



СОФТВЕРСКИ СИСТЕМ ЗА ПРЕПОРУКУ СТУДИЈСКОГ МОДУЛА НА СМЕРУ РАЧУНАРСТВО И АУТОМАТИКА

SOFTWARE SYSTEM FOR RECOMMENDING STUDY MODULE ON COURSE COMPUTER AND CONTROL ENGINEERING

Светозар Стојковић, Факултет техничких наука, Нови Сад

Област - ЕЛЕКТРОТЕХНИКА И РАЧУНАРСТВО

Кратак садржај - Овај рад бави се истраживањем перформанси два приступа системима за препоруку. Први тип система за препоруку је систем за препоруку заснован на колаборативном филтрирању, а други је заснован на моделу латентних фактора. За складиштење података користе се технологије семантичког веба где се подаци налазе на Фусеки серверу одакле се добављају SPARQL упитима. Тачност система за препоруку заснованом на колаборативном филтрирању износи 83.92% док је тачност система за препоруку засновану на моделу латентних фактора 30.04%. Закључено је да иако примитивнија метода, колаборативно филтрирање, даје боље резултате јер се користи део целокупног скупа студената.

Кључне речи: системи за препоруку, колаборативно филтрирање, модел латентних фактора, семантички веб, RDF, SPARQL, Java Spring

Abstract - This paper contains an performance analysis on two types of recommender systems. First type is a recommender system based on collaborative filtering and the other one is based on latent factor model. Semantic web technologies are used for storing data, data are stored on Fuseki server and they are accessed via SPARQL queries. Recommender system based on collaborative filtering have the accuracy of 83.92% while the accuracy of LFM recommender system is 30.04%. It was concluded that even though collaborative filtering is more primitive method because it uses the part of whole student grades dataset.

Keywords: recommender systems, collaborative filtering, latent factor model, semantic web, RDF, SPARQL, Java Spring

1. УВОД

У овом раду је описан систем за препоруку студијских модула на основу оцена студента. На почетку рада представљени су основни концепти Семантичког Веба: онтологије, RDF, SPARQL и приказани су разни типови система за препоруку. Након тога пружен је увид у организацију наставе у по студијским модулним, где је акценат стављен на модуле на смеру Рачунарство и аутоматика.

НАПОМЕНА:

Овај рад проистекао је из матер рада чији ментор је био др Милан Сегединац.

Потом следи приказ софтверског пројекта, кроз имплементацију две врсте система за препоруку и тестирање тачности појединачних система на реалним подацима са студентске службе Факултета техничких наука.

2. ТЕОРИЈСКЕ ОСНОВЕ

У овој секцији изложени су теоријски концепти потребни за разумевање пројекта, у ком је имплементиран систем за препоруку студијског модула.

2.1 Вокабулари и онтологије на семантичком вебу

У оквиру семантичког веба вокабулари дефинишу концепте и везе које се користе за описивање и представљање одређеног домена.

Вокабулари представљају основне градивне делове Семантичког веба [1].

2.2 Web ontology language (OWL)

W3C Web ontology language (OWL) је језик семантичког веба дизајниран да представи богато и комплексно знање о стварима, групама ствари и њиховим везама. OWL је формални језик заснован на дескриптивној логици. OWL документи - онтологије - могу бити објављени на World Wide Web (WWW) [2].

2.2.1 RDF

RDF стандарни је модел за размену података на вебу који проширује везну структуру веба да користи URI-је за формирање веза међу ставкама као и два краја везе (ово се назива "triple").

Везна структура креира усмерени означен граф, где гране представљају означену везу између два ресурса где се ресурси представљају као чворови графа [3].

2.2.2 SPARQL

SPARQL се може користити за вршење упита у различитим типовима података, без обзира да ли су подаци чувани у RDF или неком другом стандардном формату семантичког веба. SPARQL подржава агрегацију, подупите, негацију, креирање вредности изразима, тестирање вредности и ограничавање упита у RDF графу. Резултати SPARQL упита могу бити сетови или RDF графови [4].

2.3 Системи за препоруку

Системи за препоруку једни су од најпопуларнијих апликација науке над подацима. Користе се да

предвиде “оцену” коју ће корисник дати некој ставци или “преференције” корисника. Скоро свака већа компанија их примењује у неком облику: Amazon их користи да препоручује производе корисницима, YouTube их користи да одлучи који ће видео следећи да пусти на autoplay, Facebook их користи да препоручи странице да се “лајкују” или пријатеље са којима би корисници могли да ступе у контакт. Штавише, неке компаније као Netflix и Spotify имају бизнис модел чији успех директно зависи од могућности вршења препорука. Заправо, Netflix је понудио милион долара 2009. године ономе који може да поправи систем за 10% [5].

Системи за препоруку могу бити подељени на три типа: једноставни системи за препоруку; системи за препоруку засновани на садржају; колаборативно филтрирање [5].

2.3.1 Једноставни системи за препоруку:

Ови системи често генерализују препоруке свим корисницима, које су засноване на популарности филма или жанра. Основна идеја иза овог система је да ће филмови који су популарнији и критички прихваћенији имати већу шансу да се свиде просечном кориснику. IMDb Top 250 пример је овог система.

2.3.2 Системи за препоруку засновани на садржају

Овакви системи за одређену ставку препоручују сличне ставке. За препоручивање се користе метаподаци, као што су жанр, режисер, опис, глумци итд. Генерална идеја иза ових система за препоруку је та да ако се особи свиђа одређена ставка, њему или њој ће се свидети слична ставка.

2.3.3 Колаборативно филтрирање

Системи за препоручивање засновани на колаборативном филтрирању покушавају да предвиде оцену коју би корисник могао да да одређеној ставци на основу претходних оцена и преферитета других корисника. Такви системи за препоручивање не захтевају метаподатке о ставкама као системи за препоруку засновани на садржају.[6] Матрица на основу које се врши колаборативно филтрирање као једну димензију има све кориснике, а као другу ставке које корисници оцењују оценама у неком одређеном опсегу [5].

Неке од метода коришћене за прорачуњавање сличности су: *Jaccard*-ова сличност; косинусна сличност; центрирана косинусна сличност (Пирсонова корелација)

Најчешћи алгоритам који се користи за одређивање сличности је Пирсонова корелација (*Pearson Correlation*).

2.3.3.1 Пирсонова корелација (*Pearson Correlation*)

Коефицијент Пирсонове корелације је мера снаге линеарне везе између две варијабле.

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Пирсонова корелација може да буде у опсегу од -1 до 1. Ако је -1, онда је то савршена негативна линеарна веза између варијабли; ако је 0, онда значи да не постоји линеарна веза међу варијаблама; ако је корелација 1, онда значи да је савршена позитивна линеарна корелација међу варијаблама [6].

2.3.4 Модел латентних фактора (LFM)

Моделовање латентних фактора мапира како кориснике тако и ставке у заједнички простор латентних фактора R^K (свака димензија енкодира прикривени фактор) [7].

2.4 Студијски модули на смеру Рачунарство и аутоматика

Након што студенти заврше другу годину на Факултету техничких наука одабирају један од наредна три студијска модула: примењене рачунарске науке и информатика (ПРНИ); рачунарска техника и рачунарска комуникација (РТ-РК); аутоматика.

Студенти који су изабрали модул Примењене рачунарске науке и информатика у четвртој години бирају један од четири студијска модула: интернет и електронско пословање; информациони системи; интелигентни системи; софтверско инжењерство.

3. СКУП ПОДАТАКА

Студентске оцене добављене су од Факултета техничких наука у *Excel* фајлу.

У фајлу се налазе оцене 2420 студената при чему је задржана анонимност студената.

Свака оцена је приказана кроз три атрибута: идентификациони број студента (насумичан број - јединствен за студента) ; назив предмета; оцена студента из тог предмета. Због мењања назива предмета кроз време, првобитно су се називи предмета морали мапирати на данашње називе предмета. Пример: Пре неколико година назив данашњег предмета “Алгебра” био је “Дискретна математика и линеарна алгебра”. Након тога је креиран .csv фајл у коме редови представљају студенте, а колоне представљају предмете. У том формату се брже добављају оцене одређеног студента, такође боља је и визуелна репрезентација расподељености студентских оцена.

4. ПРЕПОРУКА СТУДИЈСКОГ МОДУЛА

У циљу помоћи студентима на трећој и четвртој години основних академских студија, који морају да бирају наредни студијски модул, креиран је систем за препоруку истог где се врши препорука на основу оцена студента.

4.1 Креирање онтологије

На основу добијеног .csv фајла врши се популисање, односно попуњавање онтологије индивидуалима у *Python* језику.

Разлог коришћења онтологије је у томе што су складиштени одвојено од било ког система, што значи да ако се направи неки други систем који користи оцене студената може да приступи овој бази података и користи те податке за своје потребе.

4.2 Архитектура система

Архитектура система састоји се од неколико микросервиса који функционишу у кластеру зарад модуларизације система на одвојене целине, које функционишу тако што сваки сервис обавља један “задатак” и они су: *service-registry*; *config-server*; *api-gateway-service*; *module-recommender-service*.

4.3 Портал за препоруку модула

Портал са препоруку модула представља *module recommender service* који се састоји од *frontend* и *backend* дела.

4.3.1 Препоруке

У раду је имплементирано неколико врста система за препоруку.

Све препоруке садрже неке заједничке кораке као што су: учитавање података са семантичког веба; креирање матрице свих студената.

Учитавање података са семантичког веба - подаци о свим студентима су претходно смештени на *Jena Fuseki* сервер, одакле се прикупљају уз помоћ *SPARQL* упита.

Креирање матрице свих студената - након што се добава подаци, они се морају приказати у неком формату, у овој ситуацији је најадекватнији приказ матрица студент/оцена.

4.3.1.1 Колаборативно филтрирање

Препорука заснована на колаборативном филтрирању вршена је на следећи начин.

Проналазак најсличнијег студента - врши се користећи подкуп матрице свих студената и свих оцена.

Одабир модула - Проналази се најсличнији студент активном студенту уз то да се као додатни фактор узима и његов просек.

$$\text{value} = (\text{value} + 1) * 5 * (\text{sum} / \text{count})$$

- **value** - вредност Пирсонове корелације
- **sum**- сума свих оцена студента
- **count**- број оцена студента

Домен вредности је на овај начин скалиран на опсег од 0 до 100 где 0 (нула) значи да су потпуно различити студенти, а 100 (сто) да је студент вредан препоруке.

Од препорученог студента преузима се модул на ком је он и препоручује се активном студенту као модул који би он требао да одабере на наредном избору модула.

4.3.1.2 Модел латентних фактора (LFM)

Препорука заснована на моделу прикривених фактора вршена је на следећи начин.

Креирање LFM матрице - *LFM* матрица креирана је по узору на препоруку филма [8].

Алгоритам коришћен при креирању излазне матрице је објашњен у области 2.3.4.

Излаз алгоритма је матрица, где су све оцене замењене оценама које представљају студентове “препоручене оцене” из свих предмета.

Одабир модула - Модул који има највиши просек оцена се препоручује активном кориснику.

4.3.2 Крајње тачке сервиса

Акцент пројекта је на *module-recommender* сервису који садржи неколико крајњих тачака које могу да се “гађају” са неког *frontend-a* или из неког програма за тестирање као што је *Postman*.

4.4.2.1 Колаборативно филтрирање (Collaborative filtering)

На основу оцена студента систем врши препоруку тако што проверава оцене свих студената на основу Пирсонове корелације. Модул коме припада најсличнији студент се приказује као препорука активном студенту.

4.3.2.2 Провера колаборативног филтрирања

Тачност система за препоруку, заснованог на колаборативном филтрирању, тестираног на студентима који се већ налазе у бази, износи **83.92%**.

4.3.2.3 Модел латентних фактора (Latent Factor Model - LFM)

На основу оцена студента врши се препорука тако што се примени *LFM* алгоритам на све студенте укључујући и активног, и из добијене матрице се прочита препорука предмета и ком модулу он припада.

4.3.2.4 Провера LFM

Тачност система за препоруку, заснованог на *LFM*, тестираног на студентима који се већ налазе у бази, износи **30.04%**.

Разлог зашто је резултат тако низак је тај што се у рачуницу узима цела матрица за разлику од колаборативног филтрирања где се узимају само оцене активног студента.

5. ЗАКЉУЧАК

У овом истраживању анализирана је могућност развоја и примене система за препоручивање студијских модула на основу студентских постигнућа. Посматрана су два приступа: колаборативно филтрирање и методе латентних фактора.

Истраживање је показало да примена методе колаборативног филтрирања, и ако је “примитивнија” даје боље резултате од методе латентних фактора.

Као најважнија предност методе колаборативног филтрирања издвојена је чињеница да се у обзир узимају све оцене студента, док се у методи латентних фактора врши редукција димензионалности при чему долази до губитка информација и препорука се врши на основу смањеног броја атрибута.

Предност методе латентних фактора је у томе што користи мању матрицу и у огромним системима може да врши препоруке на значајно мањем броју атрибута, што се може користити у случајевима ограничености хардвера.

Систем за препоруку студијског модула на основу колаборативног филтрирања врши препоруку само на основу оцена, али не може да предвиди људске преферитете односно интересовања студената.

Могуће је да би систем за препоручивање студенту препоручио модул на коме би студент остварио виши просек занемаривши интристичку мотивацију и интересовања студента. Стога један од основних праваца даљег истраживања представља укључивање података о интересовањима у систем за препоручивање студијских модула.

Такође, као један од важних праваца даљег истраживања издваја се могућност примене других техника рачунарске интелигенције за развој оваквог система, као што су препоруке вршене уз помоћ неуронских мрежа или чак неке од метода које служе за класификовање података.

Препоруку студијских модула треба посматрати као интегрални део шире слике образовног процеса па се због тога интеграција система предложеног овим истраживањем са другим дигиталним системима који су део образовног процеса издваја као важан правац даљег истраживања.

6. ЛИТЕРАТУРА

- [1] - <https://www.w3.org/standards/semanticweb/ontology>
- [2] - <https://www.w3.org/OWL/>
- [3] - <https://www.w3.org/RDF/>
- [4] - <https://www.w3.org/TR/rdf-sparql-query/>
- [5] - <https://www.datacamp.com/community/tutorials/recommender-systems-python>

[6] - : http://onlinestatbook.com/2/describing_bivariate_data/pearson.html

[7] - Learning Personal+Social Latent Factor Model for Social Recommendation; Yelong Shen, Ruoming Jin; Department of Computer Science Kent State University.

[8] - MF-Based-Recommendation
<https://github.com/ShantanuDeshpande/MF-Based-Recommendation>

Кратка биографија:



Светозар Стојковић рођен је у Новом Саду 1994. године. Основне академске студије завршио је 2017. године. Мастер рад на Факултету техничких наука из области Електротехнике и рачунарства - Софтверски систем за препоруку студијског модула на смеру рачунарство и аутоматика одбранио је 2018. године.

контакт:

svetozar.cvele.stojkovic@gmail.com