Bing-CF-IDF+: A Semantics-Driven News Recommender System

Emma Brocken[1], Aron Hartveld[1], Emma de Koning[1], Thomas van Noort[1], Frederik Hogenboom[1], Flavius Frasincar[1][0000−0002−8031−758X] (✉), and Tarmo Robal[2][0000−0002−7396−8843]

[1] Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands
[2] Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia
{emmabrockenn,aronhartveld,emmadekoning,thomasvnoort}@gmail.com,
{fhogenboom,frasincar}@ese.eur.nl
tarmo.robal@ttu.ee

Abstract. With the ever growing amount of news on the Web, the need for automatically finding the relevant content increases. Semantics-driven news recommender systems suggest unread items to users by matching user profiles, which are based on information found in previously read articles, with emerging news. This paper proposes an extension to the state-of-the-art semantics-driven CF-IDF+ news recommender system, which uses identified news item concepts and their related concepts for constructing user profiles and processing unread news messages. Due to its domain specificity and reliance on knowledge bases, such a concept-based recommender neglects many highly frequent named entities found in news items, which contain relevant information about a news item's content. Therefore, we extend the CF-IDF+ recommender by adding information found in named entities, through the employment of a Bing-based distance measure. Our Bing-CF-IDF+ recommender outperforms the classic TF-IDF and the concept-based CF-IDF and CF-IDF+ recommenders in terms of the F_1-score and the Kappa statistic.

Keywords: News recommendation system · Content-based recommender · Semantic Web · Named entities · Bing-CF-IDF+

1 Introduction

The ever growing information stream on the Web is gradually overwhelming the rapidly increasing population of Web users that try to access information matching their needs. An automated and accurate approach for distinguishing between relevant and non-relevant content is becoming of utmost importance for fulfilling the basic needs of the people accessing the Web. Recommender systems [1] have proven to be powerful tools for efficient processing of media and news content. Such systems build up user profiles by gathering information on recently viewed content, e.g., by exploiting domain models [18]. New content is analyzed in a similar fashion, so that similarities between user profiles and content can be computed, thus supporting a personalized Web experience [19,20] through efficient and intelligent procedures to deal with the information overload.

Traditionally, there are three kinds of recommender systems: content-based recommenders, collaborative filtering recommenders, and hybrid recommenders [5]. Content-based recommenders use the content of the unseen news items, media, etc., to match the interests of the user. Collaborative filtering recommenders find similar users and recommend new content of interest to the most similar users. Hybrid recommenders combine the former two methods. In this paper, a new content-based recommender is proposed that is aimed specifically towards news recommendation. Therefore, solely content-based recommender systems are discussed in the remainder of this paper.

Content-based news recommenders suggest unread news items based on similarities between the content of the news item and the user profile. The similarity can be computed in various ways, each measure utilizing different types of information. Some measures are based on terms (text strings) found in news items, while others are based on synsets or concepts. In this paper, we propose an extension to the previously proposed semantics-driven CF-IDF+ recommender [9] that has already proved to outperform the classic TF-IDF [21] and CF-IDF [12] recommenders. Where TF-IDF employs term-based similarities, CF-IDF adds the notion of concepts. CF-IDF+ additionally makes use of concepts that are related to concepts extracted from a news article or user profile, providing more accurate representations.

Another content-based recommendation method is based on named entities within a document. Named entities can be considered as real-world instantiations of objects, such as persons and locations. Typically, named entities are used for text analytics and information extraction purposes, e.g., by supporting more efficient search and question answering algorithms, text classification, and recommender systems [22]. The latter systems often have to deal with large amounts of (semi-)unstructured texts. By omitting the irrelevant words and only considering named entities, the dimensionality of similarity computations can be greatly reduced, thus allowing for less expensive, yet accurate recommendations. This is also in line with the usage of concepts and synsets employed in our news recommenders, and could be a beneficial addition to our systems.

Named entities appear often in news items, yet are mostly neglected because they are, for instance, not present in domain ontologies that underly concept-based recommenders. As a consequence, the CF-IDF+ method does not use all the information that is provided by named entities. A possible solution to this problem is the introduction of a methodology that takes into consideration page counts gathered by Web search engines such as Google or Bing for specific named entities. In earlier work, originally, we made use of Google named entities. However, we had to move to Bing as the usage of Google API was not for free anymore, while Bing API usage was still for free.

The recommender proposed in this paper extends the CF-IDF+ method by using information given in the named entities of news items. It combines the results of the CF-IDF+ method with similarities computed by the Bing search engine, which offered, at the time of conducting the research, a free API [3]. Our proposed recommender, Bing-CF-IDF+, consists of two individually weighted parts. The CF-IDF+ recommender computes the similarity based on concepts, whereas the Bing recommender computes the similarity based on named entities. Only the named entities that do not appear in the concepts are considered by the Bing-CF-IDF+ recommender. The main contribution of this work is the joint exploitation of concepts and their relationships from a domain ontology (CF-IDF+), on one side, and named entities and a search engine-based distance measure (Bing), on the other side, in a news recommender system.

The remainder of this paper is organized as follows. In Sect. 2, related work on previously proposed recommenders is discussed. Section 3 provides an introduction to our method and its implementation, and Sect. 4 evaluates the performance of Bing-CF-IDF+, compared against CF-IDF+, CF-IDF, and TF-IDF recommenders. Section 5 provides conclusions and some additional directions for future work.

2 Related Work

The research endeavours on profile-based (news) recommenders have been plentiful [14]. These recommenders compute similarity levels between news items and user profiles derived from previously read articles, and use these for recommending unseen items. In this section, we focus on recommenders employing terms, synsets, concepts, relations, and named entities.

2.1 Term-Based Recommendation

TF-IDF [21], one of the most commonly used methods for recommending news items, is based on news item terms. The method combines the Term Frequency (TF), which is the frequency of specific terms within a document, and the Inverse Document Frequency (IDF) [16], which is a measure of the fraction of documents that contain these terms. This method is often combined with the cosine similarity method to determine the similarity between users and news articles.

The term frequency of term $t \in T$ in document $d \in D$, $\mathrm{tf}(t, d)$, and its associated inverse document frequency $\mathrm{idf}(t, d)$ are computed as follows:

$$\mathrm{tf}(t, d) = \frac{n_{t,d}}{\sum_k n_{k,d}} , \tag{1}$$

$$\mathrm{idf}(t, d) = \log \frac{|D|}{|d \in D : t \in d|} , \tag{2}$$

where term frequencies are calculated by dividing the frequency that term t occurs in news item d ($n_{t,d}$) by the total number of all terms in news item d. The inverse document frequency is computed as a division of the total number of news items $|D|$ by the amount of news items in which term t can be found. Subsequently, TF-IDF is computed as a multiplication of TF and IDF:

$$\text{tf-idf}(t,d) = \text{tf}(t,d) \times \text{idf}(t,d) \ . \tag{3}$$

This TF-IDF score is large for terms that occur frequently in a particular news item d, but not often in all other news items. Last, the similarity between unread news items and the user's interest is computed according to a cosine similarity function:

$$\text{sim}_{\text{TF-IDF}}(d_u, d_r) = \frac{d_r \cdot d_u}{||d_r|| \times ||d_u||} \ , \tag{4}$$

where d_r is the vector representation of the user's interest and d_u is the vector representation of an unread news item. The larger $\text{sim}_{\text{TF-IDF}}$ is, the more similar the unread news item and user's interest are. All unread news items that have a higher similarity value with a user profile than a certain cut-off value are recommended to the corresponding user.

2.2 Synset-Based Recommendation

A similar method to the TF-IDF method is the Synset Frequency - Inverse Document Frequency (SF-IDF) method [6]. This method uses synonym sets (synsets) associated to terms rather than terms alone. Synsets are provided by a semantic lexicon such as WordNet [10]. Due to ambiguity, a single term can have multiple synsets, thus requiring word sense disambiguation, e.g., by using the adapted Lesk algorithm proposed in [2] and implemented in [15]. The SF-IDF measure and its corresponding cosine similarity scores are computed using the same equations as introduced for TF-IDF, only by replacing term t by synset s, so that $\text{sf}(s,d) = n_{s,d} / \sum_k n_{k,d}$ and $\text{idf}(s,d) = \log |D| \, / \, |d \in D : s \in d|$, and hence

$$\text{sf-idf}(s,d) = \text{sf}(s,d) \times \text{idf}(s,d) \ . \tag{5}$$

Then, the previously defined cosine similarity is used for computing $\text{sim}_{\text{SF-IDF}}$.

2.3 Concept-Based Recommendation

The Concept Frequency - Inverse Document Frequency (CF-IDF) method [12] calculates similarity measures using concepts from a domain ontology rather than terms or synsets. The concepts of an article are obtained using a Natural Language Processing (NLP) engine. For every document, the resulting concepts are then stored in a vector and these vectors can be used to calculate the CF-IDF measure. Similar to TF-IDF and SF-IDF, scores for concept c are computed as follows:

$$\text{cf-idf}(c,d) = \text{cf}(c,d) \times \text{idf}(c,d) \ , \tag{6}$$

where frequencies and inverse document frequencies are defined as $\text{cf}(c,d) = n_{c,d} / \sum_k n_{k,d}$ and $\text{idf}(c,d) = \log |D| \, / \, |d \in D : c \in d|$, respectively. Cosine similarity computations remain unchanged for $\text{sim}_{\text{CF-IDF}}$.

2.4 Relation-Based Recommendation

Both SF-IDF and CF-IDF can be extended in such a way that also related synsets or concepts are taken into consideration. For this, the semantic lexicon and ontology can be used in order to derive related elements.

In SF-IDF+ [17], related synsets are considered to be synsets that are connected through a relation (27 unique semantic relationships, e.g., hyponymy, antonymy, synonymy, etc., exist in WordNet), and are added to the vector representation from SF-IDF. For each synset, scores are computed by multiplying the original SF-IDF score with a predefined weight. Weights always range between 0 and 1, as related synsets should never be more important that the synset itself. In Eq. 7, it is shown how the related synsets are added to the vector:

$$\text{sf-idf+}(s, d, r) = \text{sf}(s, d) \times \text{idf}(s, d) \times w_r \ , \tag{7}$$

where d is the news item, s and r are the original and related synsets, respectively, and w_r is the weight corresponding to the semantic relationship type the related synset has with s.

The same rules apply also for CF-IDF in its extended form (CF-IDF+ [9]). Related concepts are retrieved by taking into account related ontology concepts by three possible relationships, as a concept can have superclasses, subclasses, and domain-specific related concepts. Similarly, the CF-IDF+ value for a concept c and its related concept r in document d is computed as follows:

$$\text{cf-idf+}(c, d, r) = \text{cf}(c, d) \times \text{idf}(c, d) \times w_r \ , \tag{8}$$

where w_r represents the weight assigned to one of the three previously mentioned relationships present between c and r. If multiple weights are computed for the same concept (or synset), only the highest value is retained in the extended vector representation. The extended vector representation is used for computing the similarity between the user profile and the unread news items using the cosine similarity measure.

2.5 Named Entity-Based Recommendation

In recent endeavours, we additionally tried combining SF-IDF+ with named entities from Bing in Bing-SF-IDF+ [7], which showed promising results. Here, named entities that are not covered by the synsets from a semantic lexicon were still taken into account by consulting the Bing search engine and computing similarities based on page counts.

Computations are based on a weighted average of SF-IDF+ and Bing similarity scores, where the latter is computed using a co-occurrence similarity measure. Similarly, we would like to investigate the merits of the application of Bing named entities to (related) concepts.

Table 1. Average F_1-measures for the recommenders

Recommender	μ	
TF-IDF	0.449	[7]
SF-IDF	0.468	[6]
CF-IDF	0.485	[12]
SF-IDF+	0.548	[17]
CF-IDF+	0.571	[9]
Bing-SF-IDF+	0.579	[7]

2.6 Performance

The discussed methods have been thoroughly tested throughout the years. Some have served as a reference, and have been tested multiple times under different conditions. Overall, the performance of the methods (in terms of F_1) is as described in Table 1. In general, we can say that concept-based methods outperform synset-based methods and the baseline TF-IDF method. Moreover, relation-based recommenders show a performance improvement over their regular counterparts. Including named entities boosts recommendation quality even more.

3 Framework

We improve the existing methods by introducing a two-step procedure, in which we compute a Bing similarity score using point-wise mutual information similarities for Bing named entities, and a CF-IDF+ similarity score using cosine similarities based on concepts and related concepts. Bing-CF-IDF+ scores are computed as a weighted average between Bing and CF-IDF+ scores. Our approach makes use of a user profile, which can be constructed manually by a user by selecting either interesting concepts or interesting news items from which concepts and named entities can be extracted. Incoming news messages are processed similarly, while eliminating named entities that are already covered by the domain ontology.

3.1 Bing

Concept-based recommendation methods only make use of named entities that are included in the domain ontology. However, there could be many more named entities in a single article, that – if they would not be taken into consideration – could skew the entire similarity analysis. Therefore, the Bing similarity measure [7] takes all these named entities into account.

Let U and R be sets of named entities in an unread news item and the user profile:

$$U = \{u_1, u_2, \ldots, u_k\} \ , \quad (9)$$

$$R = \{r_1, r_2, \ldots, r_l\} \ , \quad (10)$$

where u_i is a named entity in unread item U, r_j a named entity in user profile R, and k and l are the number of named entities in the unread item and the user profile, respectively. Now let us define the set of possible named entity pairs from the unread news item and the user profile by taking their cartesian product:

$$V = U \times R = (\langle u_1, r_1 \rangle, ..., \langle u_k, r_l \rangle) \ . \tag{11}$$

Subsequently, we compute the point-wise mutual information co-occurrence similarity measure as proposed by [4]. We search the named entities in a pair both separately and together in Bing to construct page counts. A page count is defined as the number of Web pages that are found by Bing. For every pair the similarity is computed as the difference between the actual and the expected joint probability. The similarity measure for a pair is defined as:

$$\mathrm{sim}_{\mathrm{PMI}}(u, r) = \log \frac{\frac{c(u,r)}{N}}{\frac{c(u)}{N} \times \frac{c(r)}{N}} \ , \tag{12}$$

where $c(u, r)$ is the Bing page count for pair (u, r), $c(u)$ and $c(r)$ the page counts for named entities u and r, and N the total number of Web pages that can be found by Bing. N is estimated to be around 15 billion. The Bing similarity measure $\mathrm{sim}_{\mathrm{Bing}}$ is then defined as:

$$\mathrm{sim}_{\mathrm{Bing}}(d_u, d_r) = \frac{\sum_{(u,r) \in V} \mathrm{sim}_{\mathrm{PMI}}(u, r)}{|V|} \ . \tag{13}$$

3.2 CF-IDF+

The CF-IDF+ method makes use of concepts and related concepts. A concept can be a class, which can have superclasses and subclasses. It can also be an instance and refer to other concepts using domain relationships. The relations between concepts contain valuable information about a news article and can therefore increase recommendation accuracy. Similar to the CF-IDF method, the CF-IDF+ method stores the concepts and related concepts of a news item into a vector. For every concept c, a new set of concepts is defined which contains all related concepts:

$$C(c) = \{c\} \bigcup_{r \in R(c)} r(c) \ , \tag{14}$$

where c is a concept in the news item, $r(c)$ are concepts related to concept c by relation r, and $R(c)$ is the set of relationships of concept c.

The extended sets of concepts for all news items are now unified to one large set U:

$$U = \{C(u_1), C(u_2), \ldots, C(u_m)\} \ , \tag{15}$$

where $C(u_m)$ is the m^{th} extended concept in the set of extended concepts of the news item. CF-IDF+ scores and their cosine similarities can be computed as introduced earlier using Eqs. 8 and 4. If these scores exceed a predetermined cut-off value, the news item is recommended to the user.

3.3 Bing-CF-IDF+

We can now calculate the Bing and the CF-IDF+ similarity measures between every unread news item and the user profile. Bing-CF-IDF+ is a weighed combination of the Bing and the CF-IDF+ similarity measures. For inter-comparability of the similarities, $\text{sim}_{\text{CF-IDF+}}$ and $\text{sim}_{\text{Bing}}(d_u, d_r)$ are normalized using a min-max scaling between 0 and 1:

$$\overline{\text{sim}}_{\text{CF-IDF+}}(d_u, d_r) = \frac{\text{sim}_{\text{CF-IDF+}}(d_u, d_r) - \min_u \text{sim}_{\text{CF-IDF+}}(d_u, d_r)}{\max_u \text{sim}_{\text{CF-IDF+}}(d_u, d_r) - \min_u \text{sim}_{\text{CF-IDF+}}(d_u, d_r)} \ , \quad (16)$$

$$\overline{\text{sim}}_{\text{Bing}}(d_u, d_r) = \frac{\text{sim}_{\text{Bing}}(d_u, d_r) - \min_u \text{sim}_{\text{Bing}}(d_u, d_r)}{\max_u \text{sim}_{\text{Bing}}(d_u, d_r) - \min_u \text{sim}_{\text{Bing}}(d_u, d_r)} \ , \quad (17)$$

where d_u and d_r are an unread news item and the user profile, respectively. The Bing-CF-IDF+ similarity measure $\text{sim}_{\text{Bing-CF-IDF+}}(d_u, d_r)$ is computed by taking a weighted average over both similarities:

$$\text{sim}_{\text{Bing-CF-IDF+}}(d_u, d_r) = \alpha \times \overline{\text{sim}}_{\text{Bing}} + (1 - \alpha) \times \overline{\text{sim}}_{\text{CF-IDF+}} \ , \quad (18)$$

where α is optimized using a grid search optimization on the training set. Again a news item is recommended when the similarity measures exceeds the predefined threshold value t. Please note that only named entities that are not found as denoting concepts are considered here.

3.4 Implementation

The Bing-CF-IDF+ recommender is implemented in the Hermes framework [11], which is a Java-based personalizing news service using Semantic Web technologies. Hermes ingests user queries and RSS feeds of news items, and supports multiple recommendation methods using an internal knowledge base for storing ontological concepts. Hermes provides recommendations based on user profiles that are constructed based on browsing behaviour. Hermes contains several plug-ins that extend the basic functionality. The Athena plug-in classifies and recommends news items using an internal OWL domain ontology [13]. Next to several concept-based recommender methods, Athena supports an additional profile builder, where a user is allowed to select relevant topics in a visual knowledge graph. The Ceryx plug-in [6] is an extension to Athena. Just like Athena, Ceryx works with a user profile. However, the algorithm to find related news items is slightly different. Besides classifying terms and concepts, Ceryx also determines the senses of words. Therefore, Ceryx is capable of handling recommender methods like SF-IDF+ and CF-IDF+. The Bing-CF-IDF+ recommender is also written for Ceryx.

4 Evaluation

In order to evaluate the performance of the newly proposed Bing-CF-IDF+ method, we compare it with its concept-based alternatives, i.e., CF-IDF and

CF-IDF+, as well as the TF-IDF baseline. This section starts by elaborating on the experimental setup regarding data and performance measures. Next, the weights of the semantic relationships and their properties are discussed. Last, performance measures are compared.

4.1 Experimental Setup

In our experiments, we make use of an annotated data set of 100 news items from a Reuters news feed with news on technology companies. Domain experts related news messages to given subjects with an inter-annotator agreement of at least two thirds. The subjects (i.e., the user profiles) are listed in Table 2, accompanied by their inter-annotator agreements (IAA). The reported amounts of interesting (I+) and non-interesting (I−) news items are as determined by the experts.

The data set is randomly split into a training set and a test set, with respectively 60% and 40% of the data. First, a user profile is created by adding the interesting news items from the training set. The optimal weights are determined by using a validation set which is created by splitting the training set into two equally-sized sets, i.e., a validation set and a training set. We end up having three different sets: a validation set (30%), a training set (30%), and a test set (40%). The validation set and the test set are considered to consist of 'unread' news items. The validation set can now be used to determine the optimal weights, needed to calculate performance measures by using the test set later on.

As discussed before, the CF-IDF+ recommender computes similarity measures for every unread news item. In case this similarity measure exceeds a certain cut-off value, the unread news item is recommended to the user. The results of the recommenders can be classified for news items as either true positive (TP), false positive (FP), true negative (TN), or false negative (FN). A selection of information retrieval metrics can be deduced from this confusion matrix: precision, recall (sensitivity), and specificity. Additionally, we can deduce the F_1-scores (i.e., the harmonic mean of precision and recall) and ROC-curve (i.e., the True Positive Rate or sensitivity plotted against the False Positive Rate

Table 2. Amount of interesting (I+) and non-interesting (I−) news items, and the inter-annotator agreement (IAA)

Topic	I+	I−	IAA
Asia or its countries	21	79	99%
Financial markets	24	76	72%
Google or its rivals	26	74	97%
Web services	26	74	94%
Microsoft or its rivals	29	71	98%
National economies	33	67	90%
Technology	29	71	87%
United States	45	55	85%

Table 3. Mean and variance for the parameters of the Bing-CF-IDF+ recommender

	w_{super}	w_{sub}	w_{rel}	α
μ	0.426	0.384	0.523	0.170
σ^2	0.135	0.120	0.103	0.020

or 1 − specificity) from these measures. Last, we compute the Kappa statistic [8] to verify whether the classification power is higher than a random guess. The parameters for semantic relationships are optimized individually through an incremental procedure, optimizing the global F_1-scores. Additionally, the α parameter that determines the weight of the Bing and CF-IDF+ parts is optimized similarly.

4.2 Optimizing parameters

For each cut-off value, with an increment of 0.01, we optimize the weight parameters for superclass, subclass, and domain relationships, and the α that balances the two similarity measures. The results are displayed in Table 3, where the mean and variance of each of these parameters are computed.

On average, the Bing similarity measure has a lower weight than the CF-IDF+ measure, indicating that input from Bing has a lower impact on our recommender than the semantic relationships. This can be explained by the fact that concepts contain more informational value than named entities. Moreover, 44 out of 266 identified named entities appear in our employed ontology, indicating a loss of 20% of the available named entities. Nonetheless, α is greater than zero, and thus there is a use to employing named entities from Bing in the recommendation method. As for the semantic relationships, on average, concepts retrieved through domain relationships seem (w_{rel}) to be more important than sub- and superclasses (w_{sub} and w_{super}, respectively), and concepts retrieved through superclasses are more important than those deduced from subclass relations. This corresponds to the results of [9], and match our expectations, as superclasses give more general information about the topic of interest whereas subclasses risk to be too specific.

4.3 Experimental Results

Now that the optimal values of the parameters are determined for each cut-off value, we can compute the global precision, recall, and F_1-measures. Table 4

Table 4. Average F_1-measures for the recommenders

Recommender	μ
TF-IDF	0.449
CF-IDF	0.485
CF-IDF+	0.571
Bing-CF-IDF+	0.609

displays the mean F_1-scores for each recommender, underlining that Bing-CF-IDF+ outperforms the other recommenders. In fact, the more complex the recommender, the better the average performance. As shown in Table 5, all improvements are significant, except for CF-IDF over TF-IDF.

Our observations are also supported by Fig. 1a. From the plot, it is evident that, throughout the range of cut-off values, Bing-CF-IDF+ outperforms the other recommenders consistently. TF-IDF is more performant for lower cut-off values (i.e., higher expected recall and lower expected precision) than CF-IDF and CF-IDF+. Due to the nature of CF-IDF variants, this is an expected outcome, because when using concepts rather than terms (or named entities for that matter), we enforce a much more restricted approach with a very limited amount of tokens (concepts) to match on.

This is also depicted in Figs. 1b and 1c. These figures also show that, while recall for Bing-CF-IDF+ and CF-IDF+ is very similar, the precision of Bing-CF-IDF+ clearly improves over CF-IDF+. Recall for CF-IDF (and TF-IDF) is much lower. Therefore, it seems that the addition of semantic relations improves recall, and the additional inclusion of Bing named entities improves precision, without making concessions to the recall of CF-IDF.

Next we evaluate the Receiver Operating Characteristic (ROC) curves for the Bing-CF-IDF+, CF-IDF+, CF-IDF, and TF-IDF recommenders. The ROC curve in Fig. 2 shows that the Bing-CF-IDF+ and CF-IDF+ outperform CF-IDF and TF-IDF for low False Positive Rates. This indicates that recall (True Positive Rate) is higher for (Bing-)CF-IDF+ in more difficult situations against a handful of false positives, i.e., a higher precision. However, in the grand scale of things, the areas under the curve differ only slightly between the recommenders (value is approximately 0.85). This is in line with the higher precision and lower recall of Bing-CF-IDF+ when compared to TF-IDF.

Last, we compute the Kappa statistic to measure whether the proposed classifications made by the recommender are better than classification made by a random guess. Higher values indicate more classification power, and are preferred. In Fig. 3, the results of the Kappa statistic can be found for varying cut-off values. The plot shows that overall, the Kappa statistic of the Bing-CF-IDF+ recommender is higher than the Kappa statistic of the other three recommenders. Only for a cut-off value of 0.25, the statistics of the Bing-CF-IDF+ and the TF-IDF are similar, and for cut-off value 0.70 the statistics of the Bing-CF-IDF+ and the CF-IDF+ are alike. Because the Bing-CF-IDF+ recom-

Table 5. Recommenders in columns outperforming recommenders in rows with respect to F_1 (p values where significance is $<$ 5e-02)

Recommender	TF-IDF	CF-IDF	CF-IDF+	Bing-CF-IDF+
TF-IDF		7.046e-02	1.398e-07	5.836e-11
CF-IDF			6.525e-05	6.305e-08
CF-IDF+				3.361e-02
Bing-CF-IDF+				

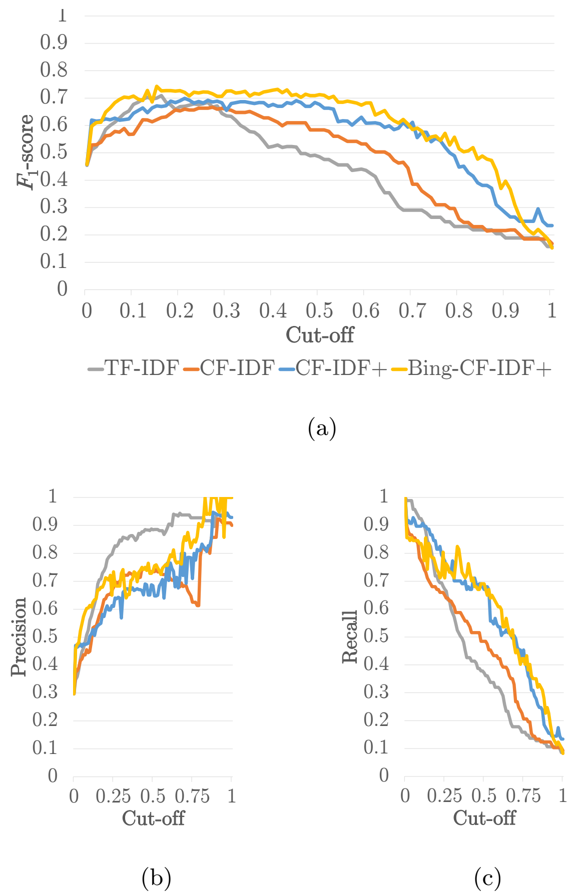

Fig. 1. Global precision, recall, and F_1 scores for the recommenders

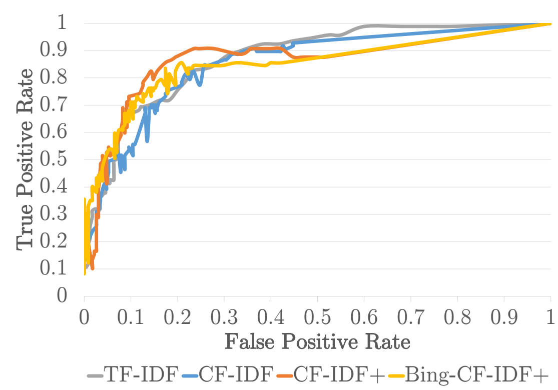

Fig. 2. ROC curve for the recommenders

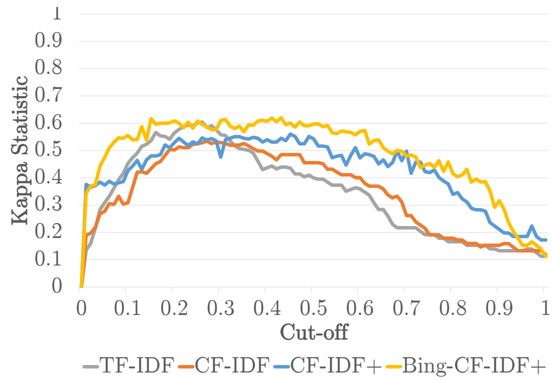

Fig. 3. Kappa statistics for the recommenders

mender clearly has higher values for the Kappa statistic over all cut-off values, we can state that overall, the Bing-CF-IDF+ has more classification power than the CF-IDF+, CF-IDF, and TF-IDF recommenders.

5 Conclusion

In previous work, several new recommendation methods have been proposed. The traditional term-based TF-IDF was improved by methods like SF-IDF and CF-IDF, which take into account synsets from a semantic lexicon and concepts from a domain ontology, respectively. The CF-IDF+ similarity measure also matches news items based on related concepts like sub- and superclasses. However, named entities are not fully covered in recommendations whenever they are omitted in the domain ontology. Therefore, we have introduced the Bing-CF-IDF+ similarity measure, which is a two-step procedure that extends the CF-IDF+ similarity measure with Bing Web search similarity scores for named entities.

In order to evaluate the performance of the new Bing-CF-IDF+ recommender, we have optimized the weights for the semantic relationships between the concepts and for the Bing and CF-IDF+ recommenders themselves. These parameters are optimized using a grid search for both the semantic relationships and the concept-based and named entity-based recommenders, while maximizing the global F_1-measure per cut-off value, i.e., the minimum score for a news item to be recommended. We have tested the performance of Bing-CF-IDF+ against existing recommenders on 100 financial news items and 8 user profiles. In our evaluation, we have shown that the Bing-CF-IDF+ similarity measure outperforms TF-IDF, CF-IDF, and CF-IDF+ in terms of the F_1 measure and the Kappa statistic.

We envision various directions for future work. Parameter optimization has been performed using an incremental grid search. This could improved by applying more advanced optimization strategies, such as genetic algorithms. Moreover, we would like to investigate a larger collection of relationships. Now, we have considered the direct super- and subclasses, but hypothetically, non-direct super- and subclasses of concepts could be valuable as well. Last, a more thorough and powerful evaluation based on a larger set of news items would further underline the strong performance of Bing-CF-IDF+.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering **17**(6), 734–749 (2005)
2. Banerjee, S., Pedersen, T.: An adapted Lesk algorithm for word sense disambiguation using WordNet. In: Gelbukh, A.F. (ed.) 4th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2002). Lecture Notes in Computer Science, vol. 2276, pp. 136–145. Springer (2002)
3. Bing: Bing API 2.0. Whitepaper. From: `http://www.bing.com/developers/s/APIBasics.html` (2018)
4. Bouma, G.: Normalized (pointwise) mutual information in collocation extraction. In: Chiarcos, C., de Castilho, R.E., Stede, M. (eds.) Biennial GSCL Conference 2009 (GSCL 2009). pp. 31–40. Gunter Narr Verlag Tübingen (2009)
5. Burke, R.: Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction **12**(4), 331–370 (2002)
6. Capelle, M., Moerland, M., Frasincar, F., Hogenboom, F.: Semantics-based news recommendation. In: Akerkar, R., Bădică, C., Dan Burdescu, D. (eds.) 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS 2012). ACM (2012)
7. Capelle, M., Moerland, M., Hogenboom, F., Frasincar, F., Vandic, D.: Bing-SF-IDF+: A hybrid semantics-driven news recommender. In: Wainwright, R.L., Corchado, J.M., Bechini, A., Hong, J. (eds.) 30th Symposium on Applied Computing (SAC 2015), Web Technologies Track. pp. 732–739. ACM (2015)
8. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**(1), 37–46 (1960)
9. de Koning, E., Hogenboom, F., Frasincar, F.: News recommendation with CF-IDF+. In: Krogstie, J., Reijers, H.A. (eds.) 30th International Conference on Advanced Information Systems Engineering (CAiSE 2018). Lecture Notes in Computer Science, vol. 10816, pp. 170–184. Springer (2018)
10. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)
11. Frasincar, F., Borsje, J., Levering, L.: A Semantic Web-based approach for building personalized news services. International Journal of E-Business Research **5**(3), 35–53 (2009)
12. Goossen, F., IJntema, W., Frasincar, F., Hogenboom, F., Kaymak, U.: News personalization using the CF-IDF semantic recommender. In: Akerkar, R. (ed.) International Conference on Web Intelligence, Mining and Semantics (WIMS 2011). ACM (2011)
13. IJntema, W., Goossen, F., Frasincar, F., Hogenboom, F.: Ontology-based news recommendation. In: Daniel, F., Delcambre, L.M.L., Fotouhi, F., Garrigós, I., Guerrini, G., Mazón, J.N., Mesiti, M., Müller-Feuerstein, S., Trujillo, J., Truta, T.M., Volz, B., Waller, E., Xiong, L., Zimányi, E. (eds.) International Workshop on Business intelligencE and the WEB (BEWEB 2010) at 13th International Conference

on Extending Database Technology and Thirteenth International Conference on Database Theory (EDBT/ICDT 2010). ACM (2010)
14. Jannach, D., Resnick, P., Tuzhilin, A., Zanker, M.: Recommender systems - beyond matrix completion. Communications of the ACM **59**(11), 94–102 (2016)
15. Jensen, A.S., Boss, N.S.: Textual Similarity: Comparing Texts in Order to Discover How Closely They Discuss the Same Topics. Bachelor's Thesis, Technical University of Denmark (2008)
16. Jones, K.S.: A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation **28**(1), 11–21 (1972)
17. Moerland, M., Hogenboom, F., Capelle, M., Frasincar, F.: Semantics-based news recommendation with SF-IDF+. In: Camacho, D., Akerkar, R., Rodríguez-Moreno, M.D. (eds.) 3rd International Conference on Web Intelligence, Mining and Semantics (WIMS 2013). ACM (2013)
18. Robal, T., Haav, H., Kalja, A.: Making Web users' domain models explicit by applying ontologies. In: Hainaut, J., Rundensteiner, E.A., Kirchberg, M., Bertolotto, M., Brochhausen, M., Chen, Y.P., Cherfi, S.S., Doerr, M., Han, H., Hartmann, S., Parsons, J., Poels, G., Rolland, C., Trujillo, J., Yu, E.S.K., Zimányi, E. (eds.) Advances in Conceptual Modeling - Foundations and Applications, ER 2007 Workshops CMLSA, FP-UML, ONISW, QoIS, RIGiM, SeCoGIS. Lecture Notes in Computer Science, vol. 4802, pp. 170–179. Springer (2007)
19. Robal, T., Kalja, A.: Conceptual Web users' actions prediction for ontology-based browsing recommendations. In: Papadopoulos, G.A., Wojtkowski, W., Wojtkowski, W.G., Wrycza, S., Zupancic, J. (eds.) 17th International Conference on Information Systems Development (ISD 2008). pp. 121–129. Springer (2010)
20. Robal, T., Kalja, A.: Applying user domain model to improve Web recommendations. In: Caplinskas, A., Dzemyda, G., Lupeikiene, A., Vasilecas, O. (eds.) Databases and Information Systems VII - Selected Papers from the Tenth International Baltic Conference (DB&IS 2012). Frontiers in Artificial Intelligence and Applications, vol. 249, pp. 118–131. IOS Press (2013)
21. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Information Processing and Management **24**(5), 513–523 (1988)
22. Sekine, S., Ranchhod, E. (eds.): Named Entities: Recognition, clasification and use. John Benjamins Publishing Company (2009)