

## СИСТЕМ ЗА ПРЕПОРУКУ ЗАСНОВАН НА ГРАФОВСКИМ НЕУРОНСКИМ МРЕЖАМА

## RECOMMENDER SYSTEM BASED ON GRAPH NEURAL NETWORKS

Марко Његомир, Јелена Сливка, Факултет техничких наука, Нови Сад

## Област – РАЧУНАРСТВО И АУТОМАТИКА

**Кратак садржај** – Главна идеја рада била је да се направе модели машинског учења утемељени на графовским неуронским мрежама који би вршили препоруке производа корисницима. Од доступних података о куповним навикама корисника направљен је хетерогени бипартитан граф који садржи чворове корисника и производа, а везе у њему представљају интеракције које су корисници имали са производима. Модели су тестирани на *Brazilian E-commerce* скупу података. Први модел је користио један енкодер са слојевима графовске неуронске мреже за енковање чворова, и један декодер ивица, чија је улога била да прави предикцију оцене коју би корисник дао производу. Други модел је користио посебне енкодере за кориснике и за производе, а такође је имао један декодер ивица. Модел са два енкодера се показао као бољи када се посматра *RMSE* метрика, а при том је имао и стабилније тренирање.

**Кључне ријечи:** графовске неуронске мреже, системи за препоруку, *GraphSAGE*.

**Abstract** – This study aimed to create a machine-learning model based on graph neural networks that can provide item recommendations for users of an e-commerce system. For model training, the data had to be converted to a heterogenous bipartite graph, where users and products were represented as nodes, while edges represented the interactions between them. Models were tested on the *Brazilian E-commerce* dataset. The first model used one encoder containing graph neural network layers for encoding nodes and one decoder for edges, which aimed to make rating predictions that the user would assign to the product. The second model used separate encoders for users and products and had an edge decoder. The model with two encoders achieved better *RMSE* than the single encoder model and had a more stable training process.

**Keywords:** graph neural networks, recommender systems, *GraphSAGE*.

## 1. УВОД

У системима електронске трговине, системи за препоруку су веома битни. У понуди често постоји веома велики број производа, и битно је препоручити

## НАПОМЕНА:

Овај рад проистекао је из мастер рада чији ментор је била др Јелена Сливка, ванр. проф.

подкуп тих производа који ће повећати задовољство корисника, а тиме и вероватноћу да ће доћи до куповине [1].

За препоруке су се до сада успешно примењивале једноставније технике као што су *Content-based filtering* [2] или мало напреднији приступ са факторизацијом матрица [3]. Предности оваквих система је што су брзи за извршавање, а мана је што не могу да користе додатне информације о интеракцији свих корисника у систему са производима.

Ове мане је делимично било могуће надоместити коришћењем ансамбла једноставнијих модела [2]. У скорије време је постало јасно да се велик број информација о куповним навикама корисника може представити у структури графа, па је на њих онда могуће применити и графовске неуронске мреже [4]. Ове архитектуре су у стању да искористе обиље информација о корисницима, производима и њиховим интеракцијама, и могу се посматрати као модели који комбинују предности *Content-based filtering* приступа са колаборативним филтрирањем, код кога се узимају и информације о интеракцијама других корисника у обзир, а при том су у стању да избегну проблеме које су претходно поменути приступи имали [2][3][4].

У овом раду су коришћени *GraphSAGE* [5] слојеви неуронске мреже да би се изградиле једно-енкодерски и дво-енкодерски модели. Ови модели су тренирани и евалуирани на *OList Brazilian E-commerce* скупу података [6]. На основу овог скупа података је направљен хетерогени бипартитан граф корисника, производа и њихових интеракција. Модели из овог рада се тренирају да предвиде коју оцену би корисник дао производима, и затим се ти производи могу сортирати да би се добила коначна листа препорука.

У овом раду ће прво бити приказана претходна решења за овај проблем. Након тога ће бити описани детаљи о имплементацији решења. Затим ће се анализирати добијени резултати. У последњем поглављу ће бити представљен закључак рада.

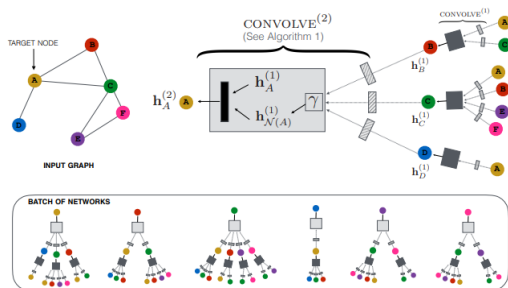
## 2. ПРЕТХОДНА РЕШЕЊА

Један од основних приступа у системима за препоруку је *Content-based filtering* [2], у коме се корисницима препоручују производи који су слични онима са којима су раније имали интеракцију. Овај приступ је брз због своје једноставности, али не узима у обзир информације о потрошачким навикама свих корисника у систему. Још једна мана овог приступа је што корисницима препоручује ствари сличне онима

које су већ купили, што не мора значити да желе да купе одмах поново сличан производ.

У раду са факторизацијом матрица [3], аутори су представили својства производа и својства корисника помоћу матрица, а коначно рангирање производа за сваког корисника се добија множењем матрица. Овакав приступ може имати проблема када корисници нису имали довољно интеракција са производима. У раду [3] су истраживани различити алгоритми у циљу да се овај проблем реши, као и да се побољша способност скалирања оваквих алгоритама.

Аутори рада [7] су покушали да реше недостатке претходно поменутих приступа коришћењем конволутивних слојева графовских неуронских мрежа (енг. *Graph Convolutional Neural Networks - GCN*). У раду је направљен систем за препоруку за веб-платформе које раде са великим количинама података. Експерименти су вршени на подацима са *Pinterest* сајта. Сви подаци у претворени у форму графа, и затим је направљен *PinSAGE* модел који је приказан на слици 1, и који има два GCN слоја које је ефикасно могао да обради велику количину података, тако што је био у стању да експлоатише локалну структуру графова и употребљава паралелизацију. Овакав модел је у стању да користи и информације о свим корисницима и производима у графу, и да умањи проблем када неки корисник нема много интеракције са производима, због тога што граф садржи обиље контекстних информација о сличним корисницима и производима. Због тога је овај модел постигао воље резултате од претходно поменутих стандардних метода за препоруку.



Слика 1. Двослојни *PinSAGE* модел [7].

У раду [8] је представљен модел за колаборативно филтрирање помоћу графовских неуронских мрежа (*Neural Graph Collaborative Filtering - NGCF*). У овом раду су интеракције између корисника и производа моделоване помоћу бипартитног графа. Модел користи GCN слојеве, и у стању је да прави персонализоване препоруке а да при том искористи и информације о куповним навикама других корисника. Овај модел је показао значајно побољшање у односу на класично колаборативно филтрирање и у односу на друге модерне приступе са којима је поређен у раду.

Аутори рада [9] су демонстрирали значај увођења мета-путања у хетерогеним графовима код система за препоруку. Мета-путања је веза између два чвора која нису директно повезана у оригиналном графу, али је таква веза додата јер их неки други тип чвора повезује. У раду је коришћена хетерогена графуска

неуронска мрежа са пажњом (*Heterogenous Graph Attention Network – HGAT*). Додавање нових веза у виду мета-путања и коришћења механизма пажње који додељује значај свим везама је довело до побољшања у препорукама у односу на друге моделе са којим су вршена поређења.

Овај рад показује значај истовременог повећања комплексности структуре података и флексибилности модела за постизање бољих резултата.

### 3. МЕТОД

У наредним поглављима су описан је скуп података кориштен у раду, начин на који је извршено претпроцесирање и архитектура испробаних модела за препоруке.

#### 3.1. Скуп података

Скуп података који је коришћен у раду је *OList Brazilian E-commerce* [6], и он се састоји од више табела са информацијама о корисницима, производима, поруџбинама, и рецензијама са оценама.

У скупу података постоји око 150 000 поруџбина, и од тога 650 поруџбина нема оцену. Подаци у скупу података су анонимизовани, тако да није могуће препознати кориснике, али ни производе.

За кориснике су доступне географске информације. Поруџбине садрже информације о цени и датуму поручивања, а такође постоје и информације о оценама поруџбина.

Производи садрже опште информације о производу, као што су димензије, тежина, цена и категорија производа.

#### 3.2. Претпроцесирање података

За рад са графовским неуронским мрежама, од информација из скупа података је направљен хетерогени бипартитни граф. У том графу чворови представљају кориснике и производе, док ивице представљају интеракцију производа са корисницима.

Географске информације о корисницима нису значајне у свом изворном облику, па су подаци о корисницима додатно обогаћени прикупљањем са интернета. Тако су додате информације о величини и густини градова корисника, као и бруто регионалном производу.

Ивице у графу садрже информације о оцени коју је корисник дао производу, као и количини производа који су купљени. За експеримент са другим типом модела су додате и мета-путање између производа у графу.

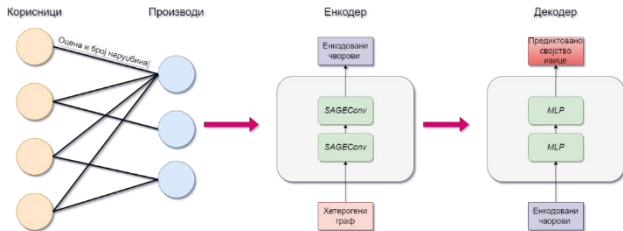
Све информације о поруџбинама које нису биле потпуне јер није постојала информација о оцени су одбачене, јер то је представљало само 0.43% укупног броја поруџбина.

Просечан број интеракција које су корисници имали са производима је само 1.21, што је последица реалне ситуације са много пасивних корисника, и мањим бројем веома активних купаца. Оваква расподела је добар тест за могућност графовских неуронских мрежа да се изборе са тим проблемом.

### 3.3. Архитектура решења

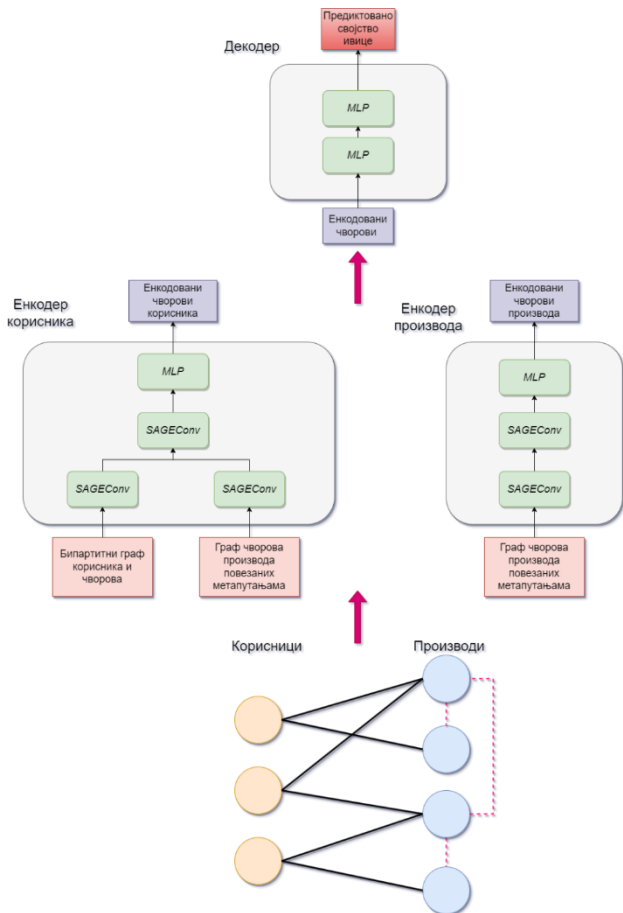
За прављење препорука су коришћена два модела графовских неуронских мрежа, са једним и са два енкодера. Обе мреже имају декодер ивица који има задатак да изврши предикцију оцене коју би корисник дао неком производу.

Једно-енкодерски модел се састоји од два SAGEConv слоја. Енкодер прима информације о производима и корисницима у виду хетерогеног бипартитног графа. Резултати енкодера се прослеђују декодеру ивица као што је приказано на слици 2.



Слика 2. Једно-енкодерски модел.

Дво-енкодерски модел има посебне енкодере за производе и ивице. У оба енкодера се прослеђује хетерогени граф обогаћен додатним информацијама у виду мета-путања. Посебно енкодирани производи и корисници се прослеђују декодеру ивица као што је приказано на слици 3.



Слика 3. Дво-енкодерски модел.

Оба модела се тренирају и евалуирају трансдуктивно, што значи да су сви чворови доступни приликом тренирања и евалуације, док су ивице подељене на три скупа, тренинг скуп који чини 80% ивица, и

валидациони и тест скуп од који сваки има 10% ивица.

Због тога што је скуп података небалансиран, јер је најзаступљенија оцена 3 на скали од 1 до 5, коришћена је отежињена средња квадратна грешка (*weighted mean squared error* - WMSE) као функција губитка приликом тренирања.

Системи за препоруку се често евалуирају помоћу  $\text{top}@k$  метрике, која мери однос препоручених производа и првих  $k$  производа сортираних по томе како их је корисник оценио. У овом скупу података, просечан број оцена је 1.21, па је због тога коришћена средња квадратна грешка (*Mean Squared Error* - MSE) као метрика за евалуацију.

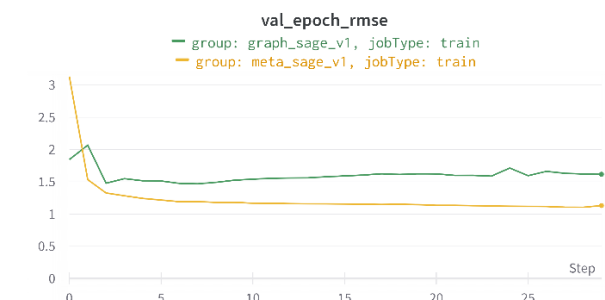
### 4. РЕЗУЛТАТИ И ДИСКУСИЈА

Оба модела су тренирана 300 епоха. Након сваке епохе се вршила и валидација, и чуван је модел који је добио најмању RMSE вредност. Приликом тренирања оба модела коришћен је *ADAM* (*Adaptive Momentum*) оптимизатор са брзином учења 0.01. У табели 1 су дати добијени резултати за оба модела на тест скупу.

Модел	RMSE
Једно-енкодерски модел	1.43
Дво-енкодерски модел	1.118

Табела 1. Перформансе модела графовских неуронских мрежа.

Из резултата се јасно види да је дво-енкодерски модел постигао боље резултате од једно-енкодерског модела. Ово је последица тога што је дво-енкодерски модел имао богатији граф података на располагању због додавања мета-путања између производа, као и чињенице да је имао већу флексибилност што му је и омогућило да створи боље репрезентације корисника и производа. При том је дво-енкодерски модел такође имао и стабилније тренирање. Са слике 4 се види да RMSE крива стабилно опада за дво-енкодерски модел током целог тренирања, док крива за једно-енкодерски модел показује одређену нестабилност због наглих скокова. Осим тога, ове две криве показују и да је дво-енкодерски модел мање склон преприлагођавању од једно-енкодерског модела, што се види у благом расту криве једно-енкодерског модела након што је достигао најбољи резултат већ после пар епоха.



Слика 4. RMSE крива током тренирања за једно и дво-енкодерски модел.

## 5. ЗАКЉУЧАК

У овом раду су представљени модели за препоруку утемељени на графовским неуронским мрежама. Идеја је била да се за проблем препоруке искористе напредне архитектуре графовских неуронских мрежа које су у стању да искористе информације о корисницима, производима и о њиховим интеракцијама. Овај приступ се због тога може посматрати као комбинација колаборативног филтрирања, које узима и куповне навике других корисника у обзир, и *content based* модела, који узимају у обзир сличност производа са производима са којима је корисник раније имао интеракцију.

За потребе ових модела, од табела из *OList Brazilian E-commerce* [6] скупа података табела је направљен хетерогени бипартитни граф корисника и производа. Везе у овом графу представљају интеракцију корисника са производима, и носе информацију о томе како је корисник оценио производ, и колику количину је претходно купио. Географске информације о корисницима су искоришћене да се прикупе додатни подаци са интернета за густину насељености места из кога корисник долази, као и информације о бруто-регионалном производу, због тога што ови додатни подаци могу носити информацију о куповним навикама корисника.

У раду су направљена два модела графовских неуронских мрежа. Једно-енкодерски модел је изграђен од GraphSAGE [5] слојева графовске неуронске мреже, и он ради са хетерогеним бипартитним графом корисника, производа и њихових интеракција. Декодеру ивица се затим пропушта излаз енкодера који врши предикцију коју оцену би корисник дао неком производу. Дво-енкодерски модел користи посебне енкодере изграђене од GraphSAGE слојева за кориснике и производе, и резултате из оба енкодера прослеђује декодеру ивица који је врши предикцију оцене као и претходни модел. У дво-енкодерски модел се шаље хетерогени граф који је додатно обогаћен метапутањима између производа са којима је неки корисник претходно имао интеракцију.

Перформансе оба модела се мере помоћу RMSE метрике, и дво-енкодерски резултат је постигао бољу вредност од 1.118 RMSE у односу на 1.48 RMSE коју је постигао једно-енкодерски модел. Ово је последица богатијег графа података који дво-енкодерски модел може да искористи због своје веће флексибилности.

Предмет даљих истраживања би могло бити убацивање додатних временских димензија за куповину, због тога што куповне навике корисника често одликује и периодичност. Ове информације се лако могу убацивати као додатна обележја у ивице графа, и графовске неуронске мреже су у стању да искористе све те додатне информације.

Још један правац истраживања би могло бити додавање додатних мета-путања између корисника, и додавање других типова мета-путања. Овакав приступ ди довео до веома повезаног графа, и због тога би се могли користити слојеви графовске неуронске мреже

са пажњом, који би били у стању да процене колико пажње треба да се обрати на који тип везе.

## 6. ЛИТЕРАТУРА

- [1] Wu, L.L., Joungh, Y.J. and Lee, J., 2013, January. Recommendation systems and consumer satisfaction online: moderating effects of consumer product awareness. In *2013 46th Hawaii International Conference on System Sciences* (pp. 2753-2762). IEEE.
- [2] Koren, Y., 2009. The bellkor solution to the netflix grand prize. *Netflix prize documentation*, 81(2009), pp.1-10.
- [3] Koren, Y., Bell, R. and Volinsky, C., 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8), pp.30-37.
- [4] Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G., 2008. The graph neural network model. *IEEE transactions on neural networks*, 20(1), pp.61-80.
- [5] Xu, K., Hu, W., Leskovec, J. and Jegelka, S., 2018. How powerful are graph neural networks?. *arXiv preprint arXiv:1810.00826*.
- [6] Olist and André Sionek, 2018, Brazilian E-Commerce Public Dataset by Olist Data set. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/195341>.
- [7] Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W.L. and Leskovec, J., 2018, July. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 974-983).
- [8] Wang, X., He, X., Wang, M., Feng, F. and Chua, T.S., 2019, July. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 165-174).
- [9] Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P. and Yu, P.S., 2019, May. Heterogeneous graph attention network. In *The world wide web conference* (pp. 2022-2032).

### Кратка биографија:



**Марко Његомир** рођен је 1991. године у Апатину, Република Србија. Основне академске студије завршио је 2022. године на Факултету техничких наука, на ком брани и мастер рад 2023. године из области Електротехнике и рачунарства – Софтверско инжењерство и информационе технологије.  
контакт:  
marko.njegomir@uns.ac.rs