

## Posudek oponenta závěrečné práce

Student: Bc. Martin Spišák

Oponent práce: Ing. Vojtěch Vančura

Název práce: Sparse Approximate Inverse for Enhanced Scalability in Recommender Systems

---

Předložená diplomová práce rozvíjí oblast tzv. shallow (mělkých, míněno jako opak hlubokých – deep) modelů pro doporučovací systémy, která v poslední době získala velkou pozornost výzkumné komunity okolo doporučovacích systémů, a to jak v akademické, tak v průmyslové sféře. Práce je logicky velmi dobře rozdělena do tří bloků o celkem šesti kapitolách.

V první kapitole autor čtenáře provede základy doporučovacích systémů, vysvětlí základní pojmy a představí běž používané algoritmy pro doporučování. Oceňuji zmínku o tzv. multi-stage přístupu, který je aplikován v typických praktických implementacích doporučovacích systémů. Následuje kapitola druhá, která uvede čtenáře do problematiky řídkých (sparse) matic. Tato kapitola je obzvláště cenná, pokud je čtenář pouhým uživatelem frameworků podporujících algebru řídkých matic (jako je například `scipy.sparse`). Dále následuje představení velmi známého modelu EASE, na který práce navazuje. Autor se v této části nebojí jít do hloubky a vysvětluje původní článek velmi edukativně. Na závěr úvodního bloku přichází čtvrtá kapitola, ve které autor nastiňuje principy inverze řídké matice. Tato kapitola je nezbytná pro pochopení představeného modelu a opět je napsána velmi srozumitelným a čtivým způsobem.

Těžištěm práce je pátá kapitola, ve které autor představuje vlastní model SANSA. Popis modelu je opět velmi jasný a srozumitelný. Je zde navíc jasně vidět podtext autorovy spolupráce s průmyslem; jak teoretický design, tak i praktická implementace nese snahu o efektivitu a maximální využitelnost v reálném prostředí.

Poslední kapitola se zabývá praktickými experimenty. Autor pro evaluaci používá pět datasetů (MovieLens, Netflix, MSD, Goodbooks, Amazon Books) a vlastní model (SANSA) porovnává s ostatními shallow modely (EASE, MRF), deep modely (RECVAE, MULT-VAE) a pro kontext přidává i maticovou faktorizaci (WMF). Zde bych rád vyzdvihl autorovu vlastní implementaci modelu MRF, který je často při porovnávání modelů neprávem přehlížen. Na druhou stranu bych doporučoval (minimálně do budoucna) doplnit do porovnání jinak zmiňovaný model ELSA, v průmyslu běžně používanou maticovou faktorizaci ALS a stále velmi silné KNN, což ostatně autor sám zmiňuje zcela na závěr při popisu dalších plánovaných prací na svém modelu.

Celkově bych rád vyzdvihnul nejen srozumitelnost a čtivost práce, kterou autor zvládl udržet i přes ne zcela jednoduché téma, zejména však celkový výsledek, který se jistě dočká bohatého průmyslového využití. To ostatně potvrzuje i přijetí článku vycházející z práce na prestižní konferenci RecSys2023.

Vzhledem k výše uvedenému nemohu jinak, než navrhnout hodnocení předložené práce **výborně**.

Protože předpokládám, že autor bude představený model SANSA dále rozvíjet, rád bych autorovi položil následující otázky, které je třeba zároveň brát jako naznačení směru, kterým by se tento rozvoj mohl dále ubírat:

1. **Časová složitost.** Autor v práci explicitně nevyjadřuje časovou složitost algoritmu. Toto by stálo za to buď vyjádřit formálně nebo alespoň prověřit na syntetických datech. V praxi lze totiž vypočítat trade off mezi dobou tréninku a výkonem modelu – čím déle model trénujeme, tím máme přesnější model; zároveň však platí, že čím déle model trénujeme, tím starší data jsou pro trénink použita, což výkon modelu snižuje. Proto je potřebné mít o této problematice lepší přehled.
2. **Business rules.** Algoritmy pro kolaborativní filtrování se v praxi používají v multi-stage doporučení, kde zastávají roli prvního modelu, který vybírá kandidáty pro doporučení. Za nimi pak většinou navazuje aplikace nějakých pravidel – např. lze doporučovat jenom položky, které máme skladě, pouze články, které jsou staré maximálně několik hodin a podobně. Co když se stane, že zrovna v těch místech té řádkové váhové matice, které potřebujeme pro zrovna doporučitelné kandidáty budou nuly? Nemůže to negativně ovlivnit výkon doporučovacího systému? Šlo by tomu nějak předejít?
3. **(A)symetrie.** I název modelu naznačuje, že výsledná matice vah je asymetrická. To má svůj význam v situaci, kdy má model při tréninku na vstupu a výstupu různé hodnoty. Typický příklad by byl obchod s dětským vybavením: lidé, kteří mají v nákupní historii plenky velikosti 2, si následně kupují plenky velikosti 3, ale ne plenky velikosti 1 (protože dítě roste). Ovšem SANSA je trénována pouze na interakční matici, kde je vše zagregováno dohromady a tato závislost se ztratí. Šlo by nějak využít asymetrie ve váhové matici k podchycení tohoto případu?
4. **Partial fit.** Uměl by model se inkrementálně doučovat? U modelu ELSA máme dobré výsledky s inkrementálním učením – na začátku model natrénujeme na všech datech a pak postupně nové položky umísťujeme do již existujícího latentního prostoru, což je velmi rychlé. Šlo by podobný přístup použít i u modelu SANSA? Pro model EASE byl podobný přístup zkoumán v [1]
5. **Cold start.** U kolaborativního filtrování je často problematické doporučení nových položek, které ještě nenasbíraly dostatek interakcí. Uměl by model generalizovat nové položky pomocí jejich atributů (obrázky, popisy produktů a podobně)? U metod generujících latentní prostor lze natrénovat funkci mapující atributy do latentního prostoru, lze něco obdobného využít i zde? Pro model EASE byl podobný přístup zkoumán v [2]

[1] Jeunen O, Van Balen J, Goethals B. Embarrassingly shallow auto-encoders for dynamic collaborative filtering. *User Modeling and User-Adapted Interaction*. 2022 Sep;32(4):509-41.

[2] Jeunen O, Van Balen J, Goethals B. Closed-form models for collaborative filtering with side-information. In *Proceedings of the 14th ACM Conference on Recommender Systems 2020 Sep 22* (pp. 651-656).

V Praze dne:

Podpis: