Integrating Hybrid Recommendation Algorithms for Enhanced Personalization: A Case Study

Abstract

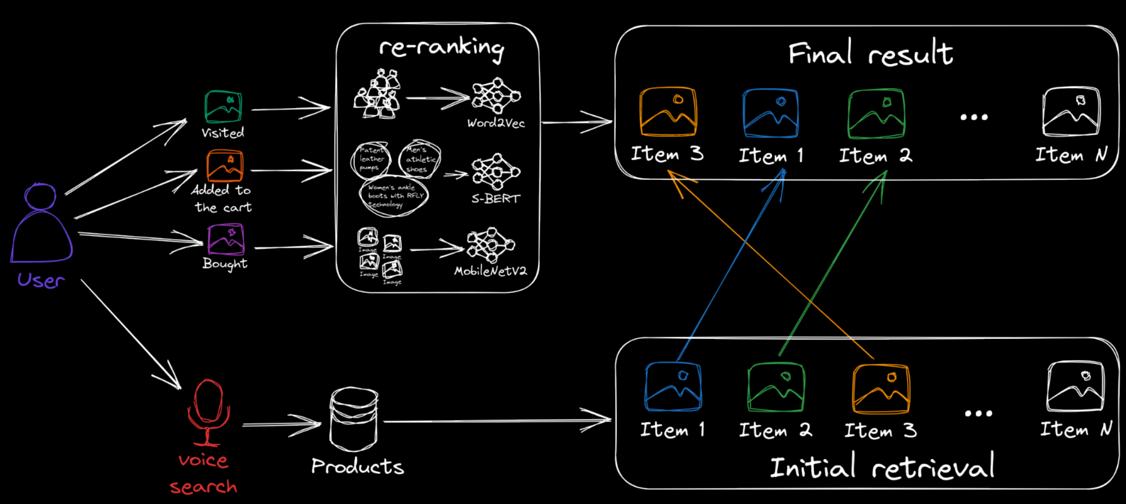
Effective recommendation systems are crucial for businesses aiming to offer personalized user experiences.
We introduce a hybrid recommendation algorithm merging content-based filtering with the BERT model and image embedding, and Word2Vec collaborative filtering. The Sentence-BERT model provides semantic embedding for textual content, producing recommendations based on semantic similarity between item descriptions. Our image-based algorithm identifies visual features from product images, recommending items visually similar to other products in which the customer was interested in. Using Word2Vec, the system taps into user journey data, promoting a collaborative recommendation approach, exposing users to new and relevant items.
We use Normalized Discounted Cumulative Gain (NDCG) to assess recommendation quality. As a recognized benchmark, NDCG evaluates our system's performance.
Our recommendation system is part of our Autonomous Voice Assistant project, suitable for e-commerce platforms. We demonstrate its utility with a case study from an online shoe store.

Methodology

Hybrid Recommendation Algorithm: Combines content-based and collaborative filtering for enhanced recommendations.
Content-based Filtering: Uses BERT for textual analysis of descriptions of products and image embeddings to match items with user preferences.

- **Sentence-BERT**: A BERT variant for deeper textual understanding beyond just keywords: the deeper meaning and context of the content.
- **Image-based Algorithm**: Extracts visual features from product images to align with user's visual preferences.

Word2Vec Collaborative Filtering: Leverages the collective intelligence of users. By analyzing user journey data, which includes items visited and interactions made, the system can make recommendations based on what similar users have liked or interacted with.



Practical Application

Case Study: an online store in the footwear industry - **rylko.com**. This showcases how our recommendation system can be integrated into e-commerce platforms to enhance user experience and drive sales.

Evaluation

We analyzed 249 sessions from **rylko.com**, encompassing 114 unique search queries. Using Natural Language Understanding, the **Autonomous Voice Assistant** AVA matched products to queries by extracting details like *product names*, *descriptions* and attributes, including *color*, *style*, *size*. These matches were then re-ordered using the Hybrid Recommendation Algorithm. An ecommerce expert assessed the accuracy of these results in the *top 5* window, both the original search and after re-ranking, assigning a rating between *1* and *5* for each position, indicating the appropriateness of the product match. Post re-ranking, some products shifted in or out of this *top 5*.

"I'm looking for women's sandals in size 37 for around 250 PLN."

Initial ranking: nDCG$_5$ = 95.01%

expert rating

Re-ranking: nDCG$_5$ = 100.00%

Normalized Discounted Cumulative Gain (*nDCG$_p$*): a metric used to evaluate the quality of a retrieval system. It measures how well the recommended items are ranked, considering the relevance of each item. The higher the *nDCG$_p$* score, the better the quality of the recommendation list. The parameter ***p*** in the *nDCG$_p$* represents the depth of evaluation, specifically the number of top results considered. In our context, *p* is set to *5*, meaning we assess the top *5* products. The *nDCG$_p$* value is normalized with ***Ideal DCG$_p$***.

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

DCG$_p$ represents the actual ranking of items, giving more weight to items at the top due to their perceived higher relevance. There are multiple formulations of the metric and we choose the most popular one. The rationale behind *DCG$_p$* is that highly relevant items appearing earlier in a ranking should be weighted more heavily than those lower down. The discounting factor, which is the logarithm of the rank position, ensures that relevant items appearing further down the list contribute less to the overall score.

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2(i+1)}$$

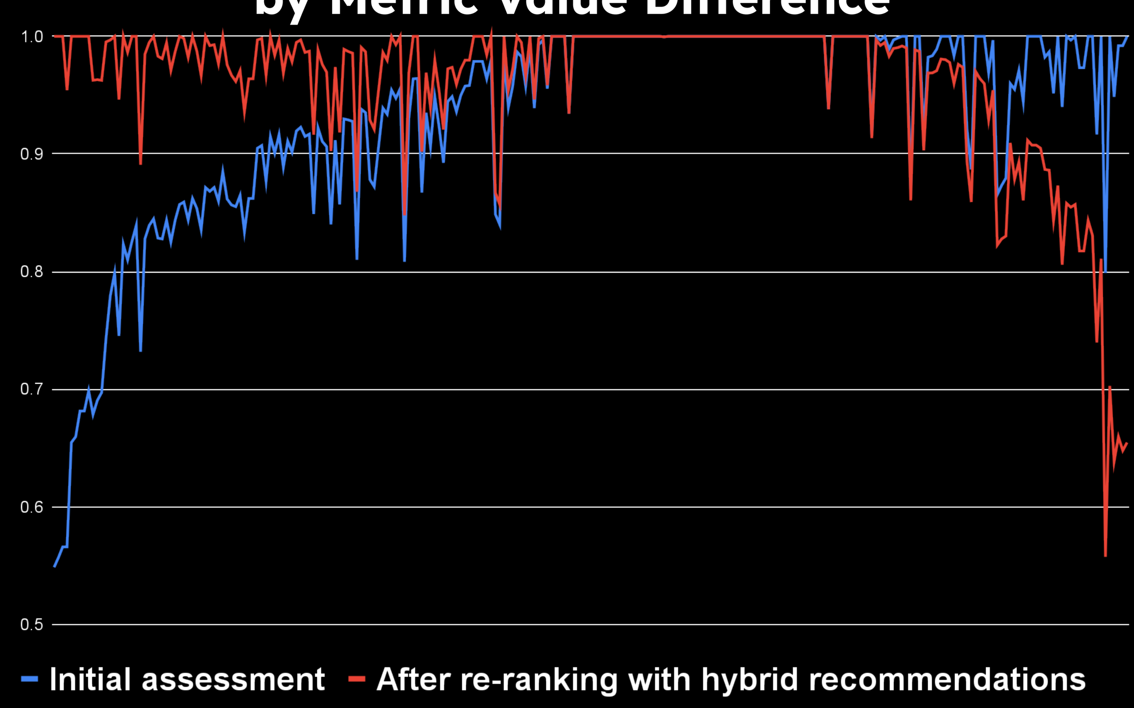
IDCG$_p$ depicts the ideal ranking scenario. We simulate the best possible ranking for the given set of products from retrieval: we take the relevance scores as in *DCG$_p$* but sort them in descending order, representing the ideal scenario where products are sorted from the highest to the lowest relevance.

i-th largest relevance score

$$IDCG_p = \sum_{i=1}^{p} \frac{2^{rel'_i} - 1}{\log_2(i+1)}$$

The baseline *nDCG$_p$* value for AVA search results stood at ***93.18%***. After re-ranking based on the recommendation algorithms, the average *nDCG$_p$* value increased to ***96.00%***. Notably, in ***46%*** of the cases, the *nDCG$_p$* value **improved** post-re-ranking, and in ***76%*** of instances, the metric was **at least as good as the baseline**. The chart illustrates the impact of applying recommendation algorithms to the search results from AVA.

***nDCG$_p$* Improvement Analysis: Sorting Cases by Metric Value Difference**



Re-ranking

$$r_k = ES * \prod_i \begin{cases} d(k, rec_i\{s\}) * B & \text{if } k \in rec_i\{s\} \\ P & \text{if } k \notin rec_i\{s\} \end{cases}$$

Updated score for the *k*-th product in the ranking

Similarity between product *k* and recommendations from session, in the range *[0, 1]*

Products recommended based on user session *s*

Boost, modifying the similarity measure and causing the product to move closer to the top of the ranking, we assumed *B = 10.0*

Original score returned by the AVA search engine

Recommendation algorithm, *i* in *Word2Vec, SENTENCE-BERT, IMAGE-BASED*

Modifier for a product that was not recommended (penalty causing the product to move to a lower position in the ranking), we assumed *P = 1.0*

Future work

We aim to combine product descriptions and images into multimodal embeddings for comprehensive item representation. Real-time feedback will dynamically refine recommendations, and analyzing user session temporal patterns will help the system adapt to changing preferences.

Project POIR.01.01.01-00-0318/20 „AVA - Autonomous Voice Assistant - development of a platform to automate the deployment of virtual assistants for the eCommerce sector" is co-financed by the European Union with funds from the European Regional Development Fund under the Operational Program Smart Growth. The project is implemented under the competition of the National Center for Research and Development: Quick Path.

European Funds Smart Growth

Republic of Poland

NCBR National Centre for Research and Development

European Union European Regional Development Fund

Contact the authors:
{m.mozolewski, g.knor}@edrone.me