

# Redukcja niepewności i obciążenia danych w systemach rekomendacyjnych

Natalia Pludra

# Problem

Systemy rekomendacyjne – personalizacja treści na podstawie interakcji użytkownika z obiektami

- Dane obarczone są niepewnością
- Dostępnych jest wiele bardzo podobnych produktów
- Użytkownik widzi ograniczoną liczbę produktów

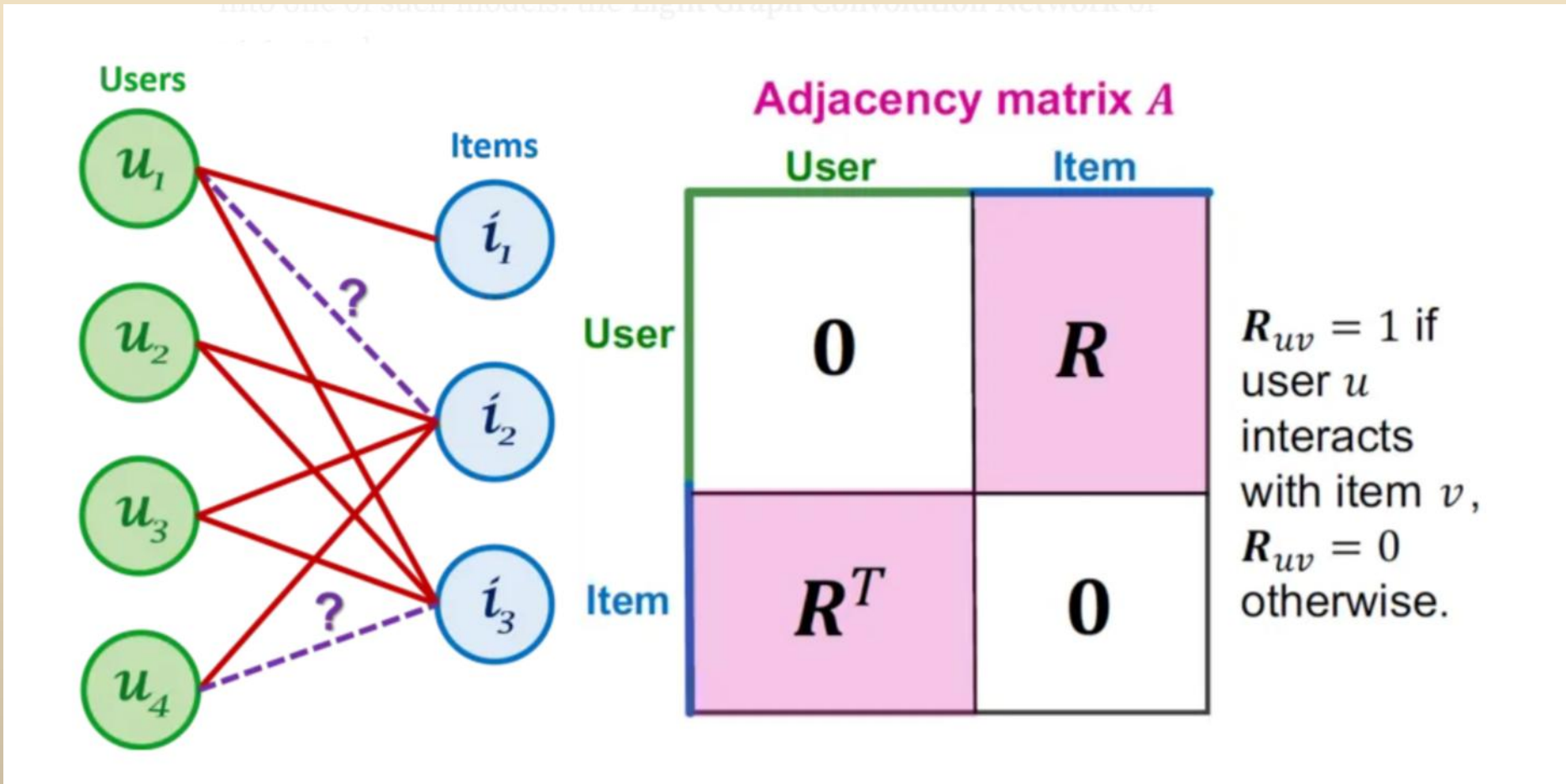
# Cel

- Poprawa odporności modeli na obciążenie i niepewność danych
- Uogólnienie modeli, żeby nie skupiały się na dokładnych interakcjach
- Podejście probabilistyczne

# Grafowe sieci neuronowe

- Dane w postaci grafów
- Wykorzystywanie cech poszczególnych elementów i powiązań między nimi
- Propagacja informacji między sąsiadującymi wierzchołkami
- Modelowania nieregularnych danych strukturalnych, takich jak sieci społeczne, systemy rekomendacyjne

# LightGCN (Light Graph Convolutional Network)



# LightGCN (Light Graph Convolutional Network)

- Model oparty na embeddingach
- Embedding (wektor osadzony) - reprezentacja danych w postaci N-wymiarowego wektora
- Cel: znalezienie optymalnych embeddingów użytkowników i produktów, które odzwierciedlają relacje i interakcje między nimi
- Zwraca K rekomendacji dla każdego użytkownika

# Jak działa LightGCN

1. Każdy użytkownik i przedmiot otrzymuje losowy embedding.
2. W każdej warstwie model aktualizuje embeddingi użytkownika/przedmiotu, uwzględniając embeddingi jego sąsiadów w grafie.

$$e_i^{(k+1)} = \sum_{u \in \mathcal{N}_i} \frac{1}{\sqrt{|\mathcal{N}_u|} \cdot \sqrt{|\mathcal{N}_i|}} e_u^{(k)}, \quad e_u^{(k+1)} = \sum_{i \in \mathcal{N}_u} \frac{1}{\sqrt{|\mathcal{N}_u|} \cdot \sqrt{|\mathcal{N}_i|}} e_i^{(k)}$$

3. Ostateczny embedding użytkownika/produktu jest kombinacją liniową embeddingów ze wszystkich warstw.
4. System używa iloczynu skalarnego, aby przewidzieć, jak bardzo dany użytkownik polubi dany przedmiot.

# Eksperymenty

- Dane: produkty z Amazona
- Ewaluacja:

- $Precision@K = \frac{\text{Liczba trafionych elementów w zbiorze K rekomendacji}}{K}$

- $Recall@K = \frac{\text{Liczba trafionych elementów w zbiorze K rekomendacji}}{\text{Liczba wszystkich trafnych elementów dla użytkownika}}$

- $NDCG@K$  - uwzględnia trafność rekomendacji oraz pozycję, na której pojawiła się trafiona rekomendacja



# Eksperymenty

- Modyfikacja zbioru treningowego:
  - Podmiana losowych produktów na produkty podobne (na podstawie odległości między embeddingami)
  - Testowanie na zaburzonym zbiorze
- Modyfikacja embeddingów w trakcie uczenia modelu:
  - Dodanie szumu z różnych rozkładów prawdopodobieństwa