

# Systemy rekomendacyjne

metody głębokie, sekwencyjne, reguły asocjacyjne

Marcin Sobański

Dawid Karczewski

Sebastian Korniewicz

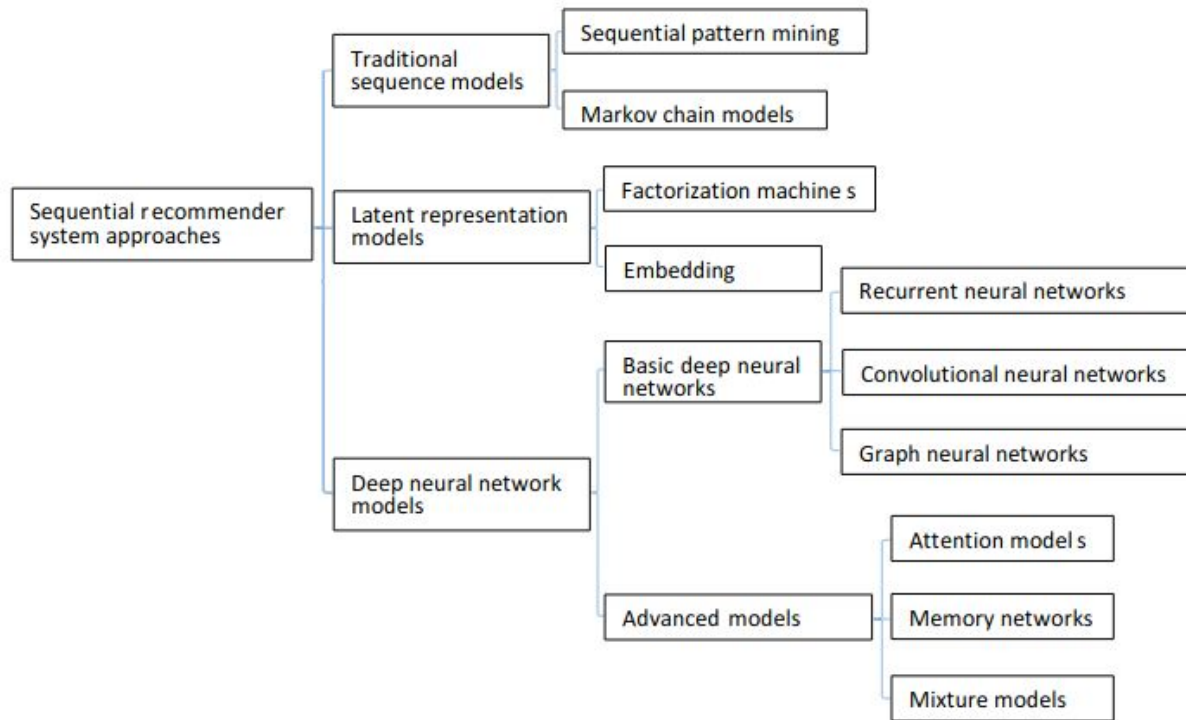
# Zbiór danych

- MovieLens 100k
- 100k ocen
- 1k użytkowników
- 1.7k filmów
- Minimum 20 ocen na użytkownika
- Dane sekwencyjne
- Ograniczony dla metod asocjacyjnych
  - 260 użytkowników
  - Filmy ponad 55 ocen

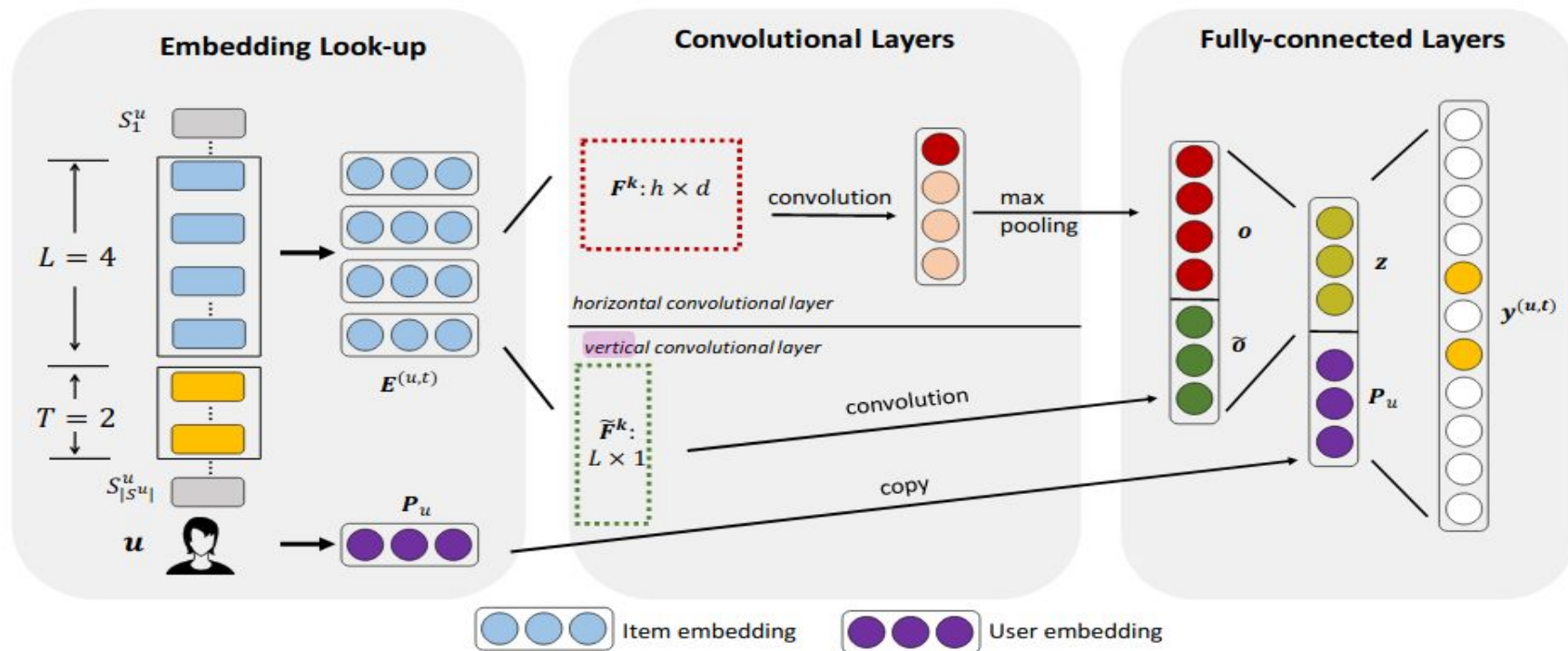
$$\text{Hit@}\ell = \frac{1}{m} \sum_{u \in \mathcal{U}} \mathbf{1}(\text{rank}_{u,g_u} \leq \ell),$$

# Sequential Recommender Systems (SRS)

- Elastyczna kolejność - np. kolejność nie ma znaczenia (mąka, masło)
- Szumy: (Masło, mleko, mąka, róża)
- Długość sekwencji.



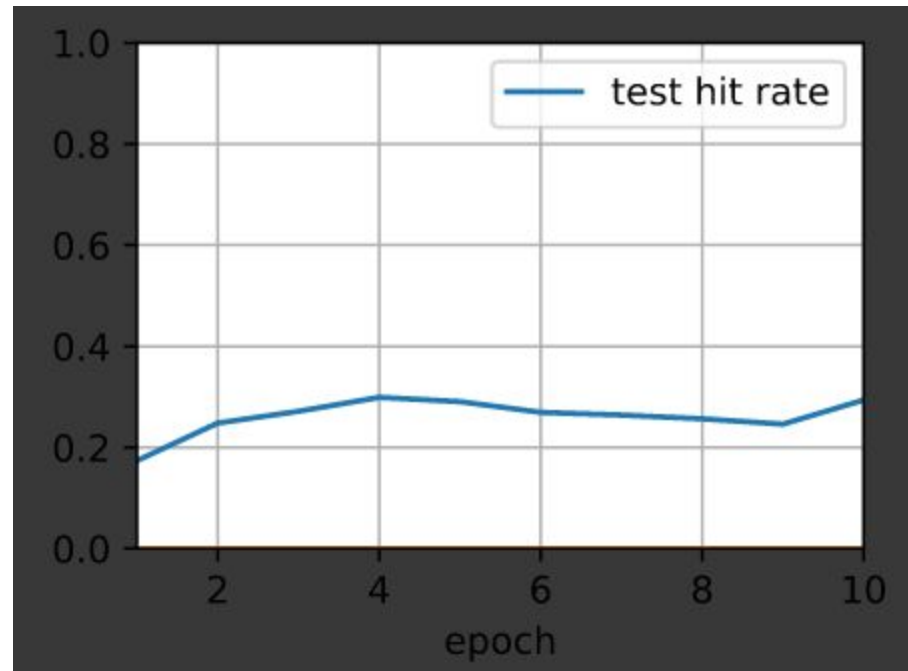
# Convolutional Sequence Embedding Recommendation Mode (Caser)



# Caser - wyniki

- Hit rate - 0.294, dla  $l=50$
- Długość sekwencji 5
- Testowanie 1 w przód

$$\text{Hit}@l = \frac{1}{m} \sum_{u \in \mathcal{U}} \mathbf{1}(\text{rank}_{u, g_u} \leq l),$$



# Modele głębokie

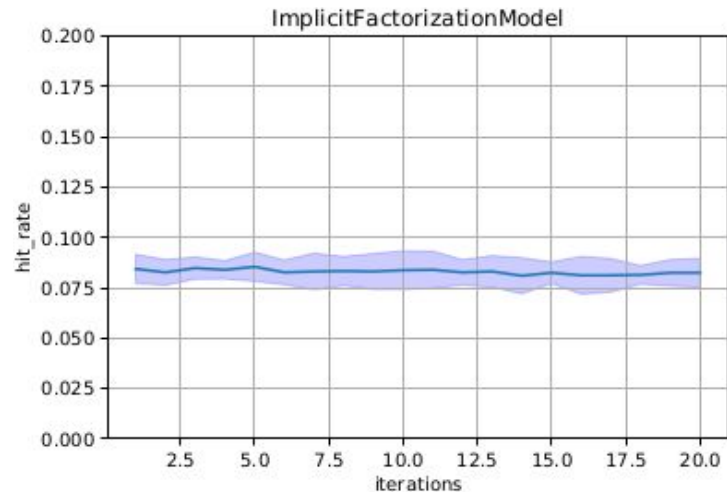
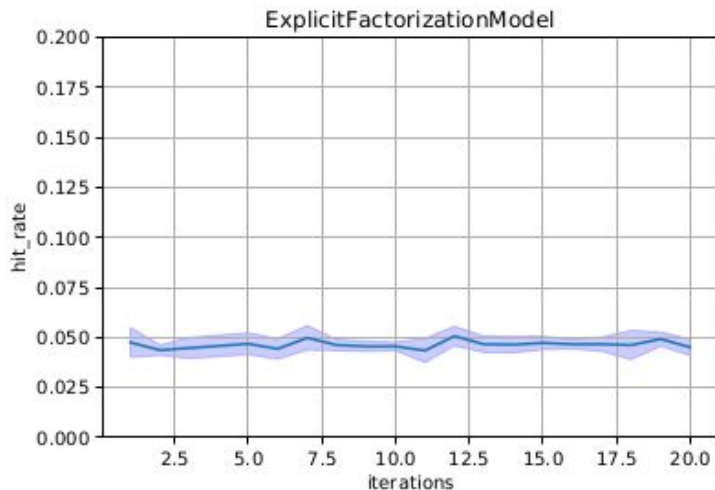


Narzędzie do szybkiego budowania i testowania systemów rekomendacji:

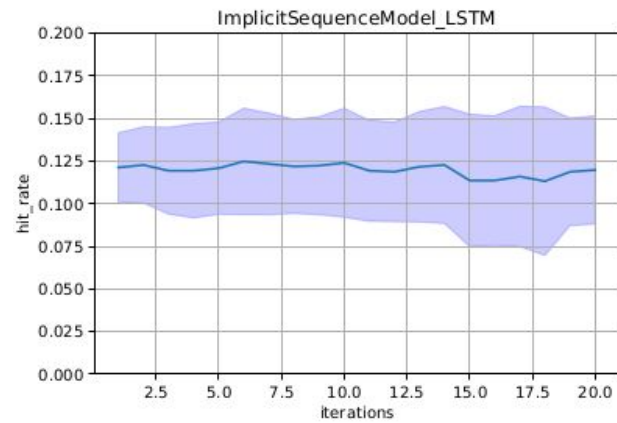
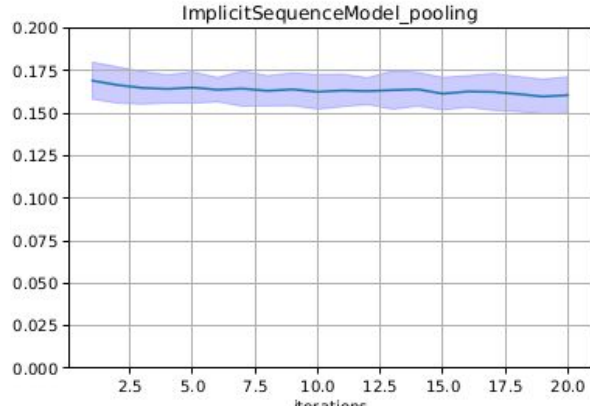
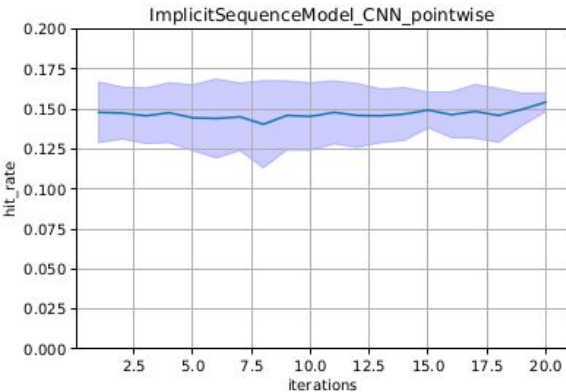
- Tworzenie płytkich (shallow) modeli reprezentacji poprzez tworzenie latent vector dla użytkowników
  - Gotowa implementacja BilinearNet
- Tworzenie głębokich modeli sekwencyjnych
  - Gotowa implementacje: CNNNet, LSTMNet, MixtureLSTMNet, PoolNet
- Pozwala na zaimplementowanie własnych sieci, gotowy zbiór popularnych funkcji strat
- Podstawowe narzędzia do testowania
- [github.com/maciejkula/spotlight/tree/master/examples](https://github.com/maciejkula/spotlight/tree/master/examples) !!!
  - Użycie Hash zamiast one hot
  - Negative sampling jako losowy wybór dla implicit

# Wyniki

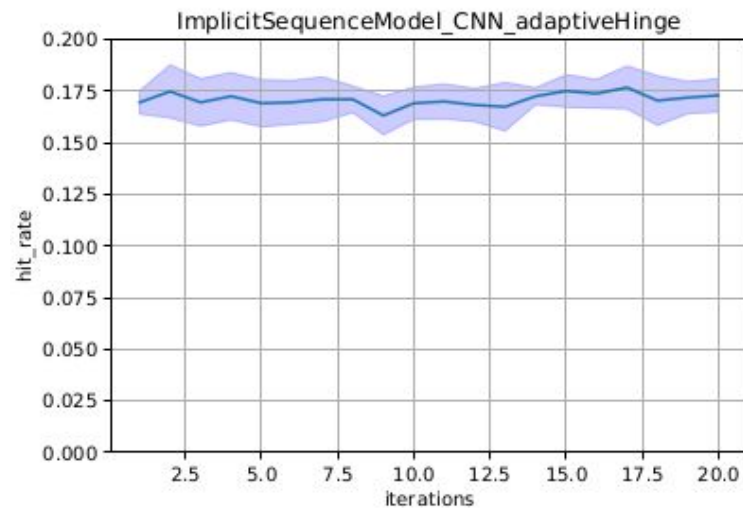
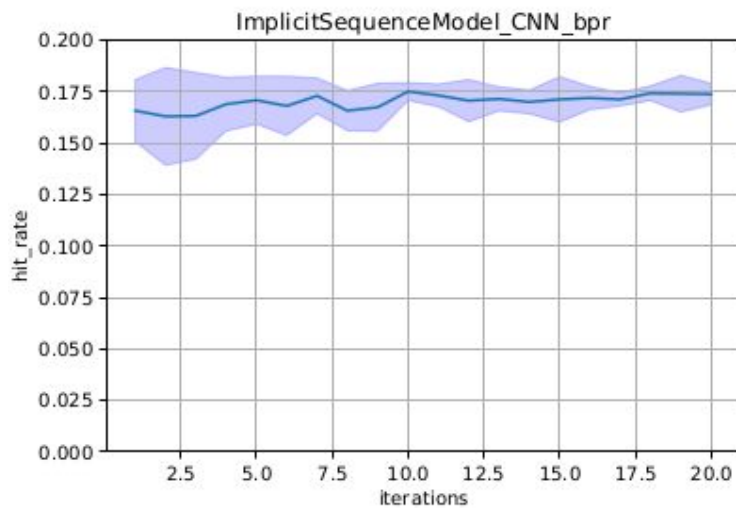
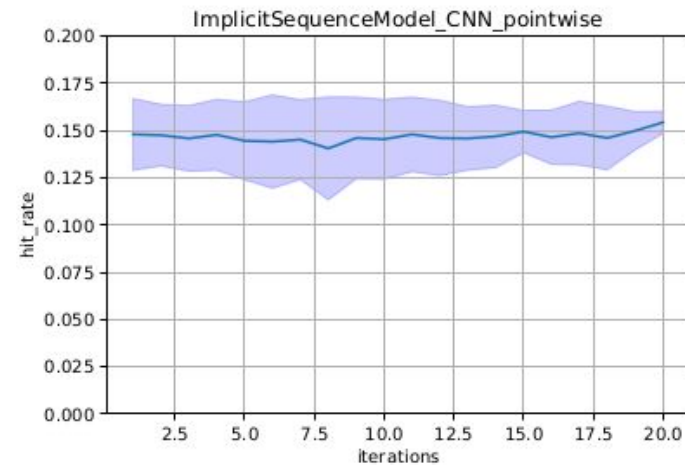
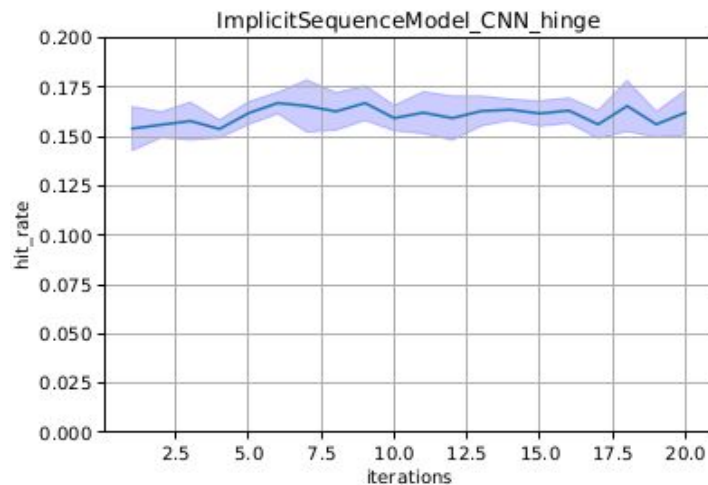
5 powtórzeń



$$\text{Hit@}\ell = \frac{1}{m} \sum_{u \in \mathcal{U}} \mathbf{1}(\text{rank}_{u, g_u} \leq \ell),$$



# Wyniki cd





# Reguły asocjacyjne:

Ograniczenia dla badań:

260 użytkowników uczenie

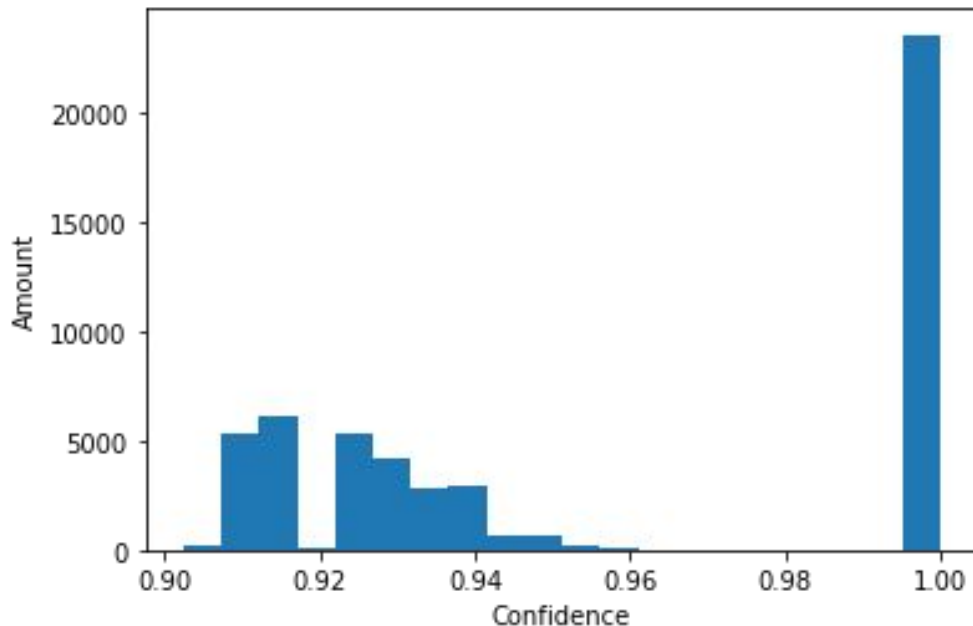
683 użytkowników testowanie

Minimum:

- 55 ocen dla filmu
- 55 wystąpień podzbioru

90% próg ufności

Hit rate top50 7.4% dla 40 ty.s wygenerowanych regół



# Wyniki podsumowanie

Metoda	Explicit Factorization	Implicit Factorization	Implicit Sequential POOLING	Implicit Sequential CNN	Implicit Sequential LSTM	Caser	Association rules
HitRate@50	0.050	0.085	0.169	0.125	0.177	0.294	0.074