

# Zastosowanie optymalizacji rojem cząstek (PSO) w procesie uczenia wielowarstwowej sieci neuronowej w problemie lokalizacyjnym, kontynuacja badań

Jan Karwowski

Wydział Matematyki i Nauk Informacyjnych PW

25 III 2014





Jan Karwowski, Michał Okulewicz, Jarosław Legierski (Orange Labs).  
Application of particle swarm optimization algorithm to neural  
network training process in the localization of the mobile terminal.  
*Engineering Applications of Neural Networks*, strony 122–131.  
Springer, 2013.



# Problem (poprzednia część)

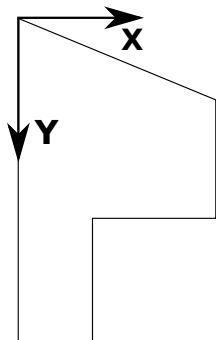
## Wejście

- Wektor sił sygnałów z punktów dostępowych WiFi, 107 punktów dostępowych w budynku.
- Wektor sił sygnałów BTS sieci komórkowych

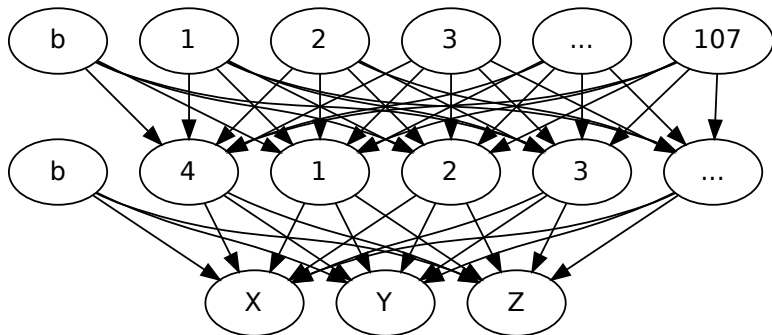
Około 1200 punktów pomiarowych, 40 pomiarów w punkcie w serii. 3 serie.

## Wyjście

- Współrzędne XYZ
- Numer pomieszczenia
- Piętro



# MLP



## BP

Propagacja wsteczna błędów, aktualizacja wag po każdej prezentacji wzorca, w każdej iteracji zmiana kolejności wzorców.



M Clerc, i in.

Standard PSO 2011.

<http://www.particleswarm.info/>, Particle Swarm Central, 2011.

Dwufazowy proces uczenia sieci:

- 1) znajdowanie wstępnych wag sieci za pomocą niedużej liczby iteracji PSO
- 2) uczenie za pomocą BP, startując z wag połączeń znalezionych przez PSO.



Dodanie fazy PSO:

- Lepsze wyniki dla dużych sieci z dużą liczbą próbek
- Lepsze wyniki dla głębokich (wiele warstw) architektur MLP
- Możliwość używania zarówno z regresją jak i klasyfikacją (sprawdzone głównie z regresją)



- Testowanie na innych problemach niż lokalizacja w MiNI
- Inne funkcje błędu niż MSE
- Regularyzacja wag i usuwanie krawędzi




# Zbiory benchmarkowe


UCI:

- iris
- glass
- new-thyroid

microarrays:

- breast2
- lymphoma

 K. Bache, M. Lichman, UCI machine learning repository, 2013. URL: <http://archive.ics.uci.edu/ml>.

 K. Moorthy, M. S. Mohamad, Random forest for gene selection and microarray data classification, Bioinformatics (2011) 142–146. URL: <http://www.bioinformatics.net/007/97320630007142.pdf>.

Odmienna charakterystyka zbiorów. Podział na 3 części.





- 30 podziałów zbioru na uczący, testowy, walidacyjny, proporcja: 60:20:20, proporcja klas taka jak w całości. Ten sam zestaw podziałów w każdym eksperymencie
- 2 powtórzenia dla danego podziału, dla każdego zestawu parametrów PSO
- Prezentowane wyniki: Średnia ze wszystkich przebiegów dla każdej pary (zbiór, parametry uczenia)



# Architektury sieci (pełne połączenia)

- breast2 → 4868—8—2
- 3lbreast2 → 4868—15—15—15—2
- glass → 9—8—6
- 3lglass → 9—8—8—8—6
- thyroid → 5—4—3



# Wyniki z grudnia

| Data        | structure         |                   | structure         |                   |
|-------------|-------------------|-------------------|-------------------|-------------------|
|             | BP                | PSO+BP            | BP                | PSO+BP            |
| iris        | 3                 |                   | 10-6              |                   |
|             | $95.52 \pm 6.41$  | $95.30 \pm 6.60$  | $49.79 \pm 33.67$ | $94.43 \pm 8.04$  |
| new-thyroid | 8                 |                   | 10-10             |                   |
|             | $96.39 \pm 4.61$  | $95.67 \pm 5.78$  | $72.26 \pm 11.34$ | $95.51 \pm 5.61$  |
| glass       | 3                 |                   | 10-3              |                   |
|             | $59.33 \pm 11.12$ | $61.19 \pm 11.13$ | $32.16 \pm 9.99$  | $54.78 \pm 11.51$ |
| breast2     | 4                 |                   | 20-5              |                   |
|             | $60.70 \pm 15.31$ | $63.27 \pm 11.64$ | $57.90 \pm 16.77$ | $59.25 \pm 15.24$ |
| lymphoma    | 8                 |                   | 10-5              |                   |
|             | $95.30 \pm 8.82$  | $99.52 \pm 2.61$  | $95.58 \pm 7.85$  | $95.63 \pm 8.58$  |

Tabela: Skuteczność klasyfikacji na zbiorach benchmarkowych. Wszystkie dane w % próbek na zbiorze testowym



# Inne funkcje błędu

$a_i$  – otrzymane wyjście,  $d_i$  – spodziewane wyjście

logloss

$$e = \sum_i -\log(1 - |d_i - a_i|)$$

Różnica ze oczekiwaną klasą

Tylko klasyfikacja kodowana jako wektory zer i jedynki!!

$c \leftarrow$  oczekiwana klasa

$t$  parametr: próg pełnej akceptacji

$$e = \max_{i \neq c} \log \min\{t, (a_i - a_c)\} + 1$$

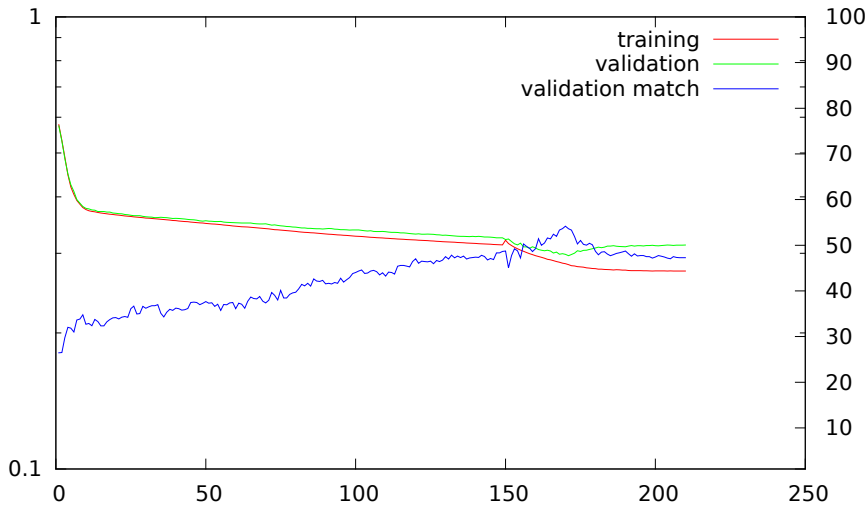


|         | breast2         | 3lbreast2       | glass           | 3lglass         | thyroid        |
|---------|-----------------|-----------------|-----------------|-----------------|----------------|
| BP      | $58.5 \pm 17.2$ | $56.8 \pm 18.1$ | $44.4 \pm 14.9$ | $37.8 \pm 13.6$ | $96.4 \pm 5.3$ |
| logloss | $56.8 \pm 17.4$ | $57.5 \pm 15.7$ | $54.2 \pm 11.4$ | $51.2 \pm 10.1$ | $95.9 \pm 4.4$ |
| MSE     | $61.8 \pm 17.0$ | $59.5 \pm 13.8$ | $55.7 \pm 11.1$ | $50.4 \pm 11.2$ | $96.4 \pm 4.4$ |
| matche  | $59.1 \pm 15.4$ | $60.1 \pm 18.4$ | $50.4 \pm 11.4$ | $31.6 \pm 10.1$ | $96.0 \pm 5.0$ |

Tabela: Eksperyment 4.6x, BP – uczenie tylko propagacją wsteczną, pso\* – uczenie najpierw PSO, a potem BP



3lglass-pso5-bp1 51.30952 11.87383



# Regularyzacja

suma modułów

$$\sum_{w_i \in \mathbb{W}} |w_i|$$

suma kwadratów

$$\sum_{w_i \in \mathbb{W}} w_i^2$$

bez rozróżnienia na duże i bardzo duże wagi

$$\sum_{w_i \in \mathbb{W}} \frac{\alpha w_i^2}{1 + \alpha w_i^2}$$



# Wyniki z regularyzacją

200 iteracji PSO

```
pso_weightsPenalty.class=\  
tuna.logic.AbsoluteWeightPenaltyFunction  
pso_weightsPenalty.penaltyCoefficient=.5
```

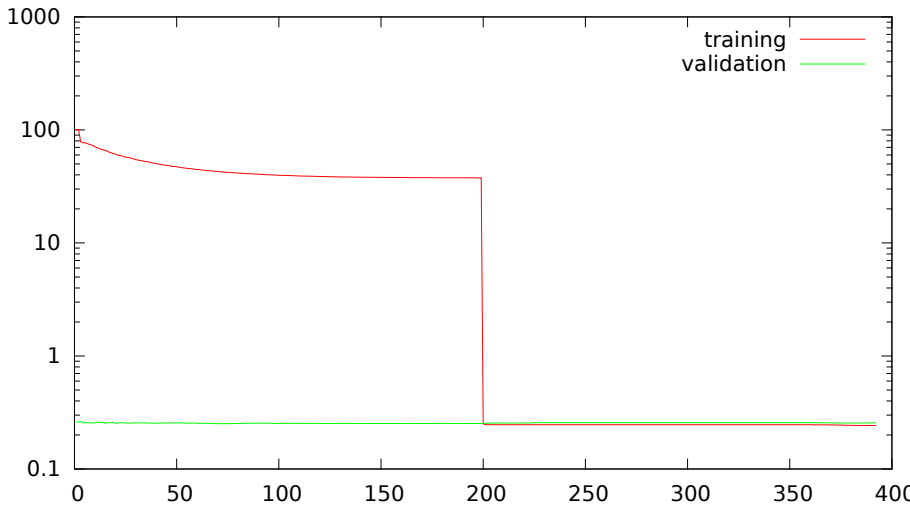
|         | breast2     | 3lbreast2   | glass       | 3lglass     | thyroid    |
|---------|-------------|-------------|-------------|-------------|------------|
| BP      | 58.6 ± 18.6 | 54.5 ± 19.8 | 39.1 ± 14.3 | 31.7 ± 9.9  | 97.0 ± 4.3 |
| logloss | 62.2 ± 16.3 | 54.3 ± 20.0 | 52.9 ± 11.6 | 31.0 ± 9.5  | 97.9 ± 2.5 |
| MSE     | 63.0 ± 17.1 | 56.1 ± 19.5 | 47.2 ± 13.2 | 31.6 ± 10.1 | 97.1 ± 4.0 |
| matche  | 61.7 ± 16.7 | 53.8 ± 20.1 | 43.3 ± 15.2 | 33.3 ± 10.3 | 97.0 ± 4.0 |

Tabela: Eksperyment 4.5t, BP – uczenie tylko propagacją wsteczną, pso\* – uczenie najpierw PSO, a potem BP





3lbreast2-pso5-bp2 56.10119 19.47094



# Usuwanie krawędzi w PSO

- Problem PSO – wiele cząstek
- Usunięcie krawędzi  $\leq t$  na koniec przebiegu
- Usuwanie krawędzi co  $k$  iteracji



# Wyniki regularyzacja+usuwanie (dwa warianty bp)

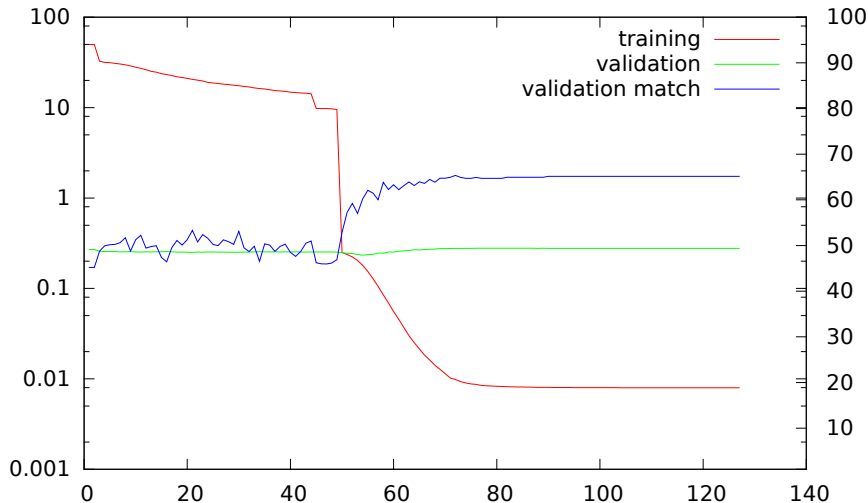
|         | breast2     | 3lbreast2 | glass       | 3lglass | thyroid    |
|---------|-------------|-----------|-------------|---------|------------|
| BP      | 59.8 ± 16.8 | X         | 40.4 ± 14.8 | X       | 95.6 ± 6.0 |
| logloss | 62.3 ± 17.1 | X         | 52.8 ± 12.4 | X       | 96.7 ± 4.1 |
| MSE     | 62.3 ± 16.4 | X         | 49.2 ± 14.1 | X       | 95.4 ± 6.4 |
| matche  | 63.0 ± 16.9 | X         | 46.3 ± 15.6 | X       | 95.5 ± 5.5 |

Tabela: Eksperyment 4.5e, BP – uczenie tylko propagacją wsteczną, pso\* – uczenie najpierw PSO, a potem BP

|         | breast2     | 3lbreast2 | glass       | 3lglass | thyroid    |
|---------|-------------|-----------|-------------|---------|------------|
| BP      | 62.4 ± 16.0 | X         | 41.1 ± 13.1 | X       | 95.9 ± 4.7 |
| logloss | 67.1 ± 15.2 | X         | 48.9 ± 9.3  | X       | 96.7 ± 4.4 |
| MSE     | 62.2 ± 14.4 | X         | 48.6 ± 11.7 | X       | 95.7 ± 4.8 |
| matche  | 62.0 ± 16.3 | X         | 48.6 ± 12.2 | X       | 96.0 ± 4.4 |



breast2-pso4-bp1 67.1131 15.22804



Zbiór walidacyjny!



# Zmiana współczynnika karania za wagi w czasie

- Liniowy wzrost współczynnika od 0 w pierwszej iteracji do 1 w ostatniej
- Modyfikacje PSO

Eksperyment 4.6h:  $p: 0 \rightarrow 0.11$

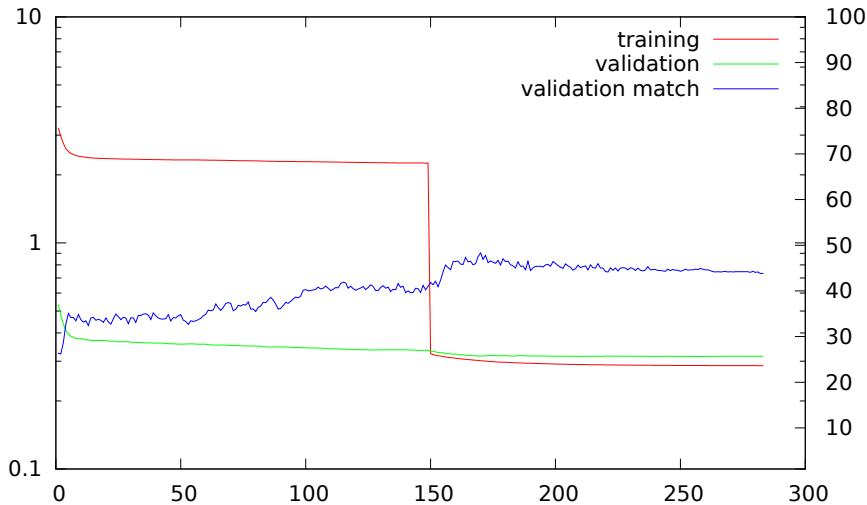
Funkcja kary: suma modułów  
150 iteracji PSO

|         | breast2         | 3lbreast2       | glass           | 3lglass         | thyroid        |
|---------|-----------------|-----------------|-----------------|-----------------|----------------|
| BP      | $58.1 \pm 18.1$ | $55.4 \pm 19.5$ | $42.9 \pm 15.6$ | $34.9 \pm 12.8$ | $96.4 \pm 5.2$ |
| logloss | $60.3 \pm 16.9$ | $55.6 \pm 19.6$ | $56.8 \pm 11.2$ | $49.3 \pm 13.1$ | $96.0 \pm 5.0$ |
| MSE     | $60.6 \pm 14.8$ | $56.2 \pm 19.1$ | $53.1 \pm 12.8$ | $34.2 \pm 12.4$ | $97.0 \pm 3.9$ |
| matche  | $59.1 \pm 18.3$ | $56.2 \pm 19.8$ | $50.7 \pm 13.9$ | $31.3 \pm 9.6$  | $96.8 \pm 4.0$ |

Tabela: Eksperyment 4.6h, BP – uczenie tylko propagacją wsteczną, pso\* – uczenie najpierw PSO, a potem BP



3lglass-pso4-bp2 49.29293 13.10987



- Korzyści ze zmiennej w czasie kary za wielkość wag (małe kary!)
- Konieczność doboru parametrów do problemu (niestety)
- Problemy z zatrzymywaniem na zbiorze walidacyjnym



- Wizualizacja wyników
- Łączone funkcje błędu
- Regularyzacja suma vs średnia, dobór współczynnika
- Regularyzacja na poziomie BP
- Powrót do problemu lokalizacyjnego
- Zaszumienie prezentacji wzorców
- Inne zbiory benchmarkowe

