ± 9.887 87.192 ± 12.357	21.338 ± 33.347 89.860 ± 19.157	energy accuracy, %
Transformer BCEloss	2.556 ± 2.585 96.256 ± 4.776	13.012 ± 11.906 91.817 ± 10.354	11.745 ± 21.117 93.811 ± 13.216	energy accuracy, %
MLP Energy	1.128 ± 1.449 95.477 ± 5.481	2.338 ± 3.124 91.347 ± 13.317	1.872 ± 3.513 81.110 ± 34.274	energy accuracy, %
Transformer Energy	1.378 ± 1.712 94.375 ± 6.129	3.577 ± 4.459 89.419 ± 13.143	2.408 ± 4.571 80.437 ± 33.126	energy accuracy, %

В ячейках таблицы отображены $\text{mean} \pm \text{std}$. Достижимый минимум энергии равен 0, математическое ожидание энергии случайных конфигураций равно 100 (т.е. если случайно выбирать x как решения задачи, то их энергия будет в среднем около 100).

Сравнение качества на выборках из других распределений

Train Dataset	Test Dataset types			Metrics
	simplified	medium	variegated	
simplified	2.556 ± 2.585	19.957 ± 15.054	57.208 ± 47.267	energy
	96.256 ± 4.776	82.525 ± 14.275	65.459 ± 25.405	accuracy, %
medium	6.994 ± 4.444	13.012 ± 11.906	37.239 ± 45.971	energy
	90.746 ± 10.354	91.817 ± 10.354	86.047 ± 23.428	accuracy, %
variegated	7.179 ± 3.575	22.100 ± 12.352	11.745 ± 21.117	energy
	91.939 ± 6.295	87.736 ± 11.475	93.811 ± 13.216	accuracy, %

Сравнение моделей Transformer обученных с BCEloss на описанных датасетах. Сравнение происходит на всех видах выборок. В ячейках таблицы отображены mean ± std. Достижимый минимум энергии равен 0, математическое ожидание энергии случайных конфигураций равно 100 (т.е. если случайно выбирать x как решения задачи, то их энергия будет в среднем около 100).

Сравнение качества на выборках из других распределений

Train Dataset	Test Dataset types			Metrics
	simplified	medium	variegated	
simplified	1.128 \pm 1.449	4.508 \pm 4.892	11.076 \pm 14.131	energy accuracy, %
	95.477 \pm 5.481	89.233 \pm 11.964	76.175 \pm 28.522	
medium	5.160 \pm 4.346	2.338 \pm 3.124	2.005 \pm 5.134	energy accuracy, %
	88.685 \pm 8.628	91.347 \pm 13.317	87.408 \pm 26.697	
variegated	3.642 \pm 2.985	4.837 \pm 5.136	1.872 \pm 3.513	energy accuracy, %
	91.567 \pm 6.792	88.679 \pm 13.928	81.110 \pm 34.274	

Сравнение моделей MLP обученных с функционалом потерь Energy на описанных датасетах. Сравнение происходит на всех видах выборок. В ячейках таблицы отображены $\text{mean} \pm \text{std}$. Достижимый минимум энергии равен 0, математическое ожидание энергии случайных конфигураций равно 100 (т.е. если случайно выбирать x как решения задачи, то их энергия будет в среднем около 100).

Список литературы

[1] Теоретическая основа. Модель Изинга.

http://www.machinelearning.ru/wiki/images/9/9e/GM_graphCuts.pdf

[2] Алгоритмы минимизации энергии на основе разрезов графов

<http://www.machinelearning.ru/wiki/images/a/aa/GraphCutNotes.pdf>

[3] Xavier Bresson, Thomas Laurent "The Transformer Network for the Traveling Salesman Problem" 2021

[4] Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, Alexey Dosovitskiy "MLP-Mixer: An all-MLP Architecture for Vision"

<https://arxiv.org/pdf/2105.01601.pdf>

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser "Attention Is All You Need" <https://arxiv.org/pdf/1706.03762.pdf>