

Doporučovací modely založené na rekurentních neuronových sítích



FAKULTA
INFORMAČNÍCH
TECHNOLÓGIÍ
ČVUT V PRAZE

Autor: Ing. Ladislav Martínek
Vedoucí práce: Ing. Tomáš Rehořek, Ph.D.

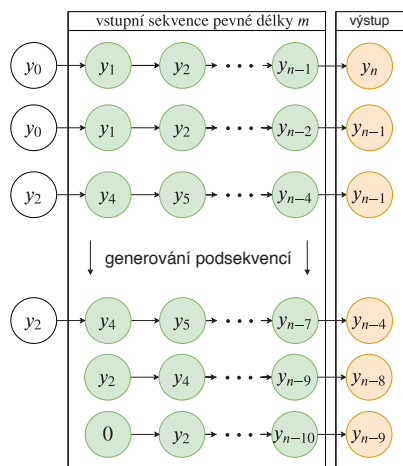
Katedra aplikované matematiky, Fakulta informačních technologií, České vysoké učení technické v Praze

Motivace

- Problematika doporučovací systémů a top-doporučování
- Aplikace doporučvacích algoritmů v offline světě
- Využití přirozené sekvenční povahy interakcí
- Vývoj uživatelských preferencí v čase
- Opakování oblíbených položek

Předzpracování sekvenčních dat

- Klíčová část pro natrénování úspěšného modelu
- Prováděno pomocí slučování interakcí v časovém okně, výběr cílového typu interakce (nákup) nebo generování podsekvencí



Datové sady

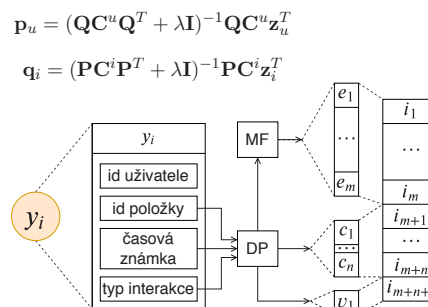
- 3 datové sady, kde 2 anonymizované sady (D1, D2) byly společně se servery pro trénování poskytnuty firmou Recombee s.r.o
- Databáze yoochoose, D1 (obchod s nábytkem), D2 (primární dataset představující obchod rozvážející jídlo)
- Parametry D2
 - # interakcí: **344 809 934**
 - # uživatelů: 3 625 903
 - # položek: 2 295
 - hustota interakční matice 2.182 %

Cíle práce

- Analýza interakčních dat a jejich zpracování
- Využití maticové faktorizace pro tvorbu vhodné reprezentace položek (embeddings)
- Návrh a implementace algoritmu pro tvorbu modelů založených na RNN
- Testování různých modelů a jejich hyperparametrizací a diskuze výsledků

Reprezentace položek - tvorba embeddingů

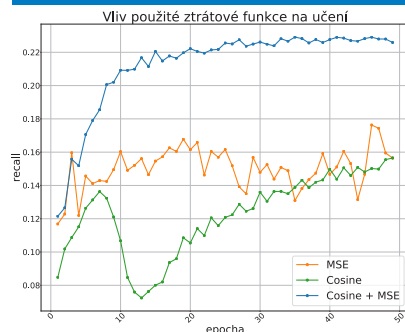
- Generování embeddingů pomocí maticové faktorizace (MF) upravené pro použití na datech tvořených pomocí implicitní zpětné vazby
- Využití kontextových dat v podobě časové známky a váhy interakce
- Navrženy 3 způsoby zpracování časové známky
 - Diskretizace v rámci sekvence nebo celého datasetu a nebo normalizovaná absolutní hodnota



$$p_u = (QC^uQ^T + \lambda I)^{-1}QC^u z_u^T$$

$$q_i = (PC^iP^T + \lambda I)^{-1}PC^i z_i^T$$

Experimenty

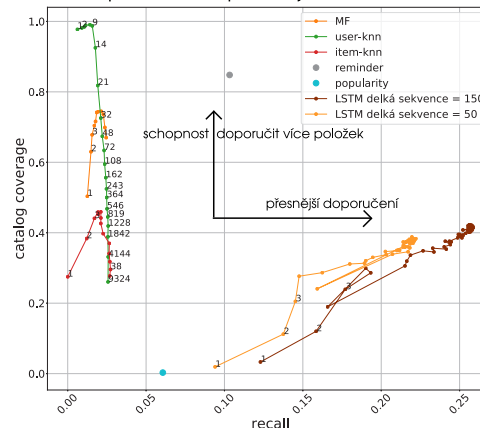


- Provedeno velké množství experimentů, především na datové sadě D2
- Použita metrika recall
- Kombinace s metrikou catalog coverage
- Vlevo je patrný vývoj metricky recall v průběhu trénování, kde bylo pro kvalitu modelu klíčové dosažení průběhu odpovídajícího modré křivce
- Navrženy byly také 3 možné způsoby vybavování, které umožňují přecházet mezi doporučováním podobných (spíše alternativ k položce, která by uživatele měla zaujmout) nebo sekvenčně a tím dosáhnout větší diverzity v doporučení (snažit se odhadnout, jak budou položky v sekvenci kupovány)
- Vyhodnocování probíhalo pomocí leave-last-one-out validace

- Provedeno velké množství experimentů, především na datové sadě D2
- Použita metrika recall

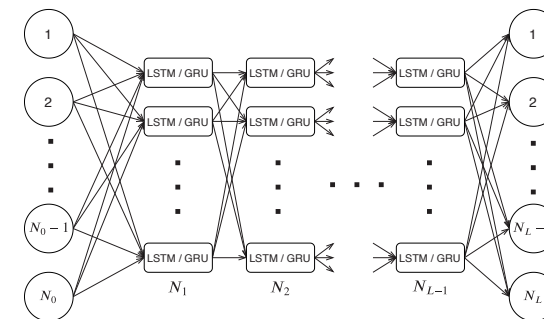
$$Recall@N = \frac{\#true-positive}{\#true-positive + \#false-negative}$$

- Kombinace s metrikou catalog coverage
- Vlevo je patrný vývoj metricky recall v průběhu trénování, kde bylo pro kvalitu modelu klíčové dosažení průběhu odpovídajícího modré křivce



Rekurentní model

- Shodná velikost vstupní a výstupní vrstvy
- Sryté vrstvy jsou rekurentní
- Vstupní vrstva doplňována o kontextová data interakcí, především o časovou známku a její váhu



- Navržena vlastní ztrátová funkce, úspěšně dosaženo silné korelace s recallem (požadovanou metrikou)
- Kombinace kosinové a střední kvadratické chyby (MSE) pro MF embedding
- Pro kontextová data využita pouze MSE
- MSE pro MF embedding umožňuje dát důraz na omezení velikosti absolutních hodnot

$$\mathcal{L} = -\omega_1 \cdot \frac{\sum_{i=1}^m y_i \cdot \hat{y}_i}{\sqrt{\sum_{i=1}^m y_i^2} \cdot \sqrt{\sum_{i=1}^m \hat{y}_i^2}} + \omega_2 \cdot \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 + \omega_3 \cdot \frac{1}{n} \sum_{i=m+1}^{m+n} (y_i - \hat{y}_i)^2 + \omega_4 \cdot (y_{m+n+1} - \hat{y}_{m+n+1})^2$$

Výsledky a závěr

- Zjištěny závislosti mezi velikostí embeddingů a velikostí modelu a jejich vliv na úspěšnost
- Zjištěn vliv délky sekvence na úspěšnost
- **Podářilo se dosáhnout více jak dvojnásobného recallu oproti zvoleným standardním metodám**
- Předzpracování má velký význam na kvalitu výsledného modelu
- Nutnost optimalizace hyperparametrů MF pro dosažení kvalitních embeddingů
- Model je již delší dobu nasazen při real-time doporučováním a také real-time vybavování
- Diplomová práce byla zpracována pod záštitou firmy Recombee s.r.o