Healthcare Analytics 3 (2023) 100148

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

A recommender system for occupational hygiene services using natural language processing☆

Nicolás Barrera [a,*], Rodolfo Torres [a], Jhonathan Rodríguez [a,b], Oscar Espinosa [a,b], Salín Avellaneda [c], Jorge Ramírez [c]

[a] Directorate of Analytical, Economic and Actuarial Studies in Health, Instituto de Evaluación Tecnológica en Salud (IETS), Colombia
[b] Economic Models and Quantitative Methods Research Group, Centro de Investigaciones para el Desarrollo, Universidad Nacional de Colombia, Colombia
[c] Vice Presidency of Promotion and Prevention, Positiva Compañía de Seguros, Colombia

ARTICLE INFO

Keywords:
Recommender system
Insurance sector
Natural language processing
Collaborative filter

ABSTRACT

Although recommender systems have been used in a wide variety of products, they have not had the same adoption in the insurance sector due to the particular aspects of the industry. The present research proposes a novel model that uses natural language processing techniques and a deep learning-based model to recommend occupational hygiene services. The results of ranking methods showed that the model combining memorization and generalization of companies' activities preferences performed better compared to other models. These findings have significant implications for decision-making as they can improve the welfare of both companies and workers, while also limiting the options offered to clients based on their unique profiles.

1. Introduction

Recommender systems (RSs) are tools that can suggest different types of items, such as movies, books, cars, and tweets, among others, to users of a platform to satisfy their requirements without the need to search through endless lists of products [1]. Machine learning techniques that exploit the characteristics of the items, the user, or the community are used for this purpose [2–6].

However, despite their widespread popularity in many other industries [7–9], RSs have yet to be adopted for insurance applications due to several challenges specific to this industry. These challenges include the limited range of products and customers, the need to continually update products according to regulatory changes, and a lack of sufficient data for applying modern machine learning techniques [10,11]. In addition, the unique way in which insurance products are offered based on a customer's risk assessment has also posed a challenge to the widespread adoption of RSs in the insurance industry, compared to other industries [12].

Occupational hygiene services are affected by all the problems mentioned above regarding the insurance industry. In addition, they must fulfill the dual purpose of reducing hazards and retaining the loyalty of affiliated companies. In light of these challenges, we propose different collaborative filters based on items, each founded on the purchase history between the different companies and an insurance broker with expertise in the field of occupational hygiene. Collaborative filtering is a well-known technique in the field of recommender systems that are based on the preferences and behaviors of users [13].

By leveraging the information available in the purchase history, the proposed system aims to offer relevant recommendations to customers in the occupational hygiene field. The performance of the algorithms is evaluated using a real-world dataset of occupational hygiene activities, and the results are compared to identify the best-performing algorithm.

To further enhance the accuracy of the recommendations, natural language processing techniques are used to standardize semantically similar activities and homogenize the customer-implemented activities using the Word Mover's Distance (WMD) algorithm [14].

The remainder of this paper is organized as follows. Section 2 is devoted to a literature review of recommender systems; Section 3 describes the methodology; Section 4 presents the data and the results of the application; and in Section 5, we discuss the implications of the research.

2. Literature review

Multiple methodologies have been used for the construction of RS such as decision trees, neighborhood-based algorithms, associative classification, and neural networks, among others [15]. In general, the selection of the approach is based on the economic sector to which it will be applied, and the availability of information with different levels of disaggregation. However, it is rare to find documents that use

☆ Funding: This work was supported by Positiva Compañía de Seguros [contract 0155 of 2022].
* Correspondence to: Carrera 49a # 91 – 91, Bogotá, Colombia.
E-mail address: nibarrerabo@unal.edu.co (N. Barrera).

https://doi.org/10.1016/j.health.2023.100148
Received 6 September 2022; Received in revised form 14 February 2023; Accepted 15 February 2023

evidence from the real world, on the contrary, it is common to find articles and conference presentations in which an algorithm is tested on a set of experimental data, that is, data obtained from simulations.

All the studies described below have applied various measures of the model quality, with accuracy, F1-score, precision, and root mean square error (RMSE) being the most commonly used. Kamlor and Cosh [2], Chaurasiya and Sahu [3], Duy and Dinh [4] Yu et al. [5], and Zeng et al. [16] developed RSs based on collaborative filters (CFs) for the e-commerce sector and achieved an accuracy of 89.4%, an RMSE of 0.64, a precision of 0.03 and a mean absolute error (MAE) of 0.74, respectively.

In each of these studies, the authors used real-world data. In the first, information from a video game company was used, and the results were cross-checked by surveying regular buyers of the company. In the second, the authors used the user ID, product ID, and rating and review history of each Amazon user, whereas in the third, a database available from the Kaggle platform was used, which contained records for each user, the time spent and the interactions with the website, the transactions made, and the ratings assigned to products previously purchased. In the fourth, data on the type of action performed (searches, views, pre-purchase, and purchase actions), the number of users, and the products purchased by each user from a company were used.

Another sector in which RSs have been widely applied is entertainment, and especially movies. Xia [6] used the MovieLens database, which contains the satisfaction rating assigned by each user for each movie, and applied a CF, obtaining an MAE of 0.62. Using the same database, Ye et al. [17] used a neural network scheme and achieved an MAE of 0.9. Roy et al. [18] applied a content-based filter to the Movie Dataset database, which contains the same variables mentioned above, and obtained a MAE of 0.34 and a RMSE of 0.49, while Zhou [19] used the FilmTrust database and achieved an RMSE of 0.71 by combining a CF with deep and compressed neural networks.

The RS developed by Lian and Li [20] was based on transformer neural networks, and an RMSE of 0.79 was achieved. Although the authors indicated that this approach was used to identify products that could be offered through a specific financial institution, they did not mention the company or the data on which training and testing of the algorithm were performed. Turkut et al. [21] used experimental data from the textile sector, and sought to recommend products using convolutional neural networks, using a description and/or photo provided by the user; with this approach, the authors obtained an F-1 score of 82.3% and an accuracy of 82.08%.

Finally, we note the boom in the application of RSs to the insurance sector. However, these developments form part of the marketing strategies of companies, and none of the studies detailed below precisely describes the sources of these data, despite indicating that they involved data collected from insurance companies. Liu et al. [22] used the sociodemographic characteristics of policyholders, such as their gender, age, occupation, and location, together with the history of the products purchased and the rating given to them, and applied a CF using cosine as a similarity measure, with which they obtained a MAE of 0.7. Zhou et al. [23] applied their RSs (a hybrid between a CF and a user portrait profiling algorithm) to non-automotive insurance, using data on the history of products purchased, the sociodemographic characteristics of the insured (such as gender, age, and marital status), and browsing information from the company's website, and obtained an accuracy of 78%.

In this same sector, RSs have been implemented that are specifically based on supervised machine learning algorithms, in which the products offered are those that are highly likely to be purchased according to the company's purchase history. Vij and Preethi [24] used sociodemographic information on policyholders and the history of products purchased in a model based on the random forest and XGBoost algorithms and found that the latter gave the best accuracy of 72.26%. Using similar information to that mentioned above, Guo et al. [25] implemented the naive Bayes, ID3, C4.5, nearest neighbors, and random forest methods. They reported that the last gave the best approximation since its error rate was the lowest at 16%. Li [26] applied the logit, CART, ID3, Adaboosting, random forest, nearest neighbors, and support vector machine methods with polynomial kernel algorithms, and found that the one with the best performance was the last of these, with an area under the curve (AUC) of 0.64.

Specifically, regarding the occupational risk insurance industry (or workers' compensation), no studies could be found in the literature, and we, therefore, consider that this technical document represents the first scientific approach to address this issue.

3. Theoretical model structure

The activities in occupational hygiene are constantly changing, with updates, additions, or deletions, due to sector regulations, constant updates in line with the rules that govern the insurance industry, or internal policies of the insurance company. Nevertheless, some activities retain the core of their purpose. This affects the traceability and record-keeping of activities performed for different companies and hampers the recommendation of future activities.

For example, consider activities designed to eliminate or reduce mining workers' exposure to hot thermal stress. In 2019, one activity was described as "Identifying workers who are acclimatized or assessed as fit to work in hot conditions". In 2020, this activity was eliminated, and in 2021, a new activity was created: "Defining a safe environment (in terms of temperature and humidity) and specifying the length of time a worker can work in a hot environment". These changes in occupational hygiene activities demonstrate how they are constantly evolving to keep up with industry regulations and to ensure the health and safety of workers. However, not all occupational hygiene activities are simply evolutions of others, as demonstrated by the activity "Provide cool water and encourage workers to drink every 25–30 min", which focuses on a different aspect of controlling worker exposure to hot thermal stress.

Creating a list of core occupational hygiene activities (standardized activities/item list) allows for the homogenization of similar activities across different years. This list helps to ensure that the program remains consistent over time, even as regulations and internal policies evolve. Homogenizing similar activities across different years allows for improved traceability and record-keeping, which is essential for constructing a recommender system. The construction of standardized activities list is explained in Section 3.1.

Based on the standardized activity list, a pipeline is constructed to create a recommendation system for occupational hygiene activities. This pipeline (Fig. 1) standardizes the historical activities acquired by the customer using the standardized activity list. These standardized activities form the inputs to the proposed RS, which outputs a series of activities. These activities are then de-standardized (converted from the standardized form into their equivalent in the current list of activities offered) and finally suggested to the customer.

To perform the standardization and de-standardization, activity descriptions are transformed into vectors using the methodology described in the next section. From the vector representation, we calculate the WMD between the activity to be standardized or de-standardized and the different activities on the reference list, which could be either the standardized activity list or the current list of activities offered by the company, and choose the one with the smallest calculated distance.

3.1. Standardized activities list

To standardize the list of all activities offered by the insurance company over time while preserving their essence without losing their purpose due to the aforementioned changes, we transform the description of each activity into a mathematical object. This allows us to group the activities based on their similarity and standardize the list.

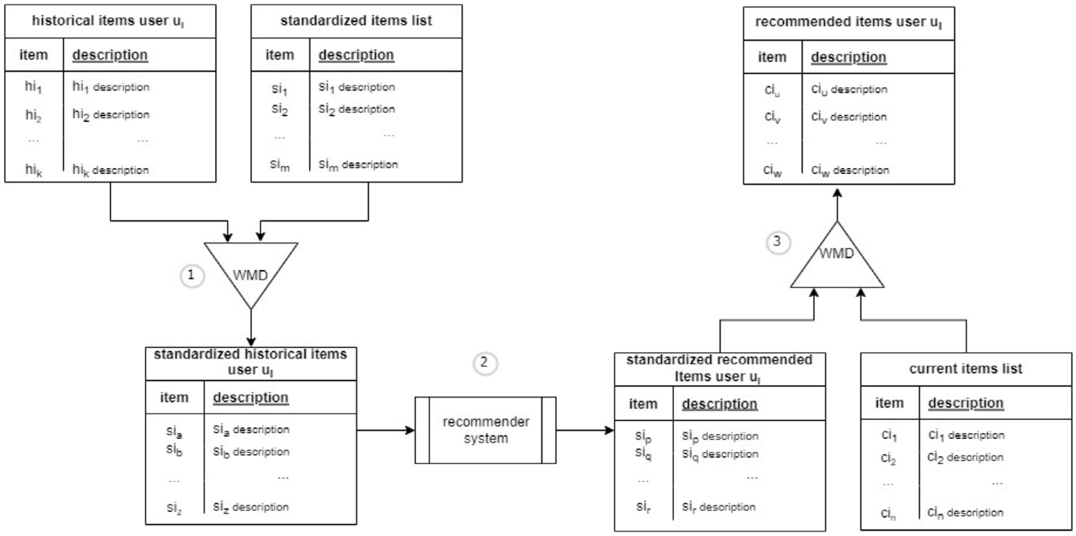

Fig. 1. Model system pipeline: (1) We use the WMD algorithm to standardize the customer-implemented historical activities, ensuring consistency and homogeneity in the activities over time. (2) We then use the proposed RS algorithm to recommend activities. (3) Finally, we use the WMD algorithm to de-standardize the recommended standardized activities back into their original form, as offered by the insurance company.

Based on the descriptions of the occupational hygiene activities, we perform embedding using FastText [27–32] to preserve its syntactic and semantic content. We then use the smooth inverse frequency (SIF) [33] to assign a weight to each of these vectors according to its relevance in the sentence. Finally, we obtain a high-dimensional vector associated with each of the activities descriptions, from which it is possible to determine the similarity between a pair of activities using different metrics.

A standardized list of activities is obtained by calculating the cosine similarity between different pairs of descriptions for all activities offered, merging those that exceed a similarity threshold, and collecting only their characteristic words. In this way, a list is obtained that condenses the core characteristics of the different activities over time.

3.2. Recommender system

In recommender systems, feedback can be either explicit or implicit. Explicit feedback refers to the user's direct feedback on a product or service, such as rating a movie or leaving a review. This type of feedback is usually collected through surveys, questionnaires, or by directly asking the user to rate a product. On the other hand, implicit feedback refers to indirect feedback gathered from the user's behavior and actions, such as how long they spend on a product page or if they added it to their shopping cart [34].

In the context of occupational hygiene activities, it is not possible to obtain explicit feedback from users in the form of satisfaction ratings. This is because it is not possible to observe an immediate change in the accident rate or safety conditions when the activity is implemented. Thus, traditional explicit feedback mechanisms, where users rate the products they have received, cannot be used to train recommender systems. Instead, alternative methods, such as implicit feedback, must be used.

Implicit feedback is based on the observed behavior of the users and the items they interact with. This can include, for example, the frequency of use, the duration of use, or the purchase history of the items [35]. By analyzing this implicit feedback, recommender systems can learn about the preferences of the users and make recommendations accordingly. In this study, we propose to use implicit feedback based on the purchase history between the companies and an insurance broker with expertise in the field of occupational hygiene to develop a collaborative filtering recommender system [1].

The recommender systems predict items that a user may have an interest in, these predictions in collaborative filters (CF) are based on finding similar users or items and using their preferences to make recommendations. There are two main approaches to generating recommendations in CF: memory-based CF and model-based CF [36]. Memory-based CF uses the entire dataset to identify similar users or items, while model-based CF creates a model based on learning from the data.

In this section, we propose different models that implement memory-based and model-based approaches [37]. In the next section, we compare their performance using a real-world dataset of occupational hygiene activities.

3.2.1. Item-based collaborative filter

Initially, we implement an Item-Based CF model, which is based on the concept of item–item similarity. Given a user, the model will recommend items that are similar to the ones the user has shown interest in the past. The item–item similarity can be represented mathematically using various similarity measures such as cosine similarity, Pearson correlation, and Jaccard similarity [38].

An item-based collaborative filtering approach can be used to recommend activities to companies based on their past purchases. This approach compares the similarity between different activities offered by the insurance broker and makes recommendations based on the activities that other similar companies have acquired in the past.

For instance, if a company has previously purchased activities related to ventilation and respiratory protection, the item-based collaborative filter may recommend activities related to ergonomics and hearing protection, if these activities have been commonly acquired together by similar companies. This approach considers the past behavior of companies and their preferences to make personalized recommendations.

The model we use is a variation of the Item-Based Collaborative Filtering model via Item-Variance Weighting (IVW) [39,40], in the algorithm, the item–item similarity score is weighted by the variance of ratings of each item. This weighting helps to address the issue of items with many ratings being preferred over those with fewer ratings. The final recommendation is then generated by taking the weighted sum of all the similar items' ratings.

In the context of occupational hygiene activities, the purchase history of these activities can then be used to calculate the item–item similarity and generate recommendations by equating the weight of the number of ratings per user with the number of activity purchases per company (Fig. 2).

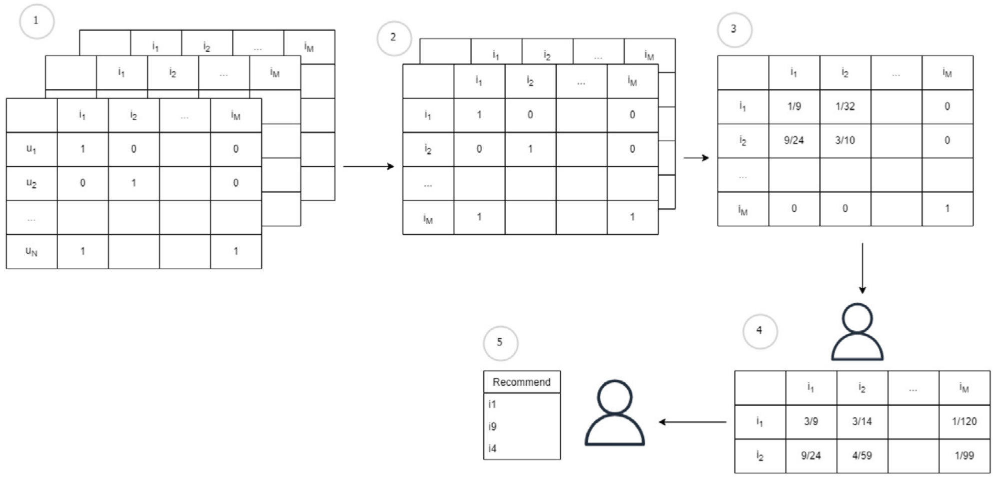

Fig. 2. IVW model-variation steps (i: activities, u: companies): (1) Collect historical data of occupational hygiene activities purchased by different companies. (2) Create an activity–activity (item–item) matrix of historical purchases; (3) Calculate the activity–activity (item–item) similarity based on the historical purchases data. (4) Use the IVW algorithm to predict a company (user's) correlation with all activities (columns) based on his historical (rows). (5) Recommend the activities with the highest predicted scores to the company.

3.2.2. Deep learning-based model

The next model we implement is an Extended Neural Matrix Factorization (Extended-NeuMF) algorithm [41–45], which is a deep learning-based recommendation system. This model consists of two components: a generalized matrix factorization (GMF) and a multilayer perceptron (MLP). These components are concatenated across the layers before their final output and serve as input to a neural network (Fig. 3).

The GMF component of the NeuMF algorithm is a wide neural network in which the input is the element-wise product of the user and item embeddings. The target label of this network is a binary indication of whether the item has been consumed. This setup captures linear relationships between users and items, and can be interpreted as memorizing the preferences of users.

The MLP component, on the other hand, utilizes a deep neural network that takes as input the concatenation of the user and item embeddings, along with the characteristics of both the users and items. The network is trained to predict whether an item has been consumed, based on the input features. This setup allows for the capture of rare or nonlinear relationships between users and items, and provides a more generalized representation of user preferences compared to the GMF component.

Embeddings in the model refer to the low-dimensional representation of high-dimensional data, such as companies and activities. The embeddings are learned by the model and capture underlying patterns and relationships in the data. The company embeddings capture individual companies' preferences, while the activity embeddings capture the features and characteristics of activities. These embeddings help to improve the performance of the model compared to traditional memory-based methods [46].

In the context of occupational hygiene activities, the NeuMF algorithm can be applied to the historical purchase data of different companies. The algorithm can learn the underlying representations of both companies and activities and make recommendations based on these representations. For example, if a company has shown interest in activities related to ergonomics and hearing protection, the NeuMF algorithm may recommend activities related to ventilation and respiratory protection based on the past behavior of similar companies.

4. Results

4.1. Experimental setup

We used to supply and demand data for industrial hygiene products from Positiva Compañía de Seguros, the Colombian state occupational risk insurer, which insures over 3 million workers. The dataset consisted of approximately 6000 companies, 500 activities, and 200,000 purchases. Each company in the dataset was associated with metadata including the number of employees, the economic sector of the company, and the company's income. The metadata for each activity included information about the composition of the industrial hygiene program (such as anticipation, recognition, evaluation, prevention, and control) and the related topic (such as chemical hazards and physical hazards).

To evaluate the model, 80% of the data were used for training and the remaining 20% for testing. When calculating this division, clients who acquired five or more activities were considered. In addition, to ensure the quality of the model, the most popular activities were not used in the test set to prevent overfitting and ensure that the model's performance is robust and generalizes well to new and unseen data.

4.2. Evaluation metrics

Ranking methods are commonly used as metrics for implicit recommendation systems, where it is only possible to determine the relevance of a recommended item to a user. The performance of these systems is often evaluated by considering the trade-off between the fraction of recommended items that are relevant to the user and the fraction of consumed items that are captured in the recommendations. This trade-off can be visualized using precision–recall and receiver operating characteristic (ROC) curves. [1].

Let Ω as the relevant elements or acquired by the user, $R(n)$ the top n of activities recommended to the user, and I the universe of all items. The precision in a top n number of recommendations is defined as the percentage of recommended items that truly turn out to be relevant:

$$Precision\,(n) = \frac{|\Omega \cap R\,(n)|}{|R\,(n)|}.$$

The recall or true positive rate *(TPR)* in a top n number of recommendations, is defined as the percentage of relevant items that have been recommended:

$$TPR\,(n) = Recall\,(n) = \frac{|\Omega \cap R\,(n)|}{|\Omega|}.$$

The false positive rate (FPR) in a top n number of recommendations, is defined as the percentage of the falsely reported positives in the recommended list:

$$FPR\,(n) = \frac{|R\,(n) - \Omega|}{|I - \Omega|}.$$

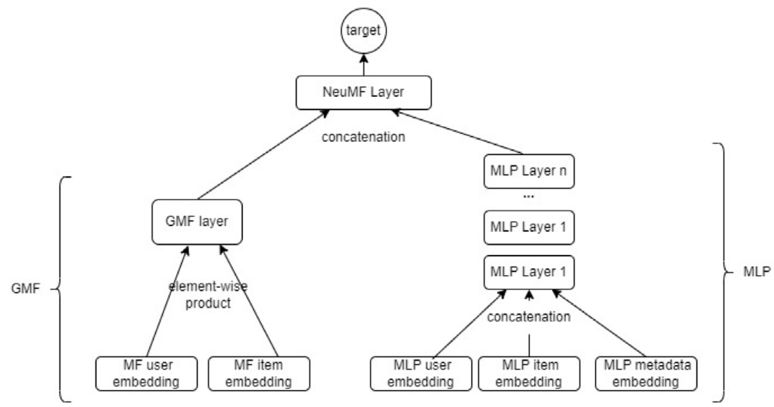

Fig. 3. Extended-NeuMF architecture: deep learning-based recommendation system that combines a Generalized Matrix Factorization (GMF) and a Multilayer Perceptron (MLP). The architecture of this model allows for different embeddings for both users and items in each component.

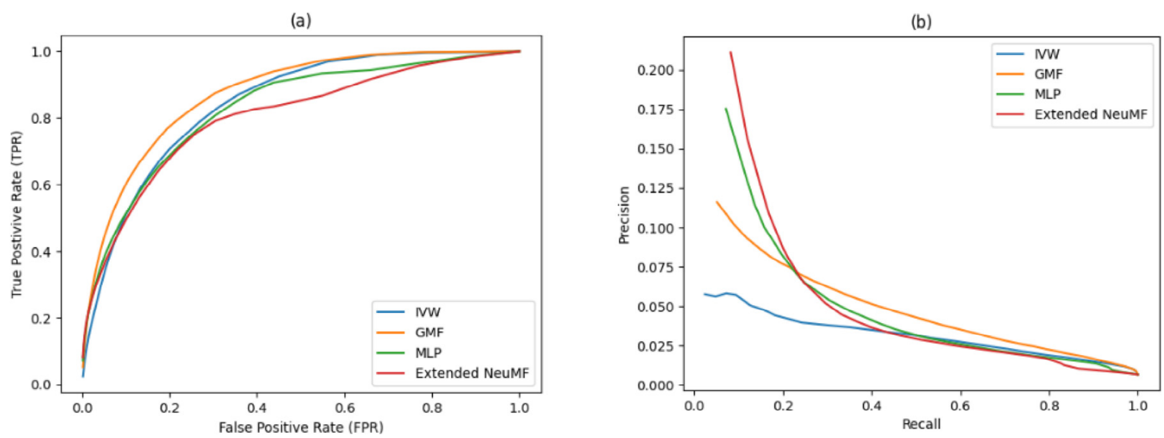

Fig. 4. (a) ROC curves and (b) precision–recall curves for different recommendation analytical models.

4.3. Application with real world evidence

We compare the IVW model, the extended NeuMF model, and each one of its components: GMF and MLP, using the ROC and precision–recall curves. From the ROC and precision–recall curves (Fig. 4), we observe different model dominances along the different parts of the curves. The NeuMF model dominated in the early part, indicating that the model ranked the most relevant activities highly, followed by its components and the IVW model. The GMF model dominated in the latter part of the curve, indicating that less relevant activities were ranked better.

Therefore, it is evident that model-based CF is superior to memory-based CF in predicting the top relevant acquisitions. This means that it is possible to generalize the preferences of companies to acquire occupational hygiene services based on the metadata of companies and activities. By combining these models, better prediction performance is obtained. Calculating the area under the ROC curve (AUC), the values for GMF, IVW, MLP, and the Extended NeuMF were found to be 0.87, 0.84, 0.82, and 0.80, respectively. These results are higher compared to what has been reported in the literature review for the insurance sector. However, it is important to note that this metric alone does not fully reflect the performance from the user's perspective, as the user may only be interested in the first few recommended items.

5. Discussion and conclusions

In this research, we have evaluated multiple models to recommend industrial hygiene services to different clients of an occupational risk insurer. From data on the historical offers, the relevant characteristics of the different products are extracted, merged, and consolidated, and a recommender system is built based on the collective expertise of the insurance brokers. We assume the activities are aimed at mitigating some risk and are executed jointly.

When we compare multiple models of recommender systems, we look that better performance is obtained with the model than combine memorization and generalization of companies' items preferences. It is also observed a better value on AUC of ROC curves than other values reported in the insurance industry. From this, it follows that it is possible to take advantage of the capabilities that RS provides in other industries for use in occupational hygiene services.

This RS does not achieve precision values that would allow the insurance broker to be replaced; however, it serves as a tool by limiting the activities that should be reviewed to be recommended to the company, and also opens the way for an exploration of other collaborative models or systems based on information that takes into account the intrinsic characteristics of the company.

Finally, in the business field, the development of analytical processes that manage to increasingly improve customer acquisition and loyalty is generating high-added value in commercial planning and prospective sales. The possible creation of induced demand, with the support of computational tools such as recommender systems, is an essential factor from the strategic context, in which companies at the forefront of technology are making constant developments. The methodological approach presented here contributes to the knowledge frontier of the design and application of recommender systems with evidence from the real world, making use of natural language processing and deep learning techniques.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] C. Aggarwal, Recommender Systems, Springer International Publishing, 2016, http://dx.doi.org/10.1007/978-3-319-29659-3.

[2] W. Kamlor, K. Cosh, Product discovery via recommendation based on user comments, in: Proceedings of the 2014 6th International Conference on Knowledge and Smart Technology, KST 2014, 2014, pp. 41–45, http://dx.doi.org/10.1109/kst.2014.6775391.

[3] R. Chaurasiya, U. Sahu, Improving performance of product recommendations using user reviews, in: 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, ICRAIE 2018, 2018, http://dx.doi.org/10.1109/ICRAIE.2018.8710414.

[4] P. Duy, L. Dinh, E-commerce recommendation system using mahout, in: 2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS, 2019, pp. 86–90.

[5] J. Yu, J. Shi, Y. Chen, D. Ji, W. Liu, Z. Xie, et al., Collaborative filtering recommendation with fluctuations of user preference, in: 2021 IEEE International Conference on Information Communication and Software Engineering, ICICSE, 2021, pp. 222–226.

[6] J. Xia, E-commerce product recommendation method based on collaborative filtering technology, in: 2016 International Conference on Smart Grid and Electrical Automation, ICSGEA, 2016, pp. 90–93, http://dx.doi.org/10.1109/ICSGEA.2016.81.

[7] S. Baby, S.L. Reddy, End to end product recommendation system with improvements in apriori algorithm, in: Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021, 2021, pp. 1357–1361, http://dx.doi.org/10.1109/ICIRCA51532.2021.9544981.

[8] T. Oswari, T. Yusnitasari, R.D. Kusumawati, S. Mittal, Music recommendation system design for online platform with the concept of unified modeling languange (UML), in: 2018 International Conference on Sustainable Energy, Electronics and CoMputing System, SEEMS 2018, 2019, http://dx.doi.org/10.1109/SEEMS.2018.8687346.

[9] N. Iswari, Rusli A. Wella, Product recommendation for e-commerce system based on ontology, in: 2019 1st International Conference on Cybernetics and Intelligent System, ICORIS 2019, 2019, pp. 105–109, http://dx.doi.org/10.1109/ICORIS.2019.8874916.

[10] W. Xu, J. Wang, Z. Zhao, C. Sun, J. Ma, A novel intelligence recommendation model for insurance products with consumer segmentation, J. Syst. Sci. Inform. 2 (2017) 16–28, http://dx.doi.org/10.1515/JSSI-2014-0016.

[11] T. Kanchinadam, M. Qazi, J. Bockhorst, M.Y. Morell, K. Meissner, G. Fung, Using discriminative graphical models for insurance recommender systems, in: Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, 2019, pp. 421–428, http://dx.doi.org/10.1109/ICMLA.2018.00069.

[12] S. Mitra, N. Chaudhari, B. Patwardhan, Leveraging hybrid recommendation system in insurance domain, Int. J. Eng. Comput. Sci. 3 (2014) 8988–8992.

[13] F. Isinkaye, Y. Folajimi, B. Ojokoh, Recommendation systems: Principles, methods and evaluation, Egypt. Inform. J. 16 (2015) 261–273, http://dx.doi.org/10.1016/J.EIJ.2015.06.005.

[14] M. Kusner, Y. Sun, N. Kolkin, K. Weinberger, From word embeddings to document distances, in: 32nd International Conference on Machine Learning, Vol. 2, ICML 2015, 2015, pp. 957–966.

[15] I. Portugal, P. Alencar, D. Cowan, The use of machine learning algorithms in recommender systems: A systematic review, Expert Syst. Appl. 97 (2018) 205–227, http://dx.doi.org/10.1016/j.eswa.2017.12.020.

[16] Y. Zeng, Y. Bi, J. Wang, Y. Lin, Collaborative filtering recommendation algorithm optimization based on user attributes, in: Proceedings - 2015 8th International Symposium on Computational Intelligence and Design, Vol. 1, ISCID 2015, 2016, pp. 580–583, http://dx.doi.org/10.1109/ISCID.2015.91.

[17] Z. Ye, S. Shu, F. Lin, Dynamic identification for personalized product recommendations based on fuzzy discrete event systems, in: Proceedings of the 2012 24th Chinese Control and Decision Conference, CCDC 2012, 2012, pp. 3110–3116, http://dx.doi.org/10.1109/CCDC.2012.6243078.

[18] S. Roy, M. Sharma, S. Kumar, Movie recommendation system using semi-supervised learning, in: 2019 Global Conference for Advancement in Technology, GCAT, 2019, pp. 1–5.

[19] Y. Zhou, A dynamically adding information recommendation system based on deep neural networks, in: Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Information Systems, ICAIIS 2020, 2020, pp. 1–4, http://dx.doi.org/10.1109/ICAIIS49377.2020.9194792.

[20] M. Lian, J. Li, Financial product recommendation system based on transformer, in: Proceedings of 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2020, 2020, pp. 2547–2551, http://dx.doi.org/10.1109/ITNEC48623.2020.9084812.

[21] U. Turkut, A. Tuncer, H. Savran, S. Yilmaz, An online recommendation system using deep learning for textile products, in: HORA 2020-2nd International Congress on Human-Computer Interaction, in: Optimization and Robotic Applications, Proceedings, 2020, http://dx.doi.org/10.1109/HORA49412.2020.9152875.

[22] Z. Liu, C. Zang, K. Kuang, H. Zou, H. Zheng, P. Cui, Causation-driven visualizations for insurance recommendation, in: 2019 IEEE International Conference on Multimedia & Expo Workshops, ICMEW, 2019, pp. 471–476.

[23] J. Zhou, Y. Guo, Y. Ye, J. Jiang, Multi-label entropy based feature selection with applications to insurance purchase prediction, in: Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2020, 2020, pp. 427–432, http://dx.doi.org/10.1109/ICAICA50127.2020.9181921.

[24] A. Vij, N. Preethi, Approaches towards a recommendation engine for life insurance products, in: 2021 IEEE International Conference on Mobile Networks and Wireless Communications, ICMNWC 2021, 2021, http://dx.doi.org/10.1109/ICMNWC52512.2021.9688436.

[25] Y. Guo, Y. Zhou, X. Hu, W. Cheng, Research on recommendation of insurance products based on random forest, in: Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2019, 2019, pp. 308–311, http://dx.doi.org/10.1109/MLBDBI48998.2019.00069.

[26] H. Li, Research on mobile caravan insurance recommendation method based on machine learning, in: Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021, 2021, pp. 9–15, http://dx.doi.org/10.1109/CAIBDA53561.2021.00010.

[27] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, in: LREC 2018-11th International Conference on Language Resources and Evaluation, 2019, pp. 3483–3487.

[28] V. Venekoski, J. Vankka, Finnish resources for evaluating language model semantics, in: NoDaLiDa 2017-21st Nordic Conference of Computational Linguistics, Proceedings of the Conference, 2017, pp. 231–236.

[29] T. Shaik, X. Tao, Y. Li, C. Dann, J. McDonald, P. Redmond, et al., A review of the trends and challenges in adopting natural language processing methods for education feedback analysis, IEEE Access 10 (2022) 56720–56739, http://dx.doi.org/10.1109/ACCESS.2022.3177752.

[30] M. Wadud, M. Mridha, M. Rahman, Word embedding methods for word representation in deep learning for natural language processing, Iraqi J. Sci. 63 (2022) 1349–1361, http://dx.doi.org/10.24996/ijs.2022.63.3.37.

[31] M. Umer, Z. Imtiaz, M. Ahmad, M. Nappi, C. Medaglia, G.S. Choi, et al., Impact of convolutional neural network and FastText embedding on text classification, Multimed. Tools Appl. (2022) 1–17, http://dx.doi.org/10.1007/s11042-022-13459-x.

[32] C. Du, Z. Chen, F. Feng, L. Zhu, T. Gan, L. Nie, Explicit interaction model towards text classification, in: 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, Vol. 33, EAAI 2019, AAAI Press, 2019, pp. 6359–6366, http://dx.doi.org/10.1609/aaai.v33i01.33016359.

[33] S. Arora, Y. Liang, T. Ma, A simple but tough-to-beat baseline for sentence embeddings, 5th International Conference on Learning Representations, ICLR 2017, https://openreview.net/pdf?id=SyK00v5xx.

[34] H. Chu, Y. Shen, User Feedback Design in AI-Driven Mood Tracker Mobile Apps, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2022, http://dx.doi.org/10.1007/978-3-031-05412-9_24/COVER, 13304 LNCS:346–58.

[35] G. Jawaheer, M. Szomszor, P. Kostkova, Comparison of implicit and explicit feedback from an online music recommendation service, in: Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, RecSys 2010, in: HetRec 2010, Held at the 4th ACM Conference on Recommender Systems, 2010, pp. 47–51, http://dx.doi.org/10.1145/1869446.1869453.

[36] P. Aditya, I. Budi, Q. Munajat, A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study PT X, in: 2016 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2016, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 303–308, http://dx.doi.org/10.1109/ICACSIS.2016.7872755.

[37] H. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, et al., Wide & deep learning for recommender systems, in: ACM International Conference Proceeding Series, 2016, pp. 7–10, http://dx.doi.org/10.1145/2988450.2988454.

[38] F. Fkih, Similarity measures for collaborative filtering-based recommender systems: Review and experimental comparison, J. King Saud Univ. - Comput. Inform. Sci. 34 (2022) 7645–7669, http://dx.doi.org/10.1016/J.JKSUCI.2021.09.014.

[39] Z. Zhang, Y. Kudo, T. Murai, Y. Ren, Enhancing recommendation accuracy of item-based collaborative filtering via item-variance weighting, Appl. Sci. (Switzerland) 9 (2019) 1928, http://dx.doi.org/10.3390/app9091928.

[40] L. Baltrunas, F. Ricci, Item weighting techniques for collaborative filtering, Stud. Comput. Intell. 220 (2009) 109–126, http://dx.doi.org/10.1007/978-3-642-01891-6_7.

[41] M. Unger, A. Tuzhilin, A. Livne, Context-aware recommendations based on deep learning frameworks, ACM Trans. Manag. Inf. Syst. (2020) 11, http://dx.doi.org/10.1145/3386243.

[42] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T. Chua, Neural collaborative filtering, in: 26th International World Wide Web Conference, WWW 2017, 2017, pp. 173–182, http://dx.doi.org/10.1145/3038912.3052569.

[43] H. Liu, W. Wang, Y. Zhang, R. Gu, Y. Hao, Neural matrix factorization recommendation for user preference prediction based on explicit and implicit feedback, Comput. Intell. Neurosci. (2022) 2022, http://dx.doi.org/10.1155/2022/9593957.

[44] C. Chen, M. Zhang, Y. Zhang, Y. Liu, S. Ma, Efficient neural matrix factorization without sampling for recommendation, ACM Trans. Inf. Syst. (2020) 38, http://dx.doi.org/10.1145/3373807.

[45] X. Sun, L. Gong, Z. Han, P. Zhao, J. Yu, S. Wang, Neural metric factorization for recommendation, Mathematics 10 (2022) 503, http://dx.doi.org/10.3390/math10030503.

[46] W. Jia, M. Sun, J. Lian, S. Hou, Feature dimensionality reduction: A review, Complex Intell. Syst. (2022) 2663–2693, http://dx.doi.org/10.1007/S40747-021-00637-X, 8:3 8 (2022).