

СОЦИАЛНИ ПРЕПОРЪКИ БАЗИРАНИ НА СЪДЪРЖАНИЕ И ДОВЕРИЕ

Милен Чечев, Иван Койчев

Софийски Университет „Св. Климент Охридски“
{milan.chechev, koychev}@fmi.uni-sofia.bg

Резюме: Статията предлага анализ на проблема за препоръчване на информация в социалните мрежи. Разглеждат се основните типове препоръчващи подходи, като се акцентира на препоръките базирани на съдържание и на доверие. Предложен е подход за изчисляване на доверието на потребител в неговите приятели. Сравнени са 3 различни подхода за препоръчване на информация – базиран на съдържание, препоръки въз основа на доверие и хибриден подход. Експериментите са направени върху реални данни събрани от социалната мрежа Фейсбук.

Ключови думи: социално препоръчване, препоръки, препоръки базирани на доверие, препоръки базирани на съдържание, социални мрежи, препоръчваща система

1. Въведение

Препоръките в социалната мрежа са основен инструмент необходим и изискван от потребителите за улесняване на тяхната работа и съкращаване на времето отделяно за търсене на новини, текстово и мултимедийно съдържание. Най-голямата социална мрежа Фейсбук, предоставя различни видове препоръки - препоръчват се новини, различни игри, страници на известни личности или компании, реклами и др., като подходите за препоръчване са скрити и с много ограничена функционалност за ръчна настройка. Липсата на обяснение защо дадени обекти са избрани от системата за препоръки и невъзможността за по прецизна ръчна настройка поражда недоверие на потребителите в тяхната достоверност. Направени проучвания [1] показват, че потребителите имат нужда от по-развита структура за търсене, сортиране и филтриране на съдържание с автоматични препоръки и възможност за ръчна фина настройка на техният профил.

Обектите споделяни в социалната мрежа имат две основни характеристики определящи разпространението им в мрежата – съдържание и доверие. Съдържанието се определя от ключовите думи описващи обекта и е статично за обекта, докато доверието е динамично и се определя от хората, които са споделили, харесали или коментирали обекта. При публикуване на нов обект в социалната мрежа първоначално доверието в него е базирано само на доверието в източника, който го е публикувал, но с харесването му и споделянето му от други потребители, доверието се променя във зависимост

от хората, които се свързват с него. Важно е да се отбележи, че доверието е индивидуална величина за всеки потребител и при измерване на доверие към даден обект винаги го измерваме като доверие от гледна точка на конкретен потребител.

2. Видове препоръчващи подходи в социалната мрежа

В социалните мрежи се използват различни препоръчващи подходи и комбинации от тях, като изборът на конкретен подход се базира на динамиката на обектите, които препоръчват и наличните исторически данни. Основните препоръчващи подходи са:

- Препоръки базирани на сътрудничество (collaborative filtering) – Основната идея на алгоритмите от този тип е използване на данни от историята на потребителите за намиране на сходност между тях, като препоръчвайки обект на потребителя избираме между обектите харесани от близките до него потребители. Основното предположение в този подход е, че близките в миналото по предпочитания потребители са близки по предпочитания и в настоящето.
- Препоръки базирани на съдържание (content based filtering) – При този подход всички обекти се разглеждат, чрез своето съдържание - ключови думи, тагове, мета-данни. Въз основа на обектите, свързани с потребител, се изгражда потребителски профил и за препоръчване на обект за потребител се оценява близостта на съдържанието му до потребителския профил. За сравняване на различните обекти и потребителския профил обикновено се използва векторно-пространствен модел, като всички обекти и потребителските профили се представят като N мерни вектори [3].
- Препоръки базирани на доверие – при този подход се препоръчва информация въз основа на това, доколко потребителят има доверие в източниците свързани с нея. Доверието в различните източници на информация е динамично и индивидуално за всеки потребител, като в рамките на социалната мрежа за определянето му използваме информацията за харесванияте, споделените и коментирани от потребителя обекти в неговата история.

3. Свързани изследвания

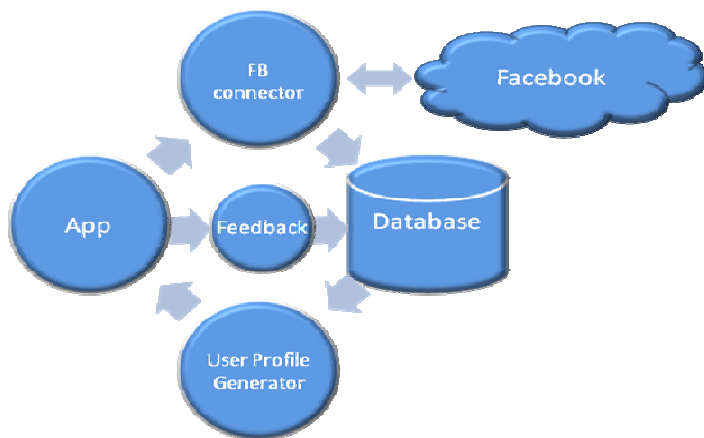
Препоръките в социалната мрежа използват данни за връзките между потребителите. Тези връзки в много изследвания са моделирани като връзки на доверието [4, 5, 6]. Изследванията в областта на препоръчващите системи базирани на доверие определят два вида доверие между потребителите – явно и неявно. Пример за мрежа на доверието, в която имаме явно указано

доверие е мрежата за ревюта Epinions. Маса и Авесани [4] изследват разпространението на доверието и недоверието в мрежата. Изследването им показва, че подходите базирани на доверие имат по-малка зависимост от обема на данните и работят дори и когато имаме сравнително малко историческа информация за потребителя.

Социалната мрежа Фейсбук също може да се разглежда като мрежа на доверието, при нея обаче за разлика от Epinions нямаме явно задаване на доверие. За извличане на доверието на потребител към негов приятел следва да се използва показаната активност към публикуваното съдържание от него. В [7] авторите използват Facebook, като мрежа на доверието, като комбинират доверието със сътрудничество между потребителите. Други препоръчващи системи за Facebook са показани в [8, 9, 10] като в [8] авторите прилагат хибриден подход и акцентират върху предоставянето на обяснения за препоръките и визуализацията им. В [9] авторите описват приложение във Фейсбук използващо за препоръчване социално факторизиране на матрици, при което се изгражда модел подобен на препоръките със сътрудничество, но използващ и социалните елементи на мрежата. В [2] авторите представят препоръчваща система, която препоръчва филми на групи от потребители.

4. Реализация

За целите на експериментите е реализирано Фейсбук приложение разработено да предостави на потребителите алтернативен начин за разглеждане на техният поток от новини. То събира данни от Фейсбук за потребителя и неговите приятели и предоставя препоръки.



Фигура 1. Архитектура на фейсбук приложение за препоръчване на информация

За реализацията му са използвани следните модули (фиг. 1): модул за комуникация със Фейсбук, модул за работа с база данни, модул за съставяне на потребителски профил и препоръки, модул за обратна връзка с потребителя, модул за представяне на данните (потребителски интерфейс).

Приложението използва програмните интерфейси Graph Api и FQL и извлича информация за новините публикувани от потребител и неговите приятели, новините попадащи в неговият поток от информация и обектите, които е харесал. Поради ограничения от социалната платформа, разполагаме с последните 1000 харесвания на потребител и около 300 публикувани новини на страницата на всеки приятел на потребителя. Експериментите са направени върху данните от 5 тестови потребители и техните 1453 приятели.

4.1. Препоръки базирани на съдържание

Съдържание наричаме формално описание на обект изградено от текстово описание в комбинация с метаданни и тагове. За изграждане на съдържание от обектите в социалната мрежа използваме текстовите им свойства: име, описание, съобщение и т.н. както и характеристики като тип на съобщение, наличност на снимка. Всички обекти се представят като N-мерни вектори, като всяка ключова дума от съдържанието е една размерност, а стойността на размерността се определя от метриката TFIDF. За изграждане на потребителски профил използваме обектите харесани, коментирани или споделени от потребителя, като от тях извличаме ключовите думи и мета данни. В направените експерименти сме приложили извличане от всяко съдържание на 10 ключови думи, спрямо уникалността им в разглежданото множество от документи, и агрегирането им в потребителски профил. След превръщането на всички обекти и потребителски профили във вектори от думи, приликата между потребител и обект изчисляваме чрез формулата за косинусова прилика (1). Където d_{α} е документ-вектор, u е потребител, m е размерността на векторното пространство, $d_{\alpha i}$ и u_i са съответно стойностите на документа и потребителският профил в измерение i .

$$\text{similarity}(d_{\alpha}, u) = \frac{d_{\alpha} \cdot u}{|d_{\alpha}| |u|} = \frac{\sum_{i=1}^m d_{\alpha i} u_i}{\sqrt{\sum_{i=1}^m d_{\alpha i}^2} \sqrt{\sum_{i=1}^m u_i^2}} \quad (1)$$

4.2. Препоръки базирани на доверие

Доверието в Фейсбук не се задава явно, но може да бъде извлечено от данните за активността на потребителите. В някои статии за простота доверието се разглежда като симетрична величина с цел да се опрости модела [7], в реалността обаче доверието е възможно да бъде асиметрично за това в нашата система то има посока и се изчисляването със формула (2).

$$\text{trust}_{ij} = \frac{\alpha \cdot \text{like}_{ij} + \beta \cdot \text{comment}_{ij} + \gamma \cdot \text{share}_{ij}}{\text{published}_j} + \mu \cdot \frac{\text{activity}_{ij}}{\text{activity}_i} \quad (2)$$

В формула (2) α , β , γ и μ са константи, like_{ij} , comment_{ij} , share_{ij} са съответно броят на харесаните, коментираните и споделените обекти от потребител i , които са публикувани от потребител j . published_j е броят на обектите публикувани от потребител j , activity_{ij} – броят на обектите, които са едновременно харесани, коментирани или споделени едновременно от потребителите i и j , а activity_i – броят на обектите харесани, коментирани или споделени от потребител i . За направените експерименти са избрани стойности на параметрите $\alpha=2$, $\beta=1$, $\gamma=3$ и $\mu=0.5$. За изчисляване на степента на харесване на обект x от потребител i се използва формула (3) където $\text{trust}_{\text{author}}$ е доверието на потребителя i в източника публикувал обекта, а $\text{trust}_{\text{related users}}$ е сумата от доверията към потребителите, коментирали или харесали публикацията.

$$\text{trust}_{ix} = \frac{\text{trust}_{\text{author}_x} + \text{trust}_{\text{related users}_x}}{\max(\text{trust}_i)} \quad (3)$$

За нормализиране на формулата разделяме на максималното доверие, което потребител е възможно да има към обект. Максималното доверие би се получило ако всичките му приятели се асоциират с този обект:

$$\max(\text{trust}_i) = \sum_{j \in \text{friends}_i} \text{trust}_{ij}.$$

4.3. Комбинирани препоръки

Направени са експерименти и с хибриден алгоритъм комбиниращ алгоритмите базирани на съдържание и доверие. За комбиниране на резултатите се използва следната формула (4):

$$\text{score}_{ix} = \lambda \text{trust}_{ix} + (1 - \lambda) \text{similarity}_{ix} \quad (4)$$

Комбинирането на двата подхода би трябвало да обедини източниците на доверие със съдържанието от което се интересува потребителя, като едновременно да се дават препоръки за обекти на подобни теми на разглежданите от нас обекти и за обекти, които са на нови, неразглеждани от потребителя до сега тематики, но са асоциирани с източниците му на доверие. При експериментите се наблюдава малко по-добра точност на хибридният алгоритъм над алгоритъма базиран на доверие при извличане на повече препоръки. В рамките на експеримента бяха изследвани различни стойности на параметъра λ като за крайните резултати е избрана стойността, която дава най-голямо увеличение на точността при разглеждане на на-добрите 300 резултата.

5. Оценяване

За извършване на експерименти за оценяване на точността на алгоритмите, наличните данни са разделени на 2 множества – обучаващо и тестващо. Разделението е направено след подреждане на данните хронологично и намиране на подмножествата от обекти съдържащи първите 90% от харесаните от потребителя обекти и следващите 10%. По-старото множество е използвано за обучение, а по-новото за тестване.

За оценяване на точността на алгоритмите са използвани метриците откриване(recall) и точност(precision). При метриката откриване се измерва каква част от релевантните резултати са намерени от системата (формула 5).

$$recall = \frac{|D_{sr}|}{|D_s|} \quad (5)$$

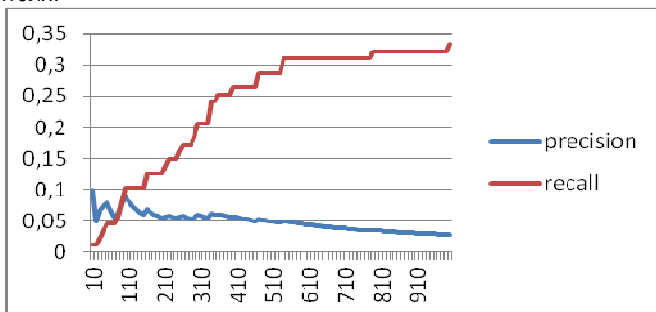
При метриката точност имаме измерване каква част от резултатите върнати от системата са релевантни(формула 6).

$$precision = \frac{|D_{sr}|}{|D_r|} \quad (6)$$

Във формулите са обозначени с $|D_{sr}|$ - броят на релевантните обекти върнати от системата, $|D_r|$ - броят на всички релевантни документи, $|D_s|$ – броят на всички документи върнати от системата.

6. Резултати

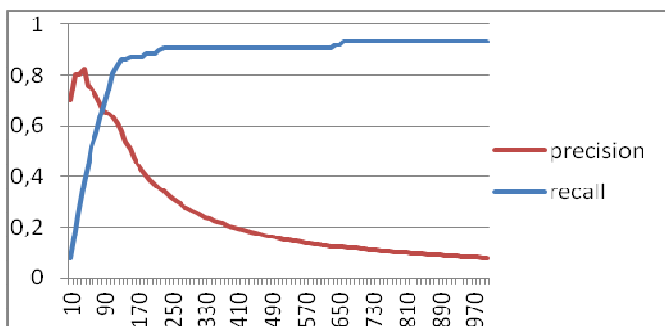
За всеки от изследваните подходи са направени експерименти за определянето на точността и откриването. Фигура 2 показва точността и откриването при препоръчване базирано на съдържание. Абцисната ос показва броя на препоръчаните примери, а ординатната стойността на точността и откриването. При препоръчване на около 100 примера имаме стойност от 0.1 едновременно за точността и откриването. Забелязва се, че броят откриването достига максимум 0.33 при препоръчани 1000 обекта на потребителя.



Фигура 2. Точност и откриване с препоръки базирани на съдържание

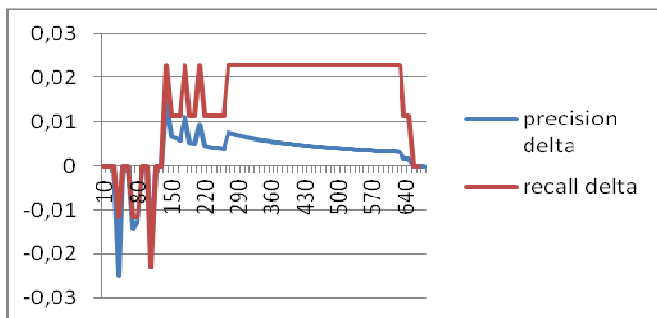
Следващата фигура показва точността и откриването при използването на доверие. При препоръките базирани на доверие се наблюдава по-добри стойности едновременно за точността и откриването. При препоръчване на около 50 обекта точността е 0,8, а при препоръчване на над 100 обекта откриването стига до внушителната стойност от 0,9.

При изследването на хибридният подход не са наблюдавани големи разлики с резултатите от подхода базиран на доверие, поради което за по-добра визуализация на фигура 4 представяме разликата между метриците на хибридният подход и подхода базиран на доверие.



Фигура 3. Точност и откриване при препоръки базирани на доверие

В резултатите показани от хибридният подход прави впечатления, че точността на хибридният подход се намаля в първите 100 препоръки, но за сметка на това след това се наблюдава скок на точността и откриването. Тези резултати може да се тълкуват като първо задължително да се показва резултатите от източниците в които имаме доверие, а след това да се прибегва до комбиниране на доверието и съдържанието.



Фигура 4. Разлика между точността/откриването на хибридният метод и препоръките базирани на доверие

Заклучение и бъдещо развитие

Направените експерименти показват голямата зависимост между харесваните от потребителя обекти и доверието в потребителите асоциирани с тях. Препоръките базирани на съдържание не генерират сами по-себе си достатъчно добри препоръки за потребителя, но имат потенциал да подпомагат системата с комбинирането им в хибридни алгоритми. Хибридният алгоритъм генерира по-ниска точност разглеждайки първите препоръки, но с увеличаване на бройката на генерираните препоръки се наблюдава по-високо откриване на релевантните документи.

Бъдещото развитие на проекта е свързано с интегриране на препоръки базирани на сътрудничество и обработка на данните за обратна връзка предоставени от потребителите.

Благодарности

Работата е частично финансирана от Европейския социален фонд чрез Оперативна програма „Развитие на човешките ресурси“, договор № BG051PO001-3.3.06 - 0052 (2012-2014).

Литература

1. Milen Chechev, Ivan Koychev. Recommendations in Social Networks: an Extra Feature or an Essential Need, In Proceedings of MIE'2013, Sofia, Bulgaria, September 2013
2. Lara Quijano-Sánchez, Juan A. Recio-García, Belén D'íaz-Agudo, and Guillermo Jiménez-Díaz, HappyMovie: A Facebook Application for Recommending Movies to Groups, ICTAI, 2011
3. Salton, G., Wong, A., Yang, C.: A vector space model for automatic indexing, Communications of the ACM, 1975
4. Massa, P., Avesani, P.: Trust-aware collaborative filtering for recommender systems. In: Proceedings of the International Conference on Cooperative Information Systems, CoopIS, pp. 492–508 (2004)
5. O'Donovan, J., Smyth, B.: Is trust robust?: an analysis of trust-based recommendation. In: IUI'06: Proceedings of the 11th international conference on Intelligent user interfaces, pp. 101–108. ACM, New York, NY, USA (2006)
6. Fuguo, Z., Shenghua, X.: Analysis of trust-based e-commerce recommender systems under recommendation attacks. In: ISDPE '07: Proceedings of the The First International Symposium on Data, Privacy, and E-Commerce, pp. 385–390. IEEE Computer Society, Washington, DC, USA (2007)
7. Wei Chen, Simon Fong, Social Network Collaborative Filtering Framework and Online Trust Factors: a Case Study on Facebook, International Journal of Web Applications, Volume 3, November 2011

8. Svetlin Bostandjiev, John O'Donovan, Tobias Höllerer. TasteWeights: A Visual Interactive Hybrid Recommender System, ACM RecSys'12, 2012
9. Joseph Noel, Scott Sanner, Khoi-Nguyen Tran, Peter Christen, Lexing Xie, Edwin V. Bonilla, Ehsan Abbasnejad, Nicolás Della Penna. New Objective Functions for Social Collaborative Filtering, WWW'12, 2012

SOCIAL RECOMMENDATION BASED ON CONTENT AND TRUST

Milen Chechev, Ivan Koychev

Abstract: *This paper present analysis on different recommendation approaches for social networks. The main accent is on content based filtering and trust aware recommendations as new formulas for calculating trust are suggested. The experiments are made on data gathered from the social network Facebook via custom made application, as 3 different approaches are compared – content-based filtering, trust based recommendation and hybrid approach. Results shows that trust calculated from the social activities of the users is good source for recommendation as with the hybrid approach the recall is increased.*

Keywords: *social recommendation, trust-based recommendation, content-based filtering, social networks, recommendation system*