BABEŞ-BOLYAI UNIVERSITY, CLUJ-NAPOCA, ROMÂNIA FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Optimizing Recommendation Systems: Strategies for Enhanced Performance and Adaptability

PhD Thesis Summary

PhD Student: Mara Deac-Petruşel
Scientific Supervisor: Professor PhD Anca Andreica

2024

Keywords: Recommendation Systems, Collaborative Filtering technique, K-Nearest Neighbors algorithm, Rating Prediction, Machine Learning

Abstract

In a world dominated by data abundance and personalized experiences, Recommendation Systems play a pivotal role in guiding decision-making processes. This thesis embarks on a focused exploration to enhance the essence of recommendations by delving into the wealth of insights embedded within data. Through meticulous investigation of various methodologies and the formulation of a novel similarity measure, the main objective of this thesis is to elevate the quality of recommendations.

The thesis begins by scrutinizing memory-based collaborative filtering techniques, with a particular emphasis on the critical role of similarity measures. Extensive experiments on diverse data sets reveal optimal similarity measures for different contexts.

Next, a novel sentiment-based similarity measure is introduced, Attractiveness-Relevance-Popularity (ARP), aimed at improving collaborative filtering by leveraging textual reviews. ARP replaces numerical ratings with sentiment scores derived from sentiment analysis lexicons, resulting in enhanced recommendation accuracy. Additionally, a robust validation framework is proposed for ARP, targeting to revolutionize the process of developing and evaluating new similarity measures.

Moreover, three approaches were designed with the goal of optimizing recommendation techniques. The first approach employs sentiment analysis techniques in conjunction with collaborative filtering, showcasing significant improvements in accuracy and recommendation quality. The second approach introduces a lexicon-based KNN collaborative filtering technique, demonstrating success in recommendation tasks with data sets containing text-based user reviews. The final section presents an unsupervised recommendation system tailored for New York Times readers, incorporating K-Means clustering to define article clusters and generate personalized recommendations.

This thesis offers valuable insights into various facets of recommendation systems, including the study and development of similarity measures, the fusion with sentiment analysis techniques, and the surprising unsupervised flavor of recommendation systems. The findings contribute to the advancement of recommendation system research, laying a robust foundation for future exploration and innovation in the field.

List of Publications

- Petruşel, Mara and Limboi, Sergiu-George. **A restaurants recommendation system: Improving rating predictions using sentiment analysis.** In 2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pages 190-197, 2019 IEEE [19] (Conference Category C - 2 points).
- Deac-Petruşel, Mara, **A comparative analysis of similarity measures in memory-based collaborative filtering.** In Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19, pages 140-151, Springer International Publishing, 2020 [6] (Conference Category C - 2 points)
- Deac-Petruşel, Mara and Limboi, Sergiu. **A sentiment-based similarity model for recommendation systems.** In 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), pages 224-230, 2020. IEEE [8] (Conference Category D - 1 point)
- Limboi, Sergiu and Deac-Petruşel, Mara. **A Validation Framework for ARP Similarity Measure**. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 1266-1271, 2021 IEEE [15] (Conference Category C - 2 points)
- Deac-Petruşel, Mara. **A Lexicon-based Collaborative Filtering Approach for Recommendation Systems.** In Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART), pages 203-210, 2022, SCITEPRESS [7] (Conference Category B - 4 points)
- Petruşel, Mara. **An Unsupervised Topic-driven New York Times Recommendation System.** In 2022 International Conference on Innovations in Intelligent Systems and Applications (INISTA), pages 1-6, 2022, IEEE [18] (Conference Category C - 2 points)

Total: 13 points.

Contents

Chapter 1

Introduction and Research Context

1.1 Motivation

In an era characterized by an unprecedented surge in data availability and a relentless pursuit of personalized experiences, the realm of Recommendation Systems (RS) emerges as a cornerstone of technological innovation. The essence of these systems lies not merely in their ability to suggest products, services, or content but in their transformative power to shape decisions, fuel economic growth, and enhance user satisfaction.

The profound impact of Recommendation Systems resonates across diverse domains, from e-commerce platforms tailoring shopping experiences to individual preferences to content streaming services curating playlists that resonate with unique tastes. In the complex landscape of modern decision-making, the significance of these systems extends beyond convenience; they encapsulate the intricate interplay between user preferences, information retrieval, and technological advancements.

The motivation to study recommendation systems arises from their dynamic and evolving nature. In a digital landscape characterized by an incessant influx of data, the ability to extract meaningful insights, discern patterns, and enhance the accuracy of recommendations is not a mere academic pursuit but a response to the evolving needs of society.

Furthermore, as digital ecosystems become increasingly complex and interconnected, the role of Recommendation Systems becomes even more critical. They serve as vital tools for managing information overload, helping users navigate vast amounts of data by presenting the most relevant options. This capability is particularly valuable in sectors such as healthcare, where personalized recommendations can significantly improve patient outcomes by suggesting tailored treatments or preventive measures based on individual health data.

The advent of big data and advancements in artificial intelligence and machine learning have propelled Recommendation Systems to new heights. These technologies enable systems to not only learn from vast datasets but also adapt to changing user behaviors and preferences in real time. This adaptability is crucial in maintaining user engagement and satisfaction in a world where preferences can shift rapidly and unpredictably.

Moreover, the ethical and societal implications of Recommendation Systems warrant close examination. As these systems wield significant influence over what users see, buy, or consume, there is an imperative to ensure that recommendations are fair, unbiased, and transparent. Addressing issues such as algorithmic bias, data privacy, and user trust forms an essential part of the ongoing development and refinement of Recommendation Systems research.

Another driving force behind the study of Recommendation Systems is their potential to foster innovation and creativity. By analyzing user preferences and behaviors, these systems can uncover latent needs and emerging trends, providing businesses with valuable insights that can guide product development and strategic planning. This not only enhances the competitive edge of businesses but also enriches the overall consumer experience by anticipating and meeting future demands.

In academic research, Recommendation Systems present a fertile ground for exploring interdisciplinary approaches that integrate insights from computer science, psychology, economics, and sociology. Understanding the nuances of user behavior and preference formation requires a holistic perspective that considers cognitive and social factors, thereby enriching the theoretical and practical dimensions of the field.

Ultimately, the motivation to delve into the Recommendation Systems domain is rooted in their profound potential to transform the way we interact with information and make decisions. Pushing the boundaries of what these systems can achieve contributes to a more personalized, efficient, and user-centric digital future.

1.2 Original Contributions

The main original contributions of this paper are presented as follows:

- The first approach [6], introduced via Chapter 3, is an in-depth comparison of diverse similarity metrics employed in memory-based collaborative filtering (MBCF), offering nuanced insights into their performance across different contexts and data set characteristics. This comparative analysis not only identifies optimal similarity measures but also provides actionable recommendations for practitioners and system designers. By bridging the gap between theory and practice, this work informs the design choices of recommendation systems, contributing to the optimization of personalized recommendations and elevating MBCF as a practical toolkit for enhancing user experiences.

- Recognizing the limitations of numerical user ratings in conveying nuanced opinions about products, leveraging text-based descriptions becomes pivotal for improving the process of recommendation. Consequently, the approach [8] systematically analyzes textual information, computing sentiment scores to better interpret users' opinions. The original contribution lies in the introduction of a sentiment-based user similarity measure, Attractiveness-Relevance-Popularity (ARP). Unlike conventional similarity measures, ARP employs sentiment ratings to determine the similarity between users, offering a unique perspective that significantly enhances the performance of the RS. It is relevant to highlight that the Sentiment Scoring Module was part of the joint research approach [8] and is not an original contribution of this thesis.

- As a follow-up to the ARP measure design, a validation framework is proposed in [15] to evaluate the added value of the new measure and to prove that it can be trustfully used instead of the traditional ones. Most of the existing approaches in literature validate new measures only in terms of evaluation techniques, for example: precision, accuracy, or mean absolute error. The original contribution within approach [15], described in detail via Chapter 4 Section 4, is the proposed validation framework for the newly designed ARP similarity measure, mainly four out of the five essential components of the validation process: metric conditions check, usefulness, expressivity, and correlation with other measures. The fifth validation component, noise robustness, is only briefly summarized in this thesis as it is not an original contribution. The details regarding the conducted numerical experiments and results can be checked in the joint research approach [15]. All in all, by introducing this validation framework, our work goes beyond conventional evaluations, ensuring a thorough examination of the ARP similarity measure's performance and its suitability for integration into recommendation processes.

- Section 5.1 of Chapter 5, brings in an approach [19] that aims to optimize the KNN recommendation technique by making use of SA in the data preprocessing phase of the recommendation process. SA is applied for classifying restaurants' text-based reviews into positive and negative. The output data set is passed to a recommendation system that, using the KNN algorithm, predicts the rating for a not-visited restaurant and generates a list of recommended restaurants for the user. This approach outperformed the results obtained when the SA step was not considered in the recommendation process. This approach [19] was developed collaboratively, with the primary focus of my research residing in the refinement and optimization of the recommendation component. To be acknowledged that the SA step is not an original contribution of this thesis.

- The approach [7], presented in Section 5.2 of Chapter 5, advocates for the implementation of a lexicon-based KNN collaborative filtering technique, integrating computed sentiment ratings in the neighborhood determination step. This methodology boasts two primary original contributions. Firstly, a thorough analysis of the textual information associated with each item is conducted and then the Vader Sentiment Analysis Lexicon [12] is applied to derive the sentiment ratings. Secondly, the lexicon-based data set is passed to the KNN user-based collaborative filtering technique. The conducted experiments reveal that the resulting text-based recommendation system produces accurate recommendations for users.

- In Section 5.3 of Chapter 5, a novel recommendation system tailored for New York Times readers is introduced, leveraging an unsupervised machine learning perspective [18]. Next, the original contributions of this work are presented. Firstly, the proposed approach integrates an unsupervised perspective by incorporating the K-Means algorithm into the traditional KNN recommendation technique. This fusion enhances the system's ability to identify clusters of articles based on similar topics or subjects, thereby refining the recommendation process. Moreover, the recommendation system is meticulously designed and personalized to cater to the extensive array of articles available on the New York Times portal. By tailoring the system to the specific content and audience of the New York Times, the efficacy and relevance of the recommendations are substantially heightened, ensuring a more enriching user experience. Furthermore, the originality of the paper extends to its rigorous numerical experiments conducted on multiple data sets sourced from the New York Times archives, encompassing articles spanning various periods. These experiments serve to validate the effectiveness and robustness of both the clustering and recommendation processes, offering empirical evidence of the system's performance across diverse data sets and temporal contexts.

Chapter 2

Key Concepts and Algorithms in Recommendation Systems

Recommendation Systems (RS) encompass software tools and methodologies designed to offer guidance to users across various decision-making scenarios: what items to purchase, which books are worth reading, or which restaurant to dine at. RS have found their inspiration in the well-known herd behavior: people often trust recommendations from others when making everyday decisions.

RS have proved in recent years to be an effective solution for the Information Overload problem. Basically a RS leads the user towards new, not yet discovered items that are most likely to be of interest to the user's current need. The item recommended by the RS to users and is most frequently part of a specific and singular category (e.g. Movies, Music, News, or Restaurants). Tailored suggestions are presented as ordered arrays of items. When computing the ranking, the RS makes use of the users' preferences, such as ratings given to products or even the navigation to a particular product page. After receiving the recommendations, the user can peruse them and decide whether to accept or decline them. Subsequently, the user may offer implicit or explicit feedback, either immediately or during a subsequent interaction. These user interactions and feedback can be retained and utilized to produce fresh recommendations during future interactions between the user and the system.

In this chapter, a detailed overview on trends of RS is presented based on different recommendation techniques. The strengths and weaknesses of these recommendation algorithms are identified, with proposed solutions offered in the form of new research opportunities.

A RS actively and continuously gathers various type of information with the goal to offer suggestions. This information references both products that are going to be recommended and those users that are receiving the suggestions. In general, the information used by RS consists: the items, the users, and the relationships between items and users.

The items recommended by the RS serve as its output and possess distinct characteristics such as complexity, value, or utility. An item's value can be either positive, indicating its relevance to the user, or negative, signifying an inappropriate selection by the user. Some items may exhibit low complexity and value, examples being news articles, websites, and movies, while other items have greater complexity and value: cameras, smart phones, computers. The items considered the most important are: the insurance policies, the financial investments, planned travels and job positions [20].

The users can have various objectives and characteristics. To personalize the suggestions and to improve the users' experience, RS exploit the gathered information about the user. The choice of information to model varies depending on the recommendation technique employed. For instance, in collaborative filtering, the user model comprises a basic list containing user evaluations of various items. Conversely, in a demographic RS, socio-demographic attributes like age, gender, education, and work experience are incorporated into the user profiles [20].

The user's preferences and requirements are stored within their model. A recommendation system acts as a tool that formulates suggestions by creating and utilizing these models. Given that personalization hinges on having an effective user model, it remains pivotal in the recommendation process. For example, in a collaborative filtering approach, the user is characterized either directly by their ratings on items, or the system deduces a vector of factor values from these ratings, reflecting variations in how users weigh each factor in their model [20].

Users can be characterized based on their behavioral patterns, such as site browsing patterns or travel search patterns. The information base about users can also include relationships between them. Therefore, the RS will use this data to suggest similar items preferred by other trustworthy users [20].

The user-item relationships represent log-like records containing significant details generated during user-system interactions, serving as input for the recommendation algorithm. Typically, ratings constitute the most common data type collected by the system [20]. Ricci et al. [20] provides a detailed classification of ratings: numerical, ordinal, binary, respectively unary.

According to Bobadilla et al. [3] the following steps are to be considered in the development of a RS:

- The data set structure (e.g numerical or textual ratings).
- The selected filtering technique: collaborative, content-based, hybrid, learning to rank, context-aware, social-based, etc.
- The selected model: memory-based or model-based.
- The data set's sparsity.
- The desired scalability.
- The RS's performance: time and memory usage.

Chapter 3

The Role of Similarity Measures in Optimizing Recommendation Accuracy

Among the numerous factors influencing recommendation performance, the selection and utilization of appropriate similarity measures have garnered substantial attention. The choice of similarity measures significantly influences the effectiveness and quality of the recommendations generated by recommendation systems.

This chapter aims to provide a comprehensive understanding of the fundamental role of similarity measures in enhancing the performance of recommendation systems.

One objective is to investigate the influence of diverse widely used similarity metrics, including but not limited to Cosine Similarity, Pearson Correlation Coefficient, and Jaccard Index, on the recommendation process. A rigorous analysis of these metrics has identified their strengths, weaknesses, and applicability across different recommendation scenarios.

Moreover, this chapter will scrutinize how the appropriate selection and integration of similarity measures can lead to optimized recommendation accuracy.

To accomplish these goals, a thorough review and analysis of relevant studies that investigated the impact of similarity measures on recommendation accuracy was conducted, including an evaluation of their methodologies, findings, and inherent limitations. Also, this chapter presents a comparative analysis of the performance of various similarity measures applied to memory-based collaborative filtering algorithms: user-based (UBCF), respectively item-based (IBCF) scenarios.

Several numerical experiments were conducted, considering the following similarity measures: PIP [2], Pearson Correlation Coefficient (PCC), Constrained Pearson Correlation Coefficient (CPC), Cosine Similarity (COS), Adjusted Cosine (ACOS), Jaccard Index (JAC), Euclidean Distance (EUC), and Spearman Rank Coefficient (SRC) [1] and memory-based collaborative filtering techniques (UBCF and IBCF).

In the evaluation step, precision and mean absolute error evaluation measures were used.

Two data sets, that are different in terms of dimensionality and sparsity, were chosen for the numerical experiments.

The MovieLens 1M[1] data set contains 1 million ratings applied to 4.000 movies by 6.000 users. The DataFiniti Hotel Reviews[2] data set consists of 10 000 reviews for 1670 hotels. Both data sets

[1] https://grouplens.org/datasets/movielens/
[2] https://data.world/datafiniti/hotel-reviews

consider one-to-five user ratings.

From a sparsity point of view, MovieLens has 95.83%, while DataFiniti Hotel Reviews 99.91%.

Results and Conclusive Remarks

In the presented study, several experiments were conducted to offer an answer to a set of essential questions in memory-based collaborative filtering approaches. The main difficulty in the design of a recommender system lies in the proper choice of similarity measure.

Table 3.1: Similarity Measures Findings.

Characteristic	**Comments**
Dataset	MovieLens and DataFiniti Hotel Reviews
Sparsity and Dimensionality impact on IBCF	JAC has the best results for small data sets with high sparsity, as it compares the presence or absence of ratings (binary values), making it applicable when dealing with data sets with a high proportion of unrated items. SRC fits the item-based approach for large data sets.
Sparsity and Dimensionality impact on UBCF	PIP is suitable for large data sets and low sparsity. Several similarities can be used for small data sets and high sparsity (COS, SRC).

The used data sets, MovieLens[3] and DataFiniti - Hotel Reviews[4], were chosen in terms of different dimensionality and sparsity features. The results of the conducted numerical experiments lead to the following conclusions. In terms of large data sets and lower data sparsity, the PIP similarity fits the user-based context, while the Spearman's Rank Coefficient could be a proper selection for the item-based context. In contrast, when having a smaller data set with high sparsity, the Jaccard similarity suits the item-based context. For the user-based scenario, multiple similarities can be chosen (COS, SRC), depending on the neighborhood size. Moreover, the main characteristics of similarity measures that positively influence the recommendation process were discussed in this analysis.

Table 3.1 reflects the summary of the presented analysis, highlighting the used data set, algorithms, and the conclusion of the similarity's impact in terms of data dimension and sparsity.

In summary, this chapter endeavors to illuminate the critical role of similarity measures in optimizing recommendation accuracy. Through the exploration of diverse similarity metrics, their influence on recommendation outcomes, and strategies for their effective utilization, we aim to contribute to the advancement of recommendation systems and offer valuable insights to researchers, practitioners,

[3] https://grouplens.org/datasets/movielens/
[4] https://data.world/datafiniti/hotel-reviews

and developers operating in this field.

To further enhance the analysis of similarity measures, a direction would be exploring additional evaluation metrics such as: coverage- to measure the proportion of items for which recommendations can be made; serendipity - to evaluate how surprising and novel the recommendations are to the user; diversity - to assess the variety of recommendations to avoid redundancy; online Measures - to analyze real-time performance metrics such as click-through rates or user interaction metrics. These metrics can provide a more comprehensive evaluation of the recommendation system's effectiveness beyond precision and MAE, ensuring a balanced and user-centric approach.

Chapter 4

Attractiveness-Relevance-Popularity (ARP) Similarity Measure

Recommendation Systems are tools that interpret the users' preferences in an attempt to generate suitable suggestions. Research studies tend to conclude that numerical user ratings are not powerful enough to truly express the users' preferences. On the other side, text-based reviews can express characteristics like sentiments, opinions, or attitudes, which are more promising for extracting valuable information. A text-based review can be used to define the sentiment or the overall opinion regarding the item. Therefore, mining the sentiment or the polarity of the textual information proves to be an essential component of a recommender system.

The similarity measure is a key concept in a wide range of domains' processes, such as Natural Language Processing, Clustering, or Recommendation Systems. In the last years, a wide range of new similarity measures have been designed and applied to different contexts. Currently, there is a deep lack in the validation and evaluation steps for novel similarities. In general, new measures are validated mostly through numerical experiments using different data sets and in terms of evaluation metrics, such as: accuracy, precision, or mean absolute error. But this is not enough, as in order to gain relevance for a domain, a more complex validation process is necessary.

This chapter presents in detail the design of a new similarity measure, **Attractiveness-Relevance-Popularity (ARP)**, which resulted from the integration of Sentiment Analysis (SA) techniques into the recommendation process to increase the accuracy of suggested items [8]. In addition, a validation framework is introduced to evaluate the added value of ARP and to prove it can be trustfully used instead of traditional similarity measures [15]. The validation process consists of five main steps: metrics conditions check, usefulness, expressivity, correlations to other measures, and noise robustness. Several criteria were considered and analyzed, such as: can the measure be applied to data sets with different data types (e.g. numerical, categorical features), or how efficient is the novel similarity?

Methodology Overview

Figure 4.1 presents the fundamental components and interactions of the designed system.

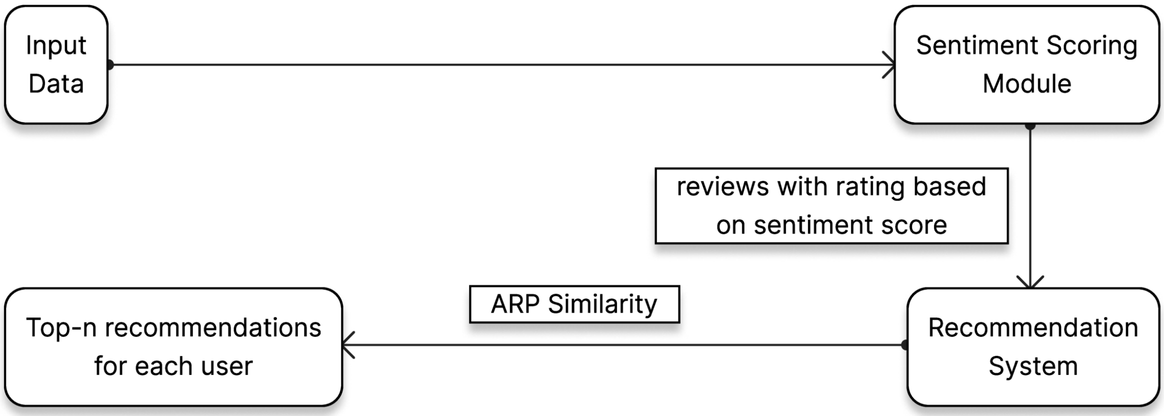

Figure 4.1: ARP: Methodology Overview.

Firstly, the collected data is passed to the Sentiment Score Module. For each text-based review, a sentiment score is computed based on the following steps. As a first step, the polarity score of a word is retrieved using the SentiWordNet[1] tool, resulting in either a positive or negative score.

Subsequently, the sentiment score of a review is computed by summing the sentiment scores of all its constituent words. A rating function is then applied to convert the sentiment score into a real one-to-five rating for each text-based review. The full details of the Sentiment Score Module are described in the joint research approach [8]. It is important to note that this aspect of the research was part of a joint research approach [8] and is not an original contribution of this thesis.

Additionally, the sentiment ratings assigned to each review replace the original user rating and are passed to the Recommender System component. The user-based KNN collaborative filtering algorithm is selected as recommendation technique. The similarity between two users is computed using the newly proposed ARP sentiment-based similarity model. Finally, the Recommendation System component outputs a top-n recommendation list for each user.

ARP: Design Principles

The sentiment-based measure ARP is defined by three factors of similarity: Attractiveness, Relevance, and Popularity.

The attractiveness of a review is dependent on the number of positive and negative scores of its containing words. In this context, it marks how appealing an item could be for users. The averages of positive and negative words' scores are considered to diminish the big differences between words with positive or negative scores from a review.

The popularity concept has been defined from three perspectives: for a review, for a user, and for a user in relation to another one. The popularity factor reveals how much the review/user deviates from the mean.

[1]http://ontotext.fbk.eu/sentiwn.html

The relevance factor is based on attractiveness and can be defined for a user, a review, and a user in relation to another one. Relevance indicates the average deviation of the j^{th} review given by the i^{th} user in terms of attractiveness.

The proposed ARP similarity measure has values ranging in [-1,1], where a value close to 1 indicates a greater similarity between users. Compared to traditional similarity measures that use the numerical one-to-five user rating, the ARP similarity is formulated based solely on the previously calculated sentiment ratings of the text-based reviews.

ARP: Application in Recommendation Systems

In the recommendation process, the user-to-user version of the KNN CF algorithm is chosen to be used. A crucial step in the development of a RS using the collaborative filtering technique is selecting the appropriate similarity measure. To determine the target user's group of k neighbors, the ARP measure calculates similarities among users' sentiment ratings. Based on the predicted rating, a top-n recommendation list is generated for the target user.

The Yelp Restaurants Reviews[2] and Datafiniti Hotel Reviews[3] data sets were used in the numerical experiments. For both, the sentiment ratings were used. Similar to the sentiment ratings approach, 20% of data was used for the testing phase of the recommendation system.

The first numerical experiment aims to determine the performance of the RS using the ARP similarity measure. Experiments are conducted on both data sets and different values for the number of neighbors k and the number of recommendations n have been considered. The results for the Yelp data set, show that the best values, in terms of MAE and RMSE, were achieved for $k = 5$ and top 15 recommendations (MAE=0.03; RMSE=0.18), respectively for $k = 10$ and top 5 recommendations (MAE=0.09; RMSE=0.10).

The results for the DataFiniti data set, show that the best values, in terms of MAE, were achieved for $k = 5$ and top 15 recommendations (MAE=0.07), respectively for $k = 10$ and top 5 recommendations (MAE=0.17). The best outcomes for RMSE were recorded for $k = 5$ and top 10 recommendations (RMSE=0.30), respectively for $k = 10$ and top 5 recommendations (RMSE=0.37).

The second experiment aims to compare ARP's performance upon several traditional similarity measures: PIP [2], Pearson Correlation Coefficient (PCC), Cosine Similarity (COS), Jaccard Index (JAC) and Spearman Rank Coefficient (SRC) [1]. Considering the best results obtained in the first experiment, the second one is conducted on $k = 5$ neighbors and top 15 recommendations, respectively $k = 10$ and top 5 recommendations. For Yelp data set, results show that the SRC similarity measure yields the best values (MAE=0.01 and RMSE=0.12). The ARP measure also produces comparable and favorable values, with MAE=0.03 and RMSE=0.18, making it the second-best measure.

For the DataFiniti dataset, considering the setup k=5 and n=15, the SRC similarity measure achieves the best values, with MAE=0.02 and RMSE =0.13. The ARP measure also yields strong results, with MAE=0.03 and RMSE=0.18, closely matching the performance of the SRC similarity measure.

[2]https://www.yelp.com/dataset
[3]https://data.world/datafiniti/hotel-reviews

The ARP Validation Framework

To prove that the ARP similarity measure can be trustfully used as an alternative to the well-known similarity measures in the collaborative filtering context, a validation framework is proposed. The validation process is described in Figure 4.2.

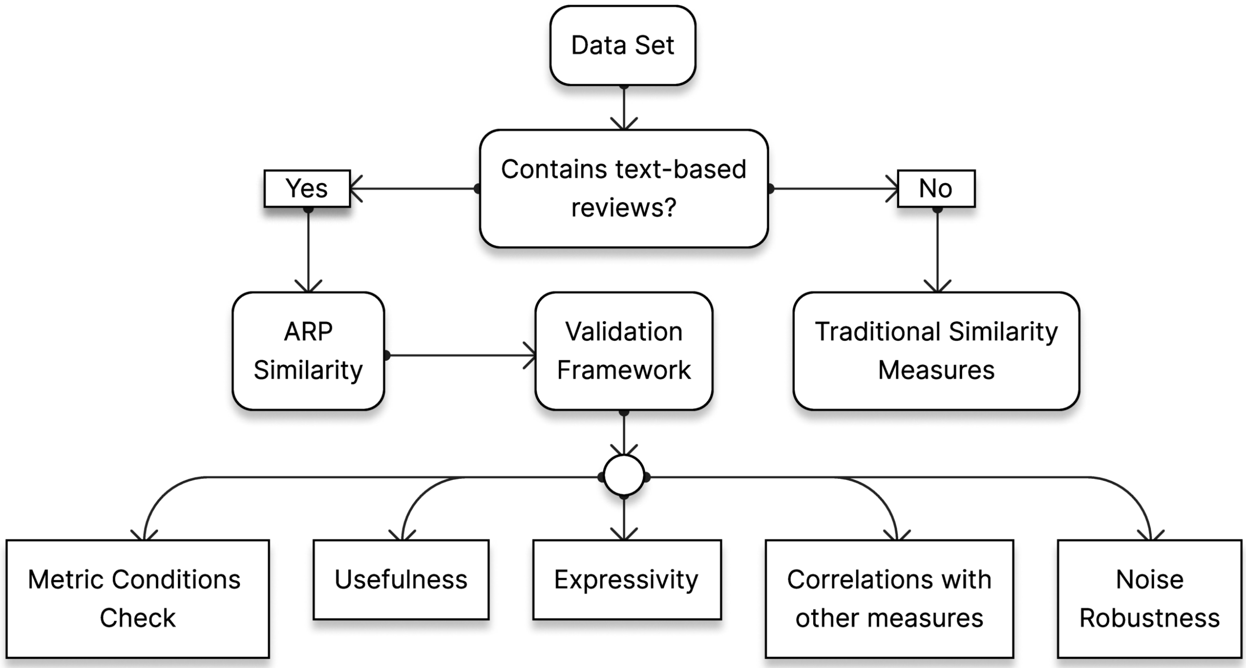

Figure 4.2: ARP: Validation Framework Components.

In the data collection phase, it must be checked if the data set contains text-based reviews for items. If this is the case, then the ARP measure can be used to improve the rating prediction of a recommendation system and the validation framework can be applied. If not, other well-known similarity measures such as the Pearson Correlation Coefficient or Cosine Similarity should be used in the recommendation process.

The validation framework consists of the following components: Metric Conditions, Check Usefulness, Expressivity, Correlations to other similarity measures, and Noise Robustness.

Concluding Remarks and Contributions

This chapter presented a newly designed sentiment-based user similarity measure (ARP), that exploits the user's opinions derived from his text-based reviews. The ARP measure was tested for the K Nearest Neighbors collaborative filtering (CF) algorithm, using two public data sets: Yelp Restaurants Reviews and Datafiniti Hotel Reviews. ARP's performance was compared to traditional similarity measures in the CF context. Results show that ARP can successfully replace traditional similarity measures like Pearson Correlation Coefficient, Cosine, or PIP [2].

Even though comparisons to related work papers [9, 5, 17, 11, 13, 14] are not possible, the clear advantage of the proposed model is that the recommendation algorithm does not suffer any adaptions, so the ARP measure can be easily integrated into the existing process. A very important aspect is that ARP similarity can be used also for input data that does not contain numerical ratings, but only textual descriptions of items. This aspect is not valid for the traditional measures.

Moreover, a validation framework was proposed for ARP, taking into account five independent components: usefulness, expressivity, correlation to other measures, metrics' conditions checks, and noise robustness.

Future work plans include applying the ARP similarity on multiple data sets to study the effects of dimensionality and sparsity features over its performance. Also, other relevant features of the textual information (e.g. part-of-speech) for an item can be explored.

Chapter 5

Optimization Frameworks for Recommendation Systems' Techniques

5.1 Leveraging Sentiment Analysis for Improved Rating Predictions in Recommendation Systems

The goal of this section is to introduce an original perspective [19] that improves the results of CF by incorporating a Sentiment Analysis (SA) preprocessing step in the recommendation process. This method was tested through experimental scenarios using a restaurants review data set as follows: the sentiment classifier performs the first level of filtering, while the CF algorithm is used at the second level of filtering. The results of the proposed approach are compared with the ones produced by the baseline CF algorithm, where the sentiment outcomes are not considered. The final output is a more accurate top n recommendation list generated for the users.

The proposed approach integrates SA techniques into a CF algorithm. In Figure 5.1, the main components and interactions of the proposed system are described. Based on the input data set, the sentiment classifier produces as output a positive/ negative labeled data set. The latest will be passed to the recommender system in order to generate a top-n items recommendation list.

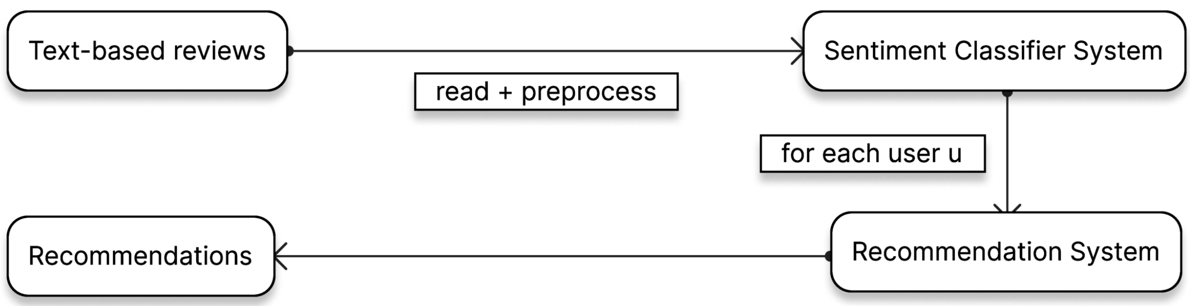

Figure 5.1: Sentiment-Enhanced Recommendation System Framework

Experimental Setup and Results

The input data set consists of 8.539 reviews positively or negatively labeled, keeping the following features of interest: business id, the stars given by a user (one-to-five values), user id and the review in free text format.

In the evaluation phase, 20% of the input data set was used as a testing set.

To evaluate the sentiment-enhanced restaurants' recommendation system (SA-enhanced RRS), several experiments have been conducted considering different sizes of neighborhoods ($k = 5$ and $k = 10$) and different numbers of generated recommendations ($n = 3$, $n = 5$, $n = 10$ and $n = 15$). In this context, the results obtained were further compared with the scenario without considering the new sentiment labels (baseline RRS). The results show that the Sentiment Analysis step increases the quality of the recommendations, as better values were obtained for precision, recall, F-score, and MAE evaluation measures.

In the recommendation phase only reviews labeled with a positive sentiment, in the SA step, are considered to be relevant and are used in the Pearson's Correlation similarity computation. An advantage is that the SA component runs independently from the recommendation process and the testing data is used as input for the recommender system.

5.2 A Lexicon-based Collaborative Filtering Approach for Recommendation Systems

This section presents an original approach [7] designed to capture the users' interests from the text-based items' reviews to produce good rating predictions for items and accurately generated recommendations for users. The items' descriptions are passed to a SA Lexicon, which outputs a sentiment score indicating the polarity of the text (positive, negative, or neutral). Based on the sentiment score, a KNN user-based CF algorithm was applied. The RS uses solely the sentiment scores (called sentiment ratings), instead of the numerical ratings. Results have proven a positive impact of the text-based approach on the performance of the recommendation system.

Since the text-based items' descriptions reveal more valuable information compared to the plain numerical ratings for the recommendation process, the focus of the proposed approach is to make use solely of the textual information when building the recommendation system, regardless of the numerical ratings. The textual input is exploited using a lexicon-based technique to determine the polarity score of a review. The resulting scores are the sentiment ratings taken into consideration for the user-based kNN collaborative filtering algorithm.

Figure 5.2 presents the proposed architecture of the designed system. After the data collection phase, the text-based items' reviews serve as input for a sentiment lexicon that determines a sentiment rating for an item. The data set enhanced with the computed sentiment rating is further passed to a recommendation system.

The proposed approach uses, for the sentiment analysis task, a sentiment lexicon, which was selected based on the complex and thorough comparison presented in [12]. The Vader Sentiment Lexicon was compared to several ones from literature (Linguistic Inquiry Word Count, General Inquirer, Affective Norms for English Words, SentiWordNet, SenticNet, Word-Sense Disambiguation) and produced, in most cases, the best results.

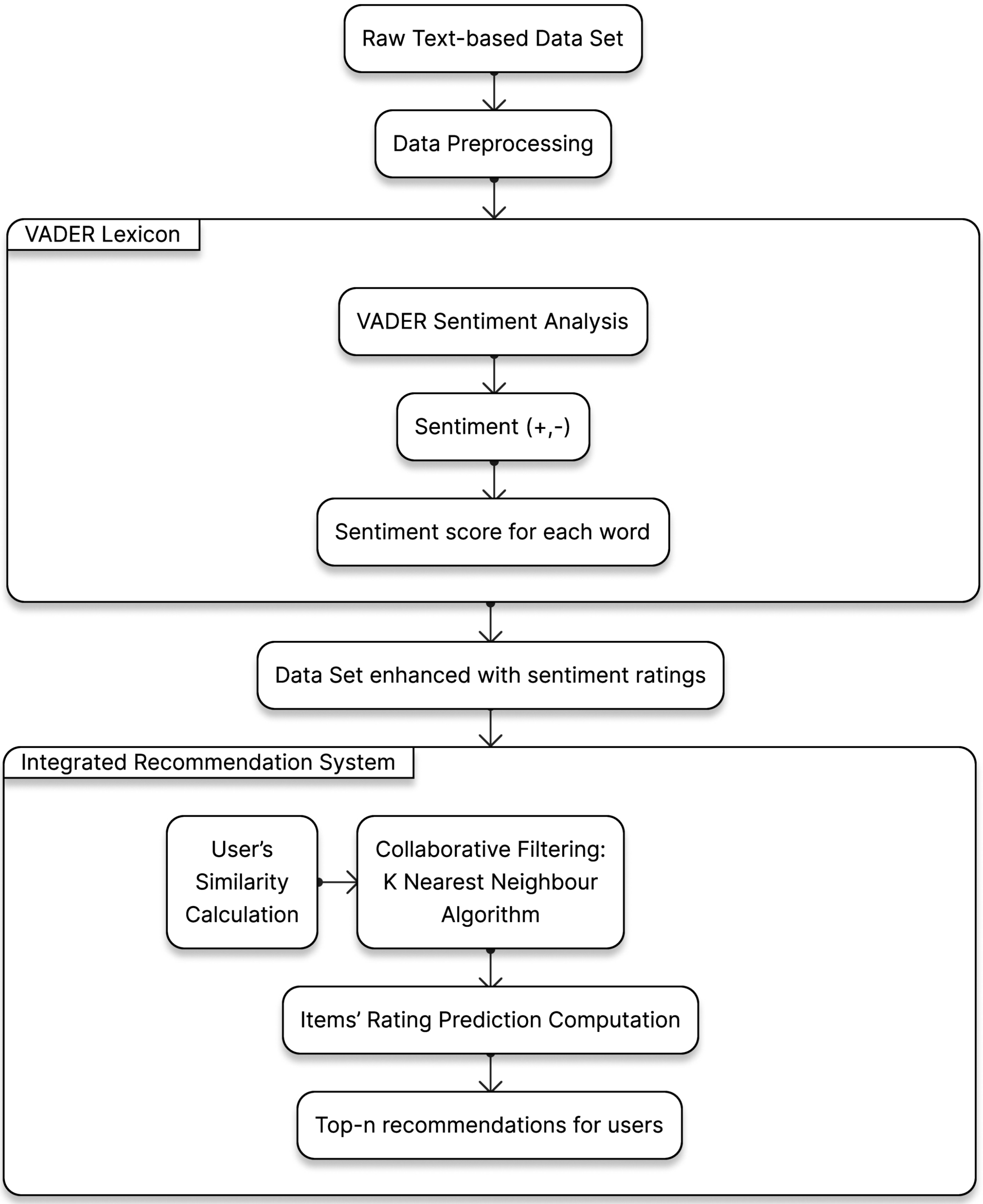

Figure 5.2: Lexicon-based RS System Architecture.

The data set containing in addition the reviews with sentiment scores represents the input data for the recommendation system. The classical KNN CF algorithm is then applied as a recommendation technique [19].

Numerical Assessment of the Lexicon-Driven Recommendation System Across Three Diverse Data Sets

To highlight the value-added by the proposed lexicon-based KNN CF approach in improving the rating prediction accuracy, several numerical experiments were conducted on three data sets containing text-based reviews for items.

For the neighborhood determination in the KNN, various popular similarity measures from literature were applied: Pearson Correlation Coefficient (PCC), Cosine (COS), Euclidean (EUC), Constrained Pearson Coefficient (CPC), Spearman Rank Coefficient (SRC), Jaccard Similarity (JAC) [1], [21] and PIP [2]. Independent scenarios are designed for different values of k (the neighborhood size) and n (the number of generated recommendations).

In the evaluation process, the MAE and RMSE measures are computed to establish the accuracy of the generated recommendations.

Moreover, the proposed lexicon-based approach is compared to another text-based KNN CF approach described in [22], in terms of RMSE performance measure. Both approaches use text-based reviews instead of numerical ones and the experiments are conducted on the Rotten Tomato Critic Reviews data set. Unfortunately, for the approach in [22], the details regarding the chosen values for the neighborhood size (k) and number of recommendations (n) are not shared in the experimental setup. Although both approaches make use of textual items' descriptions, there is a difference in the sentiment score definition (substituting the numerical rating). Terzi et al. [22] compute the distance between two words based on the shortest distance between them, while in the proposed approach the sentiment score is obtained based on the information derived from the Vader Lexicon [12].

Even though the quantitative results in [22] are better, the presented approach is different from a qualitative point of view, using a lexicon-based collaborative filtering technique. The proposed technique has value especially from the semantic point of view, considering words' polarities (positive, negative, neutral) compared to [22], which is based on the set of common words. Overall, this comparison highlights the fact that the presented approach generates good and trustworthy results and confirms again that text-based reviews indeed offer valuable information for the recommendation process.

The results obtained in the conducted numerical experiments show that the presented approach can be successfully used to solve recommendation tasks, for data sets containing text-based user reviews. As future work, the approach could be extended to also consider different types of review elements besides words, such as review topics or aspect opinions.

5.3 An Unsupervised Topic-driven Recommendation System

This section presents the newly designed New York Times Recommendation System (NYT RS), aiming to propose relevant articles for a reader on particular topics of interest. The collected papers from the

New York Times[1] website are grouped based on manually detected topics using the K-Means clustering algorithm. The resulting clusters are used in the recommendation process to suggest articles from the same group as the already-read article. This method enhances the quality of the classic k Nearest Neighbors (kNN) recommendation technique.

The proposed approach aims to increase the quality of New York Times articles' recommendations. The system defines a list of recommendations for a read article using an unsupervised topic-driven perspective. Figure 5.3 presents the architecture of the built NYT RS. There are two main components of the system: the clustering and the recommendation processes. After the data collection phase, the input data is passed through a data cleansing and pre-processing phase. Next, a data representation step is required, as the clustering algorithms can only deal with numerical input. The k-Means clustering algorithm is used to obtain groups of similar articles based on topics. The Silhouette Index [16] and Dunn Index [10] are computed to evaluate the clustering process. The second principal component of the system is the KNN content-based algorithm [4] that determines the k most similar articles for a given article read by the user on the New York Times portal, based on the previously determined topic-based clusters. Finally, the quality of the generated recommendations is evaluated using the accuracy measure, and the outcome is represented by a list of similar articles that a user can read on a specific topic of interest.

Numerical Experiments: Assessing the Performance of the New York Times Recommendation System

Several experimental setups were designed to validate the presented methodology and determine the quality of the proposed unsupervised topic-driven recommendation system.

Two data sets were used in the conducted experiments. The first[2] one consists of 16.787 articles from the well-known New York Times portal. It contains the following features:

- section is the category of the article (e.g. politics, science, game, etc.);
- material is the type of article (e.g., editorial, news);
- headline;
- abstract;
- publication date;
- keywords;

As the whole text is not public, only the abstract of each article is analyzed. Articles between the 1^{st} of January and 31^{th} of December 2020 are considered. Based on an in-depth analysis (also considering the section and keywords fields), five major topics were detected for the collected data: COVID-19, Donald Trump, the Black Lives Matter protests and movement, Joe Biden, and others (like forest fires, Oscar awards, etc.).

The second[3] data set consists of 10.732 New York Times articles from the end of 2017 and mid of 2018 with the following features: author, title, content, publication date and the url of the article.

[1] https://www.nytimes.com
[2] https://www.kaggle.com/benjaminawd/new-york-times-articles-comments-2020/
[3] https://www.kaggle.com/mathurinache/10700-articles-from-new-york-times

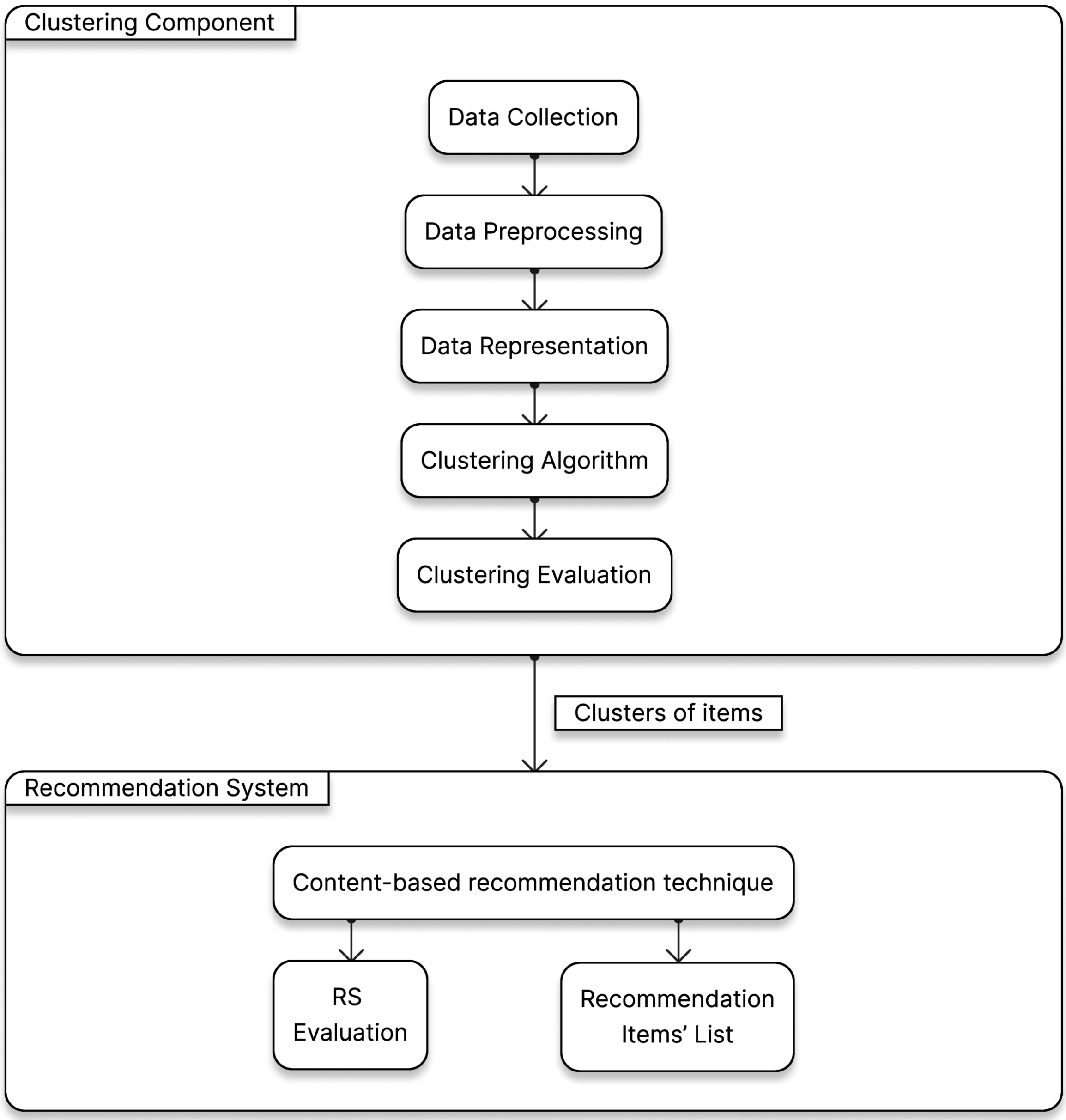

Figure 5.3: NYT Recommendation System Architecture.

The recommendation process is conducted directly for the content of the article. In addition, four essential themes were detected: the #metoo movement, Donald Trump, events related to Iran, and miscellaneous (e.g., climate change).

NYT Clustering Process

The k-Means algorithm is applied based on the Euclidean distance with several values for the k parameter. The experimental results show that the optimal number of clusters matches the number of subjects identified in the two data sets. Therefore, for the 2020 New York Times articles, k is set to 5, and for the 2017/2018 New York Times articles, to 4. The results reflect that the k-Means algorithm performs very well on the New York Times articles for the considered k values.

NYT Recommendation Process

The clusters of articles are used as input in the KNN content-based recommendation system. Multiple experimental scenarios are defined based on the selected similarity measure (Cosine, Jaccard, or Euclidean Distance) and the different values of k (the number of neighbors).

The experiments highlight the following aspects:

- The best values for accuracy were achieved using the Jaccard similarity, as it is directly applied to the textual input, without word embeddings. On the other hand, the Cosine similarity also produces good results that strengthen the idea that the system performs well.
- The most relevant value for k is five. The highest the selected k number of neighbors, the lowest the quality of the recommendations.
- For the 2017/2018 NYT articles data set, better recommendations are generated, as the whole article is considered compared to the 2020 NYT articles one, where the system gets as input only the abstracts.

The New York Times Recommendation System is a tool that offers users the possibility to read articles correlated with their interests and expectations. The original part is the unsupervised flavor that increases the quality of recommendations. Suggesting articles related to the just read one, based on their cluster belonging, shows to be an excellent path for further exploring. This aspect is sustained by the experimental results that show high accuracy values for the entire process. In addition, the clustering process itself is evaluated in terms of Silhouette and Dunn indexes.

The proposed approach is the starting point of unsupervised perspectives applied to recommendation systems, but further investigations are necessary. The plan implies using multiple and more extensive data sets, as the presented approach can be applied to any text-based data set. In addition, the impact of several similarity measures on the clustering component and the variety of word embeddings that can be used to model textual information should be analyzed further.

Chapter 6

Conclusions and Final Remarks

As we navigate through the findings and contributions documented in the preceding chapters, this concluding chapter not only summarizes the collective discoveries but also propels us toward the horizon of future possibilities, where the refinement of insightful recommendations remains an ever-persistent pursuit.

The first perspective [6], presented in Chapter 3, scrutinizes the memory-based collaborative filtering technique, with a particular focus on the critical role of similarity measures. Through extensive experiments on the Movie Lens 1M and DataFiniti Hotel Reviews data sets, the study unveiled which are the optimal similarity measures to be used in different contexts (e.g. considering the data dimensionality and sparsity). The following similarity measures are considered for analysis in the recommendation process: Pearson Correlation Coefficient, Constrained Pearson's Correlation, Cosine Similarity, Adjusted Cosine, Euclidean Distance, Spearman's Rank Correlation, Jaccard, and Proximity-Impact-Popularity (PIP) [6]. The PIP similarity emerged as fitting for user-based CF in large data sets with lower sparsity, while Spearman's Rank Coefficient showed promise in item-based CF scenarios. The Jaccard similarity performed the best for smaller data sets with high sparsity. In addition, the choice of a bigger number of neighbors (k) in KNN algorithm increases the quality of recommendations for large data sets, while a lower value is preferable when working with a reduced data set.

All in all, Chapter 3 not only explores the nuances of similarity measures but also identifies their impact on the recommendation accuracy, laying the groundwork for the subsequent chapters.

Considering the findings showcased by the comparative analysis on how similarity measures influence the recommendation process presented in Chapter 3 [6], Chapter 4 introduces a novel sentiment-based similarity measure, Attractiveness-Relevance-Popularity (ARP), aiming to improve the performance of CF by leveraging textual reviews, instead of numerical ones. The ARP measure makes use of a Sentiment Analysis lexicon (Senti Word Net) to extract the sentiment score for a given text-based review. Like this, the original one-to-five rating is replaced by the sentiment score. Then, the data set enhanced with the sentiment scores is passed to the KNN algorithm and a top n recommendation list is generated. The numerical experiments are conducted on the Yelp Restaurants Review and Datafiniti Hotel Reviews data sets and the results, evaluated in terms of MAE and RMSE, show that ARP performs better than most of the classical similarity measures, being suitable to be used for data sets containing only text-based reviews. This is a shortcoming of the traditional similarity measures.

In addition, Chapter 4 proposes a validation framework for the ARP similarity measure, based on five components: usefulness, expressivity, correlation to other measures, metrics' condition checks, and noise robustness. The design of the validation framework targets to revolutionize the process of developing new similarity measures by clearly highlighting the added value, compared to how similarity measures are currently validated in approaches from literature based solely on numerical experiments and evaluation metrics like accuracy, precision, or MAE.

Chapter 5 consists of three Sections (5.1, 5.2, and 5.3) introducing three individual approaches, having the goal to optimize different recommendation techniques.

In the first approach [19], presented in Section 5.1, SA techniques were employed in conjunction with the user-based collaborative filtering technique (the KNN algorithm), showcasing remarkable improvements in accuracy and recommendations' quality. The Sentiment Classifier receives as input a data set containing both numerical and text-based reviews and produces as output a positive/negative labeled data set (each numerical rating is enhanced by a sentiment label). The labeled data set is passed to the KNN algorithm that generates a recommendation list. To evaluate the proposed sentiment-enhanced RS, numerical experiments were conducted on the Yelp Restaurants' Reviews data set and the results were compared to a baseline approach (that does not consider the sentiment labels, just the numerical ratings). The conclusion was that the SA techniques included in the data preprocessing step of the recommendation process increses the performance of the KNN algorithm and the quality of the suggestions in terms of precision, recall, f-score and MAE.

The journey continued with the approach [7], described in Section 5.2, which introduced a lexicon-based KNN collaborative filtering technique, marking a departure from machine learning algorithms prevalent in the literature. Leveraging the Vader Lexicon for determining sentiment ratings, the approach showcased success in recommendation tasks with data sets containing text-based user reviews. This section not only emphasized the success of the lexicon-based approach in optimizing the KNN algorithm but also suggested future work involving the consideration of different review elements besides words.

The New York Times Recommendation System took center stage in Section 5.3, presenting an unsupervised flavor for recommendation systems. The K-Means algorithm defines clusters according to the most frequent topics reflected in New York Times collected articles. The resulting article clusters are used as input in the KNN collaborative filtering algorithm and recommendations are generated from the cluster the current read article belongs to. The evaluation of the clustering process, considering Silhouette and Dunn indexes, further validated the proposed approach. Future work includes the exploration of various similarity measures and word embeddings, analyzing their impact on the clustering components, and extending the proposed approach to multiple and more extensive data sets.

Collectively, this thesis delved into various facets of recommendation systems, from the intricacies of similarity measures to sentiment analysis, lexicon-based approaches, and unsupervised topic-driven recommendations. The findings and contributions lay a robust foundation for the advancement of recommendation system research. In conclusion, the thesis has not only contributed valuable insights to the field of recommendation systems but has also paved the way for future explorations and refinements. The diverse methodologies explored offer a holistic understanding of the intricate landscape of recommendation system research, propelling us toward the horizon of future possibilities.

Bibliography

[1] Ajay Agarwal and Minakshi Chauhan. Similarity measures used in recommender systems: a study. *International Journal of Engineering Technology Science and Research IJETSR, ISSN*, pages 2394–3386, 2017.

[2] Hyung Jun Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178(1):37–51, 2008.

[3] Jesus Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutierrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[4] Bei-Bei Cui. Design and implementation of movie recommendation system based on knn collaborative filtering algorithm. In *ITM web of conferences*, volume 12, page 04008. EDP Sciences, 2017.

[5] Rafael M D'Addio and Marcelo G Manzato. A sentiment-based item description approach for knn collaborative filtering. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pages 1060–1065, 2015.

[6] Mara Deac-Petruşel. A comparative analysis of similarity measures in memory-based collaborative filtering. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12-14, 2020, Proceedings, Part II 19*, pages 140–151. Springer, 2020.

[7] Mara Deac-Petruşel. A lexicon-based collaborative filtering approach for recommendation systems. In *International Conference on Agents and Artificial Intelligence (ICAART)*, pages 203–210, 2022.

[8] Mara Deac-Petruşel and Sergiu Limboi. A sentiment-based similarity model for recommendation systems. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 224–230. IEEE, 2020.

[9] Ruihai Dong, Michael P O'Mahony, Markus Schaal, Kevin McCarthy, and Barry Smyth. Sentimental product recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 411–414, 2013.

[10] Tanvi Gupta and Supriya P Panda. Clustering validation of clara and k-means using silhouette & dunn measures on iris dataset. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 10–13. IEEE, 2019.

[11] Davide Feltoni Gurini, Fabio Gasparetti, Alessandro Micarelli, and Giuseppe Sansonetti. A sentiment-based approach to twitter user recommendation. *RSWeb@ RecSys*, 1066, 2013.

[12] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.

[13] C. Jiang, L. Xia, and S. Li. A sentiment-based similarity method for cold-start recommendations. *Knowledge-Based Systems*, 189:105116, 2020.

[14] Q. Li, M. Zhang, and L. Li. A collaborative filtering recommendation algorithm based on sentiment similarity. *Mathematical Problems in Engineering*, 2019:1–9, 2019.

[15] Sergiu Limboi and Mara Deac-Petruşel. A validation framework for arp similarity measure. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1266–1271, 2021.

[16] A Rasid Mamat, F Susilawati Mohamed, M Afendee Mohamed, N Mohd Rawi, and M Isa Awang. Silhouette index for determining optimal k-means clustering on images in different color models. *Int. J. Eng. Technol*, 7(2):105–109, 2018.

[17] NA Osman, Shahrul Azman Mohd Noah, and M Darwich. Contextual sentiment based recommender system to provide recommendation in the electronic products domain. *International Journal of Machine Learning and Computing*, 9(4):425–431, 2019.

[18] Mara Petruşel. An unsupervised topic-driven new york times recommendation system. In *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6. IEEE, 2022.

[19] Mara Petruşel and Sergiu-George Limboi. A restaurants recommendation system: Improving rating predictions using sentiment analysis. In *2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 190–197. IEEE, 2019.

[20] Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. *Introduction to Recommender Systems Handbook*, pages 1–35. Springer US, Boston, MA, 2011.

[21] Mr Sridhar Dilip Sondur, Mr Amit P Chigadani, and Shantharam Nayak. Similarity measures for recommender systems: a comparative study. *Journal for Research*, 2(3), 2016.

[22] Maria Terzi, Matthew Rowe, Maria-Angela Ferrario, and Jon Whittle. Text-based user-knn: Measuring user similarity based on text reviews. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 195–206. Springer, 2014.