

**Vitaliy Khokh, postgraduate**

*Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine*

### **Methods of Automated Sentiment Analysis on Social Networks Automated System for Auditing Information Security of Computer Systems and Networks**

The article deals with the system for the audit of information security of computer systems and networks developed within the framework of the dissertation research.

The first part of the system is the control center of the system, with which the operator, the auditor works directly. The second part is a knowledge base that is presented as a SQL server in the system. The third part is a device or node, an audit tool provider, or an attack module. The control center is the main and controls all other nodes, with it directly staff. The control center of the system must be connected to the database server, which has a knowledge base, and a connection to the module of the tools provider must be provided. The knowledge base in the system is represented by a relational database of SQL. The primary tasks of the audit tool provider are to accept and execute instructions from the control center, and since the control center will attempt to contact it using the SSH protocol, the provider must be able to deploy the SSH server and accept the connection.

The system is developed in such a way as to make decisions by working closely with the user. The system allows you to capture the number of facts that people may not reach, or it becomes inconvenient. The system does not forbid inclusion in the calculation of facts that interest the user - on the contrary, it complements them, in the event that the inclusion of one custom fact will not affect the operation of the system, because the data associated with it are outside the current state of the factual database, or Products that actively use this fact are beyond the "interests" of the expert system, with the data that it currently has.

At the moment two versions of the system have been developed, both of them can be considered as early alpha versions. In the first version, attention was focused on the interaction of nodes of the system and the main mechanisms of the expert system. In the second version, the structure of the system has been redesigned so that the system becomes more flexible to scaling.

**information security, audit, fuzzy logic, expert systems, automation**

Одержано (Received) 18.05.2018

**УДК 004.9**

**Д.В. Шингалов, асп., Є.В. Мелешко, доц., канд. техн. наук, Р.М. Минайленко, доц., канд. техн. наук., В.А. Резніченко, викл.**

*Центральноукраїнський національний технічний університет, м. Кропивницький, Україна, E-mail: elismelashko@gmail.com, E-mail: dimashingalov@gmail.com*

## **Математична модель рекомендаційної системи з врахуванням емоційного забарвлення коментарів у якості контексту**

У статті пропонується математична модель рекомендаційної системи, у якій у якості контексту використовується аналіз емоційного забарвлення коментарів стосовно об'єктів рекомендацій. При відсутності явного зворотного зв'язку аналіз контексту значно підвищує точність рекомендацій та якість прогнозування вподобань користувачів.

**рекомендаційні системи, сентимент-аналіз, колаборативна фільтрація, машинне навчання, інтелектуальні системи**

**Д.В. Шингалов, асп., Є.В. Мелешко, доц., канд. техн. наук, Р.М. Минайленко, доц., канд. техн. наук, В.А. Резніченко, преп.**

*Центральноукраїнський національний технічний університет, г. Кропивницький, Україна*

© Д.В. Шингалов, Є.В. Мелешко, Р.М. Минайленко, В.А. Резніченко, 2018

## **Математическая модель рекомендационной системы с учетом эмоциональной окраски комментариев в качестве контекста**

В статье предлагается математическая модель рекомендательной системы, в которой в качестве контекста используется анализ эмоциональной окраски комментариев относительно объектов рекомендаций. При отсутствии явной обратной связи анализ контекста значительно повышает точность рекомендаций и качество прогнозирования предпочтений пользователей.

**рекомендационные системы, сентимент-анализ, коллаборативная фильтрация, машинное обучение, интеллектуальные системы**

**Постановка проблеми.** Рекомендаційна система, загалом, може бути визначена як програмний засіб, який надає корисні поради користувачам для того, щоб допомогти їм прийняти рішення щодо певного товару. Рекомендаційні системи були вперше впроваджені на сайтах електронної торгівлі. При застосуванні будь-якої рекомендаційної системи важлива наявність зворотного зв'язку (подія, за якою можна напряму судити про реакцію користувача на об'єкт).

Так, наприклад, рекомендації за оцінками за п'ятибальною шкалою – приклад задачі з явним зворотним зв'язком. Рекомендаційні системи, що керуються актами покупок, відвідуванням сторінок – приклади задач з неявним зворотним зв'язком. У разі неявного зворотного зв'язку є невизначеність в тому, позитивно чи негативно користувач оцінив об'єкт, який придбав чи переглянув. Купівля товару в Інтернет-магазині може означати досягнення користувачем своєї споживчої мети (і його реакція на товар позитивна), але в той же час покупець міг після отримання товару в ньому розчаруватися і правильно було б захувати негативну реакцію. Очевидно, що відвідування сторінок веб-сервісу користувачами може відбуватися при абсолютно різному ступені зацікавленості користувача в контенті.

Варто відзначити досить типову ситуацію, коли рекомендаційній системі подаються на вхід виключно позитивні приклади взаємодії користувачів і об'єктів. Наприклад, веб-сервіс Twitter не має функціональності, що дозволяє користувачеві виставити негативну оцінку контенту, а присутній тільки лише спосіб «заохотити» той чи інший контент, поширивши його своїм підписникам за допомогою функції «репост». Подібний зворотний зв'язок користувача дуже надійно (у порівнянні з іншими) вказує на позитивну реакцію. Надійність «репостів» у сервісі Twitter підкріплено відповідальністю користувачів перед своїми підписниками.

Але, при дослідженні реакцій користувачів на тих чи інших сервісах, де є тільки позитивні відгуки, велика кількість користувачів є не охопленою дослідженням, тому можна вважати прогнозування рекомендацій не дуже ефективним. До того ж, не можна з упевненістю говорити про ступінь прихильності користувачів до того чи іншого об'єкту рекомендацій, тому що, по-перше відсутній альтернативний спосіб висловлення прихильності, а по-друге немає впевненості у тому, що прихильність не була обрана випадково. Для більш гнучкого аналізу прихильності та подальшого прогнозування рекомендацій доцільно застосувати аналіз емоційного забарвлення тексту коментарів, для уточнення або одержання даних про оцінки контенту (у випадку їх явної відсутності).

Вирішення проблеми якісного прогнозування вподобань, при відсутності широкого спектру інструментів для явного виявлення вподобань щодо об'єктів рекомендацій є важливою задачею при створенні ефективної, гнучкої та універсальної рекомендаційної системи.

**Аналіз останніх досліджень і публікацій.** В роботах [1, 3] систематизовано методологію отримання даних при відсутності явного виявлення вподобань користувачами соціальних мереж (оцінки, лайки, дизлайки, переходи за посиланнями)

та застосовано аналіз контекстної інформації (місце, час, вік). У [2, 4] запропоновано застосування методу пониження рангу матриці оцінок та застосування факторизації до додаткових векторів ознак, з метою прогнозування вподобань. Також у роботі [5] розглянуто аналіз хештегів для розширення аудиторії застосування рекомендаційних систем. Для аналізу емоційного забарвлення тексту коментарів найчастіше застосовуються комбіновані методи, що базуються як на спеціалізованих словниках, так і на підходах машинного навчання, як видно з джерел [6, 7, 8, 9]. Для зниження складності та пришвидшення математичних обчислень застосовується перехід до афінних перетворень [10], де навчання рекомендаційної системи можна звести до канонічної покоординатної оптимізації.

**Постановка завдання.** Метою даної роботи є створення математичної моделі рекомендаційної системи на основі колаборативної фільтрації з використанням аналізу емоційного забарвлення тексту коментарів для одержання уточнюючого контексту.

Для підвищення точності виявлення вподобань користувача в системах, де не передбачено негативних оцінок контенту, або для уточнення ступеню позитивної оцінки об'єкту, доцільно застосувати оцінку емоційного забарвлення текстів коментарів користувачів, як приклад врахування контексту. Також емоційне забарвлення коментарів можна використовувати замість оцінок, у випадках, коли користувач не поставив оцінку, але залишив коментар. Задачею даного дослідження є спроба на основі матричних обчислень вивести математичну модель для подальшого створення контекстно-орієнтованої рекомендаційної системи.

**Виклад основного матеріалу.** Одним з основних підходів до моделювання рекомендацій з врахуванням контекстної інформації є факторизаційні моделі [1]. Пропонується використовувати факторизаційні машини, після аналізу тексту коментарів, контекстну інформацію про емоційне забарвлення тексту представляти у вигляді додаткових ознак для факторизаційної машини. Пропонується кодувати контекст в додаткові координати афінора [10] оцінок, здійснювати пониження рангу афінора оцінок, також можна застосовувати попарні афінні перетворення.

Розглянемо задачу колаборативної фільтрації [2], нехай є кінцева множина користувачів  $U$  і кінцева множина об'єктів  $I$ . Нехай також є набір пар  $R = \{(u, i)\} \subset U \times I$ , для яких відома реакція користувача  $r_{ui}$ . Потрібно за навчальним набором  $(R, \{r_{ui}, (u, i) \in R\})$  побудувати регресор  $\hat{r}(u, i)$ , що передбачає реакцію користувача для довільної пари  $(u, i) \in U \times I$ .

У разі, якщо кожен користувач  $u \in U$  та кожен об'єкт  $i \in I$  піддаються змістовному опису ознаковими векторами:  $x(u): U \rightarrow \mathbb{R}^d$ ,  $x(i): I \rightarrow \mathbb{R}^d$ , то задачу колаборативної фільтрації можна звести до стандартної задачі навчання на прецедентах: потрібно навчити модель, яка по вектору ознак  $x(u, i) = [x(u); x(i)]$  передбачає значення регресійної змінної  $r_{ui}$ . Такий підхід в рекомендаційних системах називається заснованим на описах (content-based) [5].

Для вирішення задачі колаборативної фільтрації: реакції (оцінки) користувача [3] можна записати у вигляді матриці  $R \in \mathbb{R}^{U \times I}$ , в якій рядки відповідають користувачам, а стовпці – об'єктам. У матриці  $R$  відомі деякі елементи, інші – пропущені. Треба побудувати правило  $\hat{r}(u, i)$ , навчитися відновлювати пропущені значення матриці  $R$ . Для заповнення пропусків в матриці  $R$  використовують припущення про її низький ранг та будують наближення  $R \approx XYT$ , де  $X \in \mathbb{R}^{|U| \times K}$ ,  $Y \in \mathbb{R}^{|I| \times K}$ ,  $K$  – передбачуваний ранг матриці  $R$ . Обмеження рангу в деякому сенсі є регуляризацією, що надає пониженню рангу узагальнюючу здатність. У такому підході передбачена реакція користувача є скалярним добутком відповідних рядків матриць  $X$  та  $Y$ :

$$\hat{r}(u, i) = x_u^T y_i, x_u \in \mathbb{R}^K, y_i \in \mathbb{R}^K, \quad (1)$$

Підхід з пониженням рангу використовує виключно інформацію про пари  $(u, i)$  та відповідні до них значення реакцій. Відомо, що в рекомендаційних системах велика частина інформації міститься саме в знеособлених трійках  $\{(r_{ui}, u, i)\}$  [2, 3]. Методи пониження рангу дають значно кращу точність передбачення  $\hat{r}(u, i)$  у порівнянні з content-based методами. Проте, облік відомої інформації про користувача та об'єкти, доповнений пониженням рангу, дає кращу точність і дозволяє впоратися з деякими проблемами рекомендаційних систем (такими, як холодний старт). Існує велика кількість моделей для колаборативної фільтрації, заснованих на базовій ідеї пониження рангу (1), але з модифікаціями для обліку зовнішньої інформації.

Модель факторизаційних машин [2, 3] покликана узагальнити різні модифікації моделі (1). Шляхом задання відповідного вектора ознак  $x(u, i) \in \mathbb{R}^d_u$  для пари користувач-об'єкт можна наводити передбачувальну модель  $\hat{y}(x)$  до різних факторизаційних моделей. Відмінність факторизаційної машини від класичних алгоритмів навчання по прецедентах – моделювання взаємодій між компонентами вектора ознак, що призводить до того, що індикаторні ознаки у вхідному векторі моделі  $x$  обробляються як в моделі пониження рангу (1).

У рекомендаційних системах найчастіше має місце проблема відсутності фіксації негативних реакцій [4, 5, 6]. Це обумовлено тим, що веб-сервісам набагато простіше і надійніше накопичувати акти про позитивні реакції користувача (наприклад, «репости» в Twitter), ніж дізнаватися у користувачів про їх негативні реакції за будь-якою шкалою (наприклад, «число зірок» в Netflix). Такі дані прийнято називати даними без явного відгуку. Без залучення додаткових припущень, навчання моделі  $\hat{r}(u, i)$  на таких даних немає сенсу – найкраща модель буде для будь-якої пари  $(u, i)$  передбачати позитивну реакцію користувача на об'єкт. Очевидно, така модель має погану узагальнюючу здатність.

Для роботи з даними без явного відгуку використовують припущення про те, що пропуски в матриці реакцій (оцінок)  $R$  частіше відбуваються через негативну реакцію, ніж позитивну.

В запропонованій моделі пропущені значення в матриці  $R$  замінюються на бали, отримані після аналізу емоційного забарвлення коментарів про об'єкт рекомендацій у разі наявності відповідних коментарів. Далі будується канонічне афінне розкладання афінора оцінок:

$$\hat{r} = (u, i, c_1, c_2, \dots, c_M) \approx r_{u, i, c_1, c_2, \dots, c_M}, \quad (2)$$

де  $c_1 \in C_1, \dots, c_M \in C_M$ , а  $C_1, \dots, C_M$  – деякі кінцеві множини, як правило вони використовуються для опису контексту рекомендації (в даній моделі вони містять дані, одержані після аналізу емоційного забарвлення тексту коментарів).

Загалом задача прогнозування вподобань користувачів рекомендаційною системою зводиться до заповнення матриці  $R$  даними про реакції користувачів, уточненням їх після сентимент-аналізу коментарів та покоординатної оптимізації регуляризованих квадратичних втрат. Будується канонічне афінне розкладання рангу  $K$ , тобто кожен елемент афінора  $T$  розміром  $S_I \times \dots \times S_D$  представляється у вигляді:

$$\dot{T}_{i_1, \dots, i_D} = 1^T M_{:i_1}^{(1)} \circ M_{:i_2}^{(2)} \circ \dots \circ M_{:i_D}^{(D)}, \quad (3)$$

де  $\{M^{(d)}\}_{d=1}^p$  – фактори канонічного розкладення,  $M^{(d)} \in R^{K \times S_d}$  матриці знаходяться шляхом мінімізації зваженого квадратичного відхилення елементів афінора:

$$\mathcal{L}(M^{(1)}, \dots, M^{(D)}) = \sum_{i_1=1}^{S_1} \dots \sum_{i_D=1}^{S_D} W_{i_1, \dots, i_D} \left( \dot{T}_{i_1, \dots, i_D} - T_{i_1, \dots, i_D} \right)^2, \quad (4)$$

де  $W$  – афінор вагів, що має розмір афінора  $T$ .

**Висновки.** В роботі запропонована математична модель рекомендаційної системи, у якій в якості контексту використовується аналіз емоційного забарвлення коментарів стосовно об'єктів рекомендацій. Також пропонується перетворювати емоційне забарвлення коментаря в оцінку для колаборативної фільтрації з метою заповнення пустих комірок у матриці оцінок або для уточнення ступеню позитивної/негативної реакції користувача у системах з неявним зворотним зв'язком.

## Список літератури

1. Bal'azs H. Fast als-based tensor factorization for context-aware recommendation from implicit feedback [Text] / H. Bal'azs, T. Domonkos // Machine Learning and Knowledge Discovery in Databases – 2012. – pp. 67–82.
2. Recommender systems handbook [Text] / Editors F. Ricci, L. Rokach, B. Shapira, P. Kantor. – Berlin: Springer, 2011. – 842 pages.
3. Kolda T. Tensor decompositions and applications [Text] / T. Kolda, B. Bader. // SIAM review. – 2009. – №51(3). – pp. 455–500.
4. Zhouxiao B. Movie Rating Estimation and Recommendation [Text] / B. Zhouxiao, X. Haiying. – 2012. – pp. 1–4.
5. Cantador I. Content-based recommendation in social tagging systems. [Text] / I. Cantador, A. Bellogn, D. Vallet. // ACM RecSys. – 2010. – №10. – pp. 237–240.
6. Burke R. Hybrid recommender systems: survey and experiments [Text] / R. Burke. // User Modelling and UserAdapted Interaction, vol. 12, no. 4. – 2002. – pp. 331–370.
7. Stuart D. What are Libraries Doing on Twitter? [Text] / D. Stuart. // Online 34, no. 1. – 2010. – pp. 45–47.
8. Mitchell T. Machine Learning [Text] / T. Mitchell. – New York: McGraw-Hill, 1997. – 414 p.
9. Narayanan V. Fast and accurate sentiment classification using an enhanced naive bayes model. [Text] / V. Narayanan, I. Arora, A. Bhatia // Intelligent Data Engineering and Automated Learning IDEAL – Berlin: Springer, 2013. – (volume 8206 of Lecture Notes in Computer Science). – pp. 194–201.
10. Норден А. П. Пространства аффинной связности [Текст] / А. П. Норден. – М.: Наука, 1976. – 432 стр.

## Список літератури

1. Bal'azs, H. & Domonkos, T. (2012). *Fast als-based tensor factorization for context-aware recommendation from implicit feedback*. 67–82.
2. Ricci, Editors F., Rokach, L., Shapira, B. & Kantor, P. (2011). *Recommender systems handbook*. Berlin: Springer.
3. Kolda, T. & Bader, B. (2009). Tensor decompositions and applications. *SIAM review*, №51(3), 455-500.
4. Zhouxiao, B. & Haiying, X. (2012). *Movie Rating Estimation and Recommendation*.
5. Cantador, I., Bellogn, A. & Vallet, D. (2010). Content-based recommendation in social tagging systems. *ACM RecSys*, 10, 237–240.
6. Burke, R. (2002). Hybrid recommender systems: survey and experiments. *User Modelling and UserAdapted Interaction*, vol. 12, 4, 331–370.
7. Stuart, D. (2010). *What are Libraries Doing on Twitter?* *Online* 34, 1, 45–47.
8. Mitchell, T. (1997). *Machine Learning*. New York: McGraw-Hill.
9. Narayanan, V., Arora, I. & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced naive bayes model. *Intelligent Data Engineering and Automated Learning IDEAL*, Vol. 8206, 194–201.
10. Norden, A. P. (1976). *Prostranstva affinnoj svjaznosti [Affine Connected Spaces]*. Moscow: Nauka.

**Dmitry Shyngalov, postgraduate, Yelyzaveta Meleshko, Assoc. Prof., Phd tech. sci., Roman Mynaylenko, Assoc. Prof., Phd tech. sci., Vitaliy Reznichenko, Lect.**

*Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine*

### **The Mathematical Model of the Recommendation System, Taking Into Account the Emotional Coloring of the Comments as a Context**

The article proposes the mathematical model of a recommendation system, in which a sentiment-analysis of comments related to a objects of recommendations is used as the context. An attempt is made to draw the mathematical model based on matrix calculations for the further creation of a context-oriented advisory system. In addition, an analysis of the feasibility of using the factorization model method to contextual filtration is used.

The paper investigates the structure of the recommendation system, in which the analysis of the emotional color of the comments to the objects of the recommendation is taken into account as a context. In the absence of explicit feedback, context analysis greatly improves the accuracy of the recommendations and the quality of prediction of user benefits. An attempt is also to deduce a mathematical model describing the work of such a recommendation system. Row down algorithms and the use of factoring machines for data processing are considered in data conditions without explicit response. Further practical application of hybrid joint filtration is considered as a solution to the practical implementation of the system of recommendations.

Rating models have huge potential for solving common filtration tasks, as well as for other subject areas in which there are interactions between objects of different types. Also, the proposed method allows you to change the structure of the model by changing the spaces of the space. The advantage of adding attributes can be applied to real data when modeling user preferences based on the analysis of the comment text. Applying an analysis of the emotional color of the text of comments by social network users on recommended objects can greatly improve the quality of prediction recommendations, and the transition to affine transformations and matrix calculations simplifies the machine complexity of calculations and increases their speed.

**recommendation systems, sentiment-analysis, collaborative filtration, machine learning, intelligent systems**

Одержано (Received) 08.05.2018