Retraction

Retraction: Truth discovery in healthcare prediction using machine learning (J. Phys.: Conf. Ser. 1916 012068)

Published 23 February 2022

This article (and all articles in the proceedings volume relating to the same conference) has been retracted by IOP Publishing following an extensive investigation in line with the COPE guidelines. This investigation has uncovered evidence of systematic manipulation of the publication process and considerable citation manipulation.

IOP Publishing respectfully requests that readers consider all work within this volume potentially unreliable, as the volume has not been through a credible peer review process.

IOP Publishing regrets that our usual quality checks did not identify these issues before publication, and have since put additional measures in place to try to prevent these issues from reoccurring. IOP Publishing wishes to credit anonymous whistleblowers and the Problematic Paper Screener [1] for bringing some of the above issues to our attention, prompting us to investigate further.

[1] Cabanac G, Labbé C and Magazinov A 2021 arXiv:2107.06751v1

Retraction published: 23 February 2022
Truth discovery in healthcare prediction using machine learning

G Priyanka¹, M Mahalakshmi², G Nivedha² and J Roshini³
¹Assistant Professor, Department of Computer Science, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India.
²Student, Department of Computer Science, Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India.
nivedhaaag@gmail.com

Abstract. Healthcare is one of the core areas of the medical field in today's digital world. By analyzing large amounts of patient data, healthcare systems are needed to gain insights and help disease prediction. The truth discovery of the Health Recommendation System (TDHRS) is becoming an important platform for medical services. The proposed use of Boltzmann Machine (RBM) - FP-Growth Health Pattern Recognition (RBMG) smart HRS can provide insights into how big data analysis can be used to implement an effective health recommendation engine, and will be used in Transition from a standard solution to a more personalized paradigm in the telemedicine environment. By considering the Root Mean Square Error (RSME) and Mean Absolute Error (MAE) values, the proposed deep learning method (RBM-FP-Growth) exhibits fewer errors than other methods.

1. Introduction
In this contemporary computerized world, search engines are designed to carry out web search. For a particular information specified in a web search are meant to show the world wide information in a systematic manner. These are multi-sourced information, which creates conflicts in the data instead of showing precise information. These multi-sourced data also create conflicts in the health care system where patients try to diagnose their disease through their symptoms. In the existing system, HRS consists of different stages, through which specific projects can be recommended. These stages are the training stage, patient file processing stage, sentiment analysis stage, privacy protection stage and recommender stage.

Hospitals use surveys to assess patient satisfaction and improve general hospital operations. Through the use of statistical graphs and charts, the collected satisfaction data can be visually presented to the hospital management department. In order to determine the important factors that contribute to patient satisfaction, no more in-depth data analysis was performed. Once the user approves the collection of usage data in the implicit model, the user does not need any other explicit feedback (for example, scoring). This work searched for algorithms specifically suited to handle implicit feedback.

2. Related work
[1] researched an optimized model for the management and redesign of patient groups in primary care. [2] introduced health recommendation systems. [3] A clustering method is proposed to effectively
incorporate multi-criteria scoring into the standard recommendation system. [4] A probabilistic matrix factorization (PMF) model is proposed, which varies linearly with observations, and a constrained version of the PMF model is introduced, which is supported by the assumption that users who rate similar movie sets may have similar preferences.

[5] proposed a system where modern consumers are inundated with choices. [6] attempted to present an overview of the role of different MF models to address the challenges of CF algorithms, which can serve as a roadmap for research in this area. [7] Investigated the most common situations through implicit feedback.

3. Methodology

3.1. Architecture diagram

The dataset has been collected and stored in a repository. The Restricted Boltzmann Machine (RBM) algorithm has been implemented. The most accurate disease is predicted using truth discovery. The architecture diagram of this project is given below as figure 1:

![Architecture diagram](image)

**Figure 1. Architecture diagram.**

3.2. Dataset

The following is a disease-symptom association database generated from information supported by automatic methods in the text discharge summary of patients admitted to the New York Presbyterian Hospital in 2004. The first column shows the disease, the second column shows the number of carryout summaries from the hospital containing a definite and present mention of the disease, and the associated symptom. Table 1 shows the attributes diseases and their symptoms.

| Disease       | Symptom    |
|---------------|------------|
| Decubitus ulcer | systolic murmur |
|               | frail      |
|               | fever      |

3.3. Data preprocessing

The data being used in the Real-world lacks attribute values, these data contain errors and variation in code and names. Data preprocessing mostly refers to the technique of preparing (cleaning and
organizing) the information to make it suitable for building and training Machine Learning models.

3.4. Restricted boltzmann (RBM) machine
This paper proposes a model based on RBM. RBM is called a random neural network. The layers of the Restricted Boltzmann Machine are given in figure 2.

![Figure 2. Layers of Restricted Boltzmann Machine.](image)

Hidden bias and visual bias are two other layers of bias units in RBM. The hidden bias RBM generates activation on the antenna, so the visible bias can help the RBM reconstruct the input during the backward pass. Figure 3 shows the Activation function of Restricted Boltzmann Machine (RBM). Figure 4 shows the Pseudo code of CD algorithm.

![Figure 3. Activation function of Restricted Boltzmann Machine (RBM).](image)

The two main training steps are as follows:

- **Gibbs Sampling**

  Input is vector \( v \), use \( p(h | v) \) to predict the hidden value \( h \).
Contrastive Divergence step

The weighting matrix is updated in this step.

\[ p(v_t = 1 | h) = \frac{1}{1 + e^{-(a_i + w_i h_i)}} = \tilde{O} (a_i + \sum h_i w_i) \]

\[ p(h_t = 1 | v) = \frac{1}{1 + e^{-(b_j + w_j v_j)}} = \tilde{O} (b_j + \sum v_j w_j) \]

\[ \Delta W = v_0 \otimes P(h_0 | v_0) - v_j \otimes P(h_j | v_j) - v_j \]

From the updated matrix, the new weight is computed.

\[ W_{new} = W_{old} + \Delta W \]

**Figure 4.** Pseudo code of CD algorithm.

3.5. Steps of algorithm

3.5.1. Load symptoms calculation entropy. Calculate the predictability of a particular event. Use the frequency table of each attribute to calculate two types of entropy to build a decision tree as follows in equations (1) and (2).

\[ E(C, A) = [P(h) \ast E(C)] \]  

where C is the current state (existing effect), and P (h) the probability of event h state C.

\[ E(C) = hH - P(h)\log 2P(h) \]  

(1)
where \( C \) is the present condition with attribute \( A \), and \( A \) attribute is considered \( (h) \) the probability of event \( H \) of attribute \( A \). Table 2 shows the calculation of \( E(\cdot) \).

\[ E(\cdot) \text{ is the entropy of the whole set, and the second word} \ E(\cdot, A) \text{ corresponds to attribute } A. \]

3.5.3. **Probability-based information gain.** Regarding the symptoms, they are as follows in equations (3-7)

\[
IG(C, A) = E(\cdot) - E(\cdot, A)
\]

**Table 2. Calculation of \( E(\cdot) \).**

|       | Dengue | Malaria | Total |
|-------|--------|---------|-------|
| 8     | 4      | 12      |

Use formula (1)

\[
E(\cdot) = hH \cdot P(h) \log_2 P(h)
\]

3.5.4 **Find the root node.** The most valuable symbol of information is considered to be the root node. Starting from the symptoms of "high fever", count \( E(\cdot, \text{high fever}) \) and \( IG(\cdot, \text{high fever}) \):

\[
IG(C, \text{high fever}) = E(\cdot) - hH \cdot [P(C) \cdot E(C) - P(C) \cdot E(\cdot)]
\]

\[
IG(C, \text{high fever}) = [E(\cdot) - P(C) \cdot E(C) - P(C) \cdot E(\cdot)]
\]

where "\( h \)" is the possible value of the symptom. Here, the symptom "fever" is assumed to have two feasible results (presence (1) and absence (0)).

\[
E(\text{high fever}) = hH \cdot P(h) \log_2 P(h)
\]

Similarly, the information gain value is calculated in the following way without other symptoms:

\[
\begin{align*}
\text{IG (presence of high fever, vomiting)} & = 0.6518 \\
\text{IG (high fever, shaking)} & = 0.0772 \\
\text{IG (high fever, muscle wasting)} & = 0.0772
\end{align*}
\]

4. **Results and discussions**

Perform two-tier data analysis. The first step is to use the FP-Growth method to cluster patients formulated on their answers to the questions. In the second step, the clusters of the first layer are split into multiple sub-clusters based on patient demographic data. The correlation between the important features of each cluster and its sub-clusters can be obtained using the Restricted Boltzmann Machine method. Use standard statistical tests to rank and verify such associations.

Thus, the first layer contains a list of symptoms from which the user is asked to select the relevant symptoms, the user can also remove a particular selected symptom.

For example, the user selects the following symptoms (figure 5(a)).
- body pain
- loose motion
- weakness

After performing data analysis using the FP-Growth and RBM approach, the result of pattern matching is displayed (figure 5(b)). Further, the differential diagnosis (figure 5(c)) and visualization of differential diagnosis (figure 5(d)) is also displayed.

5. Conclusion
The model finds reliable solutions based on the general principles of truth discovery. Therefore, through the implementation of the Restricted Boltzmann Machine (RBM) algorithm, the truth is used to discover the most accurate disease prediction, and the output is visually represented using a pie chart and a table with diseases and their respective probability values. As for future work, a more
powerful optimization formula will be produced that can handle more complex situations.

References

[1] Bodenheimer T and Pham H H 2010 Primary care: Current problems and proposed solutions, Health Affairs, vol. 29, no. 5, pp. 799-805.

[2] Wiesner M and Pfeifer D 2014 Health recommender systems: Concepts, requirements, technical basics and challenges, Int. J. Environ. Res. Public Health, vol. 11, no. 3, pp. 2580-2607.

[3] Shi Y, Larson M, and Hanjalic A May. 2014 Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges, ACM Comput. Surv. vol. 47, pp. 3:1-3:45.

[4] A. Haldorai and A. Ramu, Security and channel noise management in cognitive radio networks, Computers & Electrical Engineering, vol. 87, p. 106784, Oct. 2020. doi:10.1016/j.compeleceng.2020.106784

[5] A. Haldorai and A. Ramu, Canonical Correlation Analysis Based Hyper Basis Feedforward Neural Network Classification for Urban Sustainability, Neural Processing Letters, Aug. 2020. doi:10.1007/s11063-020-10327-3

[6] Zhong J and Li X 2010 United collaborative filtering model based on a combination of latent features, Expert Syst. Appl., vol. 37, no. 8, pp. 5666-5672.

[7] Rendle S, Freudenthaler C, Gantner Z, and Schmidt-Thieme L 2009 BPR: Bayesian personalized ranking from implicit feedback, in Proc. 25th Conf. Uncertainty Artif. Intell. pp. 452-461.