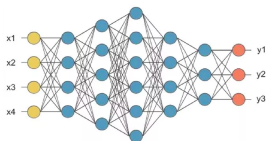


WANNs: metrics and genetic algorithms

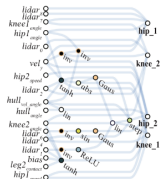
D.Skachkov, M.Sodikov

MIPT

2020



((a)) WANN №1



((b)) WANN №2

Weight Agnostic Neural Network (WANN) или Агностическая нейронная сеть - нейронная сеть, у которой веса всех связей одинаковы. В данной статье рассмотрены WANN'ы двух типов: WANN №1 - обычная полносвязная нейронная сеть, нейроны могут быть соединены только если они расположены в соседних слоях.

WANN №2 - нейронная сеть, у которой присутствуют только входной и выходной слои, остальные нейроны могут быть соединены как угодно, главное, чтобы не было направленных циклов.

Problem statement

- ▶ (X, y) , $X \in \mathbb{R}^{n \times m}$, $y \in \mathbb{R}^{n \times k}$, строки матриц X и y находятся в некоторой неизвестной функциональной зависимости
- ▶ \mathfrak{G} - множество функций активации
- ▶ $\tilde{\mathfrak{G}}$ - множество агностических нейронных сетей, у которых функции активации содержатся в \mathfrak{G}
- ▶ $(f, w) = \arg \min_{(f, w) \in \tilde{\mathfrak{G}} \times \mathbb{W}} S(f^w(x), y)$, где S - функция ошибки
- ▶ в данной работе $S(y_{pred}, y_{true}) = \|y_{pred} - y_{true}\|_2$

Classic genetic algorithm

Algorithm 1 Basic genetic algorithm

Require: grammar \mathfrak{G} , required value α of MAP

Ensure: superposition f of elements from G with $\text{MAP} \leq \alpha$;

create a set of initial, random superpositions \mathfrak{M}_0 ,

repeat

- crossover random pairs of stored superpositions \mathfrak{M} ,
- mutate random superpositions from the population \mathfrak{M} ,
- consider these generated superpositions and the ones stored in \mathfrak{M} . Select the best of them according to MAP,
- store the best generated superpositions in the population \mathfrak{M} and pass it to the next iteration,

until the required value of MAP is reached;

Modified genetic algorithm: detecting stagnations

Algorithm 2 Modified genetic algorithm

Require: grammar \mathfrak{G} , required value α of MAP

Ensure: superposition f of elements from G with $\text{MAP} \leq \alpha$;

create a set of initial, random superpositions \mathfrak{M}_0 ,

repeat

- crossover random pairs of stored superpositions \mathfrak{M} ,
- mutate random superpositions from the population \mathfrak{M} ,
- consider these generated superpositions and the ones stored in \mathfrak{M} . Select the best of them according to the quality function \mathcal{S} (2),
- store the best superpositions in a population \mathfrak{M}' and pass it to the next iteration,
- **if** $d_e(\mathfrak{M}') < \text{Thresh}$ **then**
 - evolutionary stagnation is detected and we replace the worst superpositions from the population \mathfrak{M}' by random superpositions,
- **end if**
- $\mathfrak{M} = \mathfrak{M}'$.

until the required value of MAP is reached;

Genetic algorithm: more details

Напомним, что нашей исходной задачей являлось отыскание пары (f, w) , т.ч.

$$(f, w) = \arg \min_{(f, w) \in \mathfrak{F} \times \mathbb{W}} S(f^w(x), y)$$

Тем не менее мы считаем слишком затратным на каждой итерации генетического алгоритма точно подбирать параметр. Поэтому используются 3 различных функции ошибки, по значению которых происходит отбор в конце каждой эпохи:

- ▶ $\min_{w \in \mathbb{W}'} S(f^w(x), y)$
- ▶ $\frac{1}{|\mathbb{W}'|} \sum_{w \in \mathbb{W}'} S(f^w(x), y)$
- ▶ $\max_{w \in \mathbb{W}'} S(f^w(x), y)$

Здесь \mathbb{W}' - некоторое небольшое множество возможных значений параметра (в данной работе $\{-2, -1, -0.5, 0.5, 1, 2\}$). В конце работы ген. алгоритма полученный WANN оптимизируется по параметру.

Standart metrics

Эталонные метрики - метрики, задаваемые из некоторых эмпирических соображений.

В данной работе вводятся две таких метрики:

$$\blacktriangleright \eta_1(f_1, f_2) = \frac{1}{n|W|} \sum_{w \in W} \sum_{i=1}^n \|f_1^w(x_i) - f_2^w(x_i)\|_2$$

$$\blacktriangleright \eta_2(f_1, f_2) = \max_{w_1, w_2 \in W} \frac{1}{n} \sum_{i=1}^n \|(f_1^{w_1}(x_i) - f_2^{w_2}(x_i))\|_2$$

f_j - WANN

f_j^w - функция, задаваемая WANN при значении параметра равном w

Comparing metrics

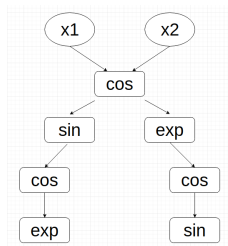
Пусть известны значения метрик μ и η на $T \times T$, T - некоторое конечное множество. Методы сравнения метрик:

- ▶ Коэффициент корреляции Пирсона
- ▶ $\nu(\mu(t_1, t_2) < \mu(t_1, t_3) \wedge \eta(t_1, t_2) > \eta(t_1, t_3)), \nu(A)$ - частота события A
- ▶
$$\frac{\inf \frac{\mu(t_1, t_2)}{\eta(t_1, t_2)}}{\sup \frac{\mu(t_1, t_2)}{\eta(t_1, t_2)}}$$
- ▶ $\|v_\mu - v_\eta\|_2$, где v_μ, v_η - векторизованные нормализованные матрицы попарных расстояний

Structural metrics

Структурная метрика - метрика, зависящая только от структуры модели. В идеале как можно ближе к эталонной.

Convert WANN to matrix: example



$$\begin{bmatrix} & \sin & \cos & \exp \\ \sin & 0 & 1 & 0 \\ \cos & 2 & 0 & 2 \\ \exp & 0 & 1 & 0 \\ x_1 & 0 & 1 & 0 \\ x_2 & 0 & 1 & 0 \end{bmatrix}$$

Structural metric №1

Пусть f_1, f_2 - агностические нейронные сети, M_1, M_2 - построенные по ним матрицы. Тогда

$$\mu_1(f_1, f_2) = \|v_1 - v_2\|_2$$

v_1, v_2 - векторизованные M_1, M_2 .

Generalization

Пусть дана пара $(f_1, f_2) \in \mathfrak{F}_{\mathfrak{G}} \times \mathfrak{F}_{\mathfrak{G}}$. Скажем, что $v_{(f_1, f_2)} = \text{abs}(v_1 - v_2)$, где v_1, v_2 из предыдущего слайда, abs - покомпонентный модуль, - признаковое описание объекта (f_1, f_2) .

Цель: найти $g : \{v_{(f, f')} | f, f' \in \mathfrak{F}_{\mathfrak{G}}\} \longrightarrow \mathbb{R}_{\geq 0}$, которое как можно лучше согласовывается с эталонной метрикой.

В структурной метрике №1 $g(v) = \|v\|_2$

Вопрос: можно ли лучше?

Structural metric №2

Для ответа на поставленный вопрос подходящая функция g ищется с помощью обычной полносвязной нейростети с несколькими скрытыми слоями, которая по входному вектору признаков учится предсказывать одну из эталонных метрик.

Genetic algorithm modification

Result: WANN

```
generate_population();
```

```
while  $error > \varepsilon$  or  $iteration < max\_iterations$  do
```

```
    mutate();
```

```
    crossover(); //for WANN №1;
```

```
    select();
```

```
    if  $\frac{error[iteration-1]}{error[iteration]} \geq \delta$  then
```

```
        for  $wann$  in  $population[\frac{n}{2} :]$  do
```

```
            regenerate( $wann$ );
```

```
        end
```

```
    end
```

```
end
```

```
return population[0];
```

Algorithm 1: Генетический алгоритм с детектированием стагнаций по значению ошибки

Genetic algorithm modification

Result: WANN

generate_population();

while $error > \varepsilon$ or $iteration < max_iterations$ **do**

 mutate();

 crossover(); //for WANN №1;

 select();

if $\frac{error[iteration-1]}{error[iteration]} \geq \delta$ **then**

 clusters := cluster(population);

for cluster in clusters **do**

for wann in cluster \ {best_in_cluster} **do**

 wann.regenerate();

end

end

end

end

return population[0];

Algorithm 2: Генетический алгоритм с кластеризацией

Results: structural metric №2

Нейросеть net_j^i обучалась предсказывать эталонную метрику j для WANN № i .

С помощью каждого из четырех способов сравнения метрик была выбрана метрика, наиболее близкая к эталонным.

	1 способ	2 способ	3 способ	4 способ
standart 1	structural 1	structural 1	net_1^1	structural 1
standart 2	structural 1	structural 1	net_1^2	net_1^2

Таблица: Соответствие между эталонными и структурными метриками для WANN №1

	1 способ	2 способ	3 способ	4 способ
standart 1	structural 1	structural 1/ net_2^1	—	net_2^1
standart 2	structural 1	—	net_2^2	structural 1

Таблица: Соответствие между эталонными и структурными метриками для WANN №2

Experiment description: genetic algorithms

Все 4 модификации гентического алгоритма была запущены по 100 раз с каждой из трех функций ошибки со следующими параметрами:

- ▶ 100 эпох
- ▶ размер популяции - 10
- ▶ количество мутаций за одну эпоху - 5/10 (WANN №1/WANN №2)
- ▶ количество скрещиваний 5/0 (WANN №1/WANN №2)

Результат усреднялся по всем 100 экспериментам.

Results: genetic algorithms

error	min	mean	max
classic	0.70	0.37	0.33
diameter	0.70	0.36	0.33
error	0.57	0.39	0.32
clustering	0.61	0.37	0.33

Таблица: Accuracies for WANN №1

error	min	mean	max
classic	0.88	0.77	0.64
diameter	0.87	0.77	0.63
error	0.88	0.81	0.62
clustering	0.88	0.80	0.65

Таблица: Accuracies for WANN №2

Conclusion

Улучшить результаты классического генетического алгоритма не удалось. Напротив, замена половины популяции в случае стагнаций на случайные WANN'ы в большинстве случаев даже ухудшило результат.