Introducing a Weighted Ontology to Improve the Graph-based Semantic Similarity Measures

Roberto Saia, Ludovico Boratto, Salvatore Carta
Dip.to di Matematica e Informatica, Università di Cagliari
Via Ospedale 72 - 09124 Cagliari, Italy
Email: {roberto.saia, ludovico.boratto, salvatore}@unica.it

***Abstract*—The semantic similarity measures are designed to compare terms that belong to the same ontology. Many of these are based on a graph structure, such as the well-known lexical database for the English language, named WordNet, which groups the words into sets of synonyms called synsets. Each synset represents a unique vertex of the WordNet semantic graph, through which is possible to get information about the relations between the different synsets. The literature shows several ways to determine the similarity between words or sentences through WordNet (e.g., by measuring the distance among the words, by counting the number of edges between the correspondent synsets), but almost all of them do not take into account the peculiar aspects of the used dataset. In some contexts this strategy could lead toward bad results, because it considers only the relationship between vertexes of the WordNet semantic graph, without giving them a different weight based on the synsets frequency within the considered datasets. In other words, common synsets and rare synsets are valued equally. This could create problems in some applications, such as those of recommender systems, where WordNet is exploited to evaluate the semantic similarity between the textual descriptions of the items positively evaluated by the users, and the descriptions of the other ones not evaluated yet. In this context, we need to identify the user preferences as best as possible, and not taking into account the synsets frequency, we risk to not recommend certain items to the users, since the semantic similarity generated by the most common synsets present in the description of other items could prevail. This work faces this problem, by introducing a novel criterion of evaluation of the similarity between words (and sentences) that exploits the WordNet semantic graph, adding to it the weight information of the synsets. The effectiveness of the proposed strategy is verified in the recommender systems context, where the recommendations are generated on the basis of the semantic similarity between the items stored in the user profiles, and the items not evaluated yet.

***Index Terms*—Semantic Graph, Semantic Analysis, Ontology, Graph Theory, Metrics

I. Introduction

The use of the semantic measures of similarity [1] has spread over the past decades, and this is related with the coming of the so-called Semantic Web [2], as well as, more generally, with the needs to interpret the users preferences in a non-schematic mode, in order to understand the concepts connected with a text, instead of using the single terms, disjointed from the concepts that they express.

When operating through a metric, in order to determine the level of semantic similarity between concepts, it is assumed that this takes place within a specific ontology [3], [4], related with the terms used in the operating environment. The level of similarity between two or more terms, is usually performed by measuring their distance within an ontology. The main objective of these semantic operations is to provide a standard (and non supervised) approach of evaluation of the information. This evaluation is crucial in many environments, such as the commercial ones that provide forms of personalization and have to interpret the preferences of the users [5], or the medical applications that have to analyze the medical reports automatically [6].

Many approaches map the terms of an ontology exploiting a graph structure, such as WordNet[1], the widespread approach considered in this work, which is a semantic graph were each vertex represent a distinct set of synonyms called synset (i.e., a set of words that denote the same concept). The WordNet graph is a *Directed Acyclic Graph* (DAG), where each vertex v is an integer that identifies a synset, and each directed edge that connects v with w denotes that w is a hypernym of v. The literature proposes several approaches able to evaluate the semantic similarity among concepts, i.e., *Jiang and Conrath* [7], *Leacock and Chodorow* [8], *Lin* [9], *Resnik* [10], and *Wu and Palmer* [11]. Some of them exploit graph structures such as WordNet, and are based on the measure of the shortest path length between vertexes (synsets). A limit of these approaches is that they consider only the relationship between vertexes of the WordNet semantic graph, without giving them a different weight based on the synsets frequency within the considered datasets (i.e., common synsets and rare synsets are valued equally). This could create problems in some contexts, where it is important to take into account the synsets frequency, such as the recommender systems, where the semantic similarity generated by the items with most common synsets in their description could prevent the recommendation of other relevant items with rare synsets.

In this work, we present a strategy aimed to evaluate the semantic similarity between words or sentences, which introduces a novel way to define and use the ontology of synsets used to build the WordNet semantic graph. The proposed approach, instead of a DAG graph, uses a *Weighted Graph* (WG) [12], in order to introduce the weight of the synsets on the edges, which is calculated through an inverse frequency criterion. The new WordNet weighted graph gives the possibility to characterize the operative context, attributing more importance to some terms, and less to others, during the computation of the semantic similarity. There are many

[1] http://wordnet.princeton.edu/

contexts where the proposed approach would produce benefits, e.g., those where it is important to refine the ontology in accord with the specificity of the operating area. We test it in a very widespread context, that of Recommender System [13], and to perform the experiments, we adopt the real-world dataset Yahoo! Webscope R4[2], which contains a large amount of data related to users preferences expressed on the Yahoo! Movies community. Based on the occurrences of the synsets in the considered ontology, we define their weight, which we use during the evaluation of the semantic similarity through the WordNet functionalities.

Although the proposed approach can be applied to any metric of similarity based on WordNet, in this work we will consider only the *Wu and Palmer* metric [11] to evaluate the semantic similarity between terms. We made this choice because in the literature this metric (based on the path lengths between a pair of concepts) is considered to be one the most accurate in terms of measurement of semantic similarity [14], [15]. Considering that the task of a recommender system is to infer the interest of the users for the new items, based on the information stored in their profiles, we use our strategy (compared to the canonical approach) in order evaluate the semantic similarity between the description of the items that the users have already positively evaluated, and the description of the others not yet evaluated.

The contributions of our work are the following:

- introduction of a new approach able to extract and weigh a synset ontology from a specified dataset, using an *Inverse Synset Frequency* (ISF) criterion, which gives more weight to the less frequent synsets and a lower one to the most frequent ones;
- creation of a *Weighted Ontology* (WO), implemented as a weighted graph structure which reproduces the WordNet relationships between synsets, adding them the weight information;
- definition of a new *Wu and Palmer* (WP) metric able to exploit the weight information of the *Weighted Ontology based on the Inverse Synset Frequency* (WO/ISF), which we named *Weighted Wu and Palmer* (WWP);
- application of the new WWP metric in the context of a recommender system based on the semantic similarities of WordNet, comparing the results with those of a canonical approach based on the standard WP metric that does not exploit a weighted ontology of synsets.

The rest of the paper is organized as follows: the first part of our work introduces the literature related with the proposed strategy (Section II), then we define the notation and the problem definition (Section III), we continue with the implementation details (Section IV) and with an adoption of our approach in the recommender systems application domain (Section V), ending with the description of the performed experiments (Section VI) and with some concluding remarks (Section VII).

[2]http://webscope.sandbox.yahoo.com

II. Background and Related Work

This section provides background information on the four main focus areas in our research, namely the WordNet environment, the concept of ontology, the *Wu and Palmer* metric, and the weighted graph data structure.

A. WordNet Environment

For many years the item descriptions were analyzed with a word vector space model, where all the terms of each item description are processed by TF-IDF [16] and stored in a weighted vector of terms. Due to the fact that this approach based on a simple *bag of words* is not able to perform a semantic disambiguation of the words in a text, nowadays more sophisticated approaches are largely used, such as that used in this work, which exploits the functionalities offered by the WordNet environment. This one is a large lexical database of English, where *nouns*, *verbs*, *adjectives*, and *adverbs* are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Wordnet currently contains about 155,287 words, organized into 117,659 synsets, for a total of 206,941 word-sense pairs [17]. In a short, the main relation among words in WordNet is the synonymy and the synsets are unordered sets of grouped words that denote the same concept and are interchangeable in many contexts. Each synset is linked to other synsets through a small number of *conceptual relations*. Word forms with several distinct meanings are represented in as many distinct synsets, in this way each form-meaning pair in WordNet is unique (e.g., the *fly* insect and the *fly* verb belong to two distinct synsets). Most of the WordNet relations connect words that belong to the same *part-of-speech* (POS). There are four POS: *nouns*, *verbs*, *adjectives*, and *adverbs*. Both nouns and verbs are organized into precise hierarchies, defined by hypernym or *is-a* relationships. For example, the first sense of the word *radio* would have the following hypernym hierarchy, where the words at the same level are synonyms of each other: as shown in the following, some sense of *radio* is synonymous with some other senses of *radiocommunication* or *wireless*, and so on.

- **POS=*noun***
  - *radio, radiocommunication, wireless (medium for communication)*
  - *radio receiver, receiving set, radio set, radio, tuner, wireless (an electronic receiver that detects and demodulates and amplifies transmitted signals)*
  - *radio, wireless (a communication system based on broadcasting electromagnetic waves)*
- **POS=*verb***
  - *radio (transmit messages via radio waves)*

Each synset has a unique index and shares its properties, such as a gloss or dictionary definition. In the case of *nouns* and *verbs* (the organization of adjectives and adverbs is slightly different) the WordNet hierarchies are organized into several base types (25 primitive groups for the nouns and 15 for the verbs), and all primitive groups ultimately go up to an abstract root vertex. As we can imagine, the network of nouns is far deeper than that of the other *parts-of-speech*. The verbs instead present a more bushy structure, and the adjectives are distributed into many clusters, as well as the adverbs, since these last are defined in terms of the adjectives (they are derived from adjectives and thus inherit their structure).

B. Ontology

During the last years, there has been a growth on the use of the ontologies, thanks to their ability to explicitly describe the semantic information in a common way, regardless of their characteristics, providing a model that allows the interchanging among heterogeneous data. An ontology is a conceptual model that can be applied in order to describe a certain domain, defining this as a set of concepts and relations [18]. Ontologies are generally adopted to give a uniform conceptualization of the terms used in a dataset. In a recommender system context, they are used to support the different approaches of recommendation (e.g., those based on a content-based strategy). Profiles based on the same domain ontology are not affected by problems of synonymy or homonymy, and the ontologies may also be used to define a common way to describe and classify the items involved in a recommender system [19]. In order to define an ontology, several approaches were developed (readers can refer to [20] for a survey on the topic). We can identify two main categories: the first is related to the experience-based strategies, such as that proposed in [21], or that exposed in [22], which are both based on the Enterprise Model [23]; the strategies that belong to the second main category implement evolutive prototypes models, such as that presented in [24], which proposes a set of actions to perform in order to build ontologies based on their life cycle and the prototype refinement, or the strategy proposed in [25], based on an iterative approach to build the ontology.

It should be noted that, regardless of the approach adopted to define an ontology, it is necessary to identify the best way to maximize the results during its use. In the case study of this work, related to a recommender system which operates within a context of a movies seller, we introduce the concept of weight, in order to give a different value to the items, based on their rarity/ordinary of the terms that describe them in the adopted dataset.

C. Wu and Palmer Metric

The Wu and Palmer metric [11] calculates the similarity by considering the depths of two synsets (synonym sets) in the WordNet taxonomies, along with the depth of the *Least Common Subsumer*. Assuming that the *Least Common Subsumer* (**LCS**) of two concepts x and y is *the most specific concept that is an ancestor of both x and y*, where the concept tree is defined by the *is-a* relation, in Equation 1 we have that *A=depth(LCS(x,y))*, *B=length(x,LCS(x,y))*, *C=length(y,LCS(x,y))*.

$$sim_{WP}\left(x,y\right) = \frac{2 \cdot A}{B + C + (2 \cdot A)} \tag{1}$$

We can note that $B + C$ represents the path length from x and y, while A indicates the global depth of the path in the taxonomy. In the example of Figure 1, D is the parent (and also ancestor) of E, while B is an ancestor of E. B is also an ancestor of C. In this case, the LCS of C and E is B, since it is the most specific concept that is an ancestor of both C and E. Note that while A is a common subsumer of both C and E, it is not the least, since there is still a child of A (in this case it is B), which is also a common subsumer of both E and C. D is not the least common subsumer since it is not an ancestor of C.

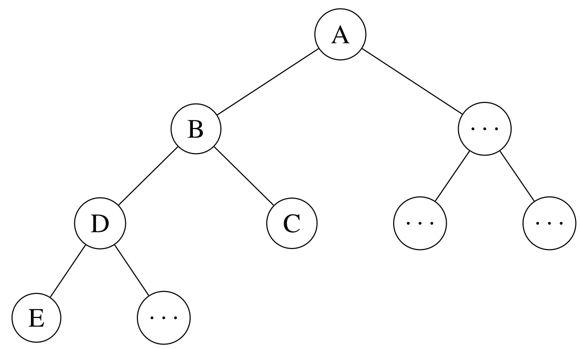

Fig. 1: WordNet Relationships Tree

In order to calculate the Wu and Palmer similarity between C and E, we first determine that the least common subsumer of C and E is B. Next, we determine that the length of the path from C to B is 1, that the length of the path from E to B is 2, and that the depth of B is 1 (distance from B to root vertex A). Now we can determine the similarity between the synsets C and E (as in Equation 2).

$$sim_{WP}\left(C,E\right) = \frac{2 \cdot 1}{2 + 1 + (2 \cdot 1)} = 0.40 \tag{2}$$

D. Weighted Graphs

In the mathematical field, and more precisely in that related to the so-called *graph theory* [26], a graph is a set of entities named vertices and connected by edges. A particular category of graphs are the weighted graphs, where each edge has an associated numerical value w, named weight. These weights are usually expressed as positive integers, and a weighted graph can be either directed or undirected (i.e., a graph in which the edges have no orientation, such as that shown in Figure 2). More formally, a weighted graph G is defined by the triple $G = (V, E, W)$, where V is a set of vertices, E is a set of edges $\{v, w\}$ (where $v, w \in V$), W is a map from edges to numbers that depends on the operative context, and should be a positive integer number.

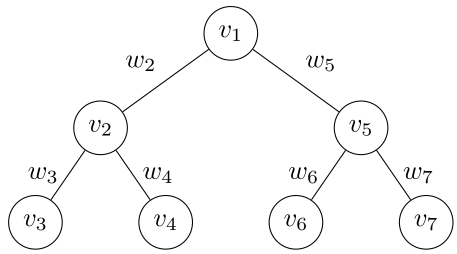

Fig. 2: Weighted Graph Structure

In their typical application, the weights associated with the edges are often related to the concept of *cost*, e.g., in an application of network routing, this cost could be the length of a route, or in a context of an electric power distribution, it could be the cost required to deliver energy in a certain place.

III. Notation and Problem Definition

For clarity, the mathematical notation used in this work, and the terms of the problem that we face through our novel strategy, are both recalled in the following:

A. Notation

We are given a set of users $U = \{u_1, \dots, u_N\}$, a set of items $I = \{i_1, \dots, i_M\}$, and a set V of values used to express the user preferences (e.g., $V = [1, 5]$ or $V = \{like, dislike\}$). The set of all possible preferences expressed by the users is a ternary relation $P \subseteq U \times I \times V$. We denote as $P_+ \subseteq P$ the subset of preferences with a positive value (i.e., $P_+ = \{(u, i, v) \in P | v \geq \overline{v} \vee v = like\}$), where $\overline{v}$ indicates the mean value (in the previous example, $\overline{v} = 3$), and in the same way we denote as $P_- \subseteq P$ the subset of preferences with a negative value (i.e., $v < 3$).

Moreover, we denote as $I_+ = \{i \in I | \exists (u, i, v) \in P_+\}$ the set of items for which there is a positive preferences, and as $I_- = \{i \in I | \exists (u, i, v) \in P_-\}$ the set of items for which there is a negative preferences. We also denote as $I_u = \{i \in I | \exists (u, i, v) \in P_+ \wedge u \in U\}$ the set of items a user u likes (user profile).

Let $BoW = \{t_1, \dots, t_W\}$ be the bag of words used to describe the items in I; we define as $S = \{s_1, \dots, s_W\}$ the set of synsets associated to BoW (that is, for each term used to describe an item, we consider the associated synsets), and as sd_i the semantic description of i. The set of semantic descriptions is denoted as $D = \{sd_1, \dots, sd_M\}$ (note that we have a semantic description for each item, so $|D| = |I|$). The approach used to extract sd_i from d_i is described in detail in Section IV. We also denote as $R_u = \{u \in U \wedge R \subseteq I\}$ the set of items i that recommend to a user u.

B. Problem definition

Given a set of items $I_u = \{i \in I | \exists (u, i, v) \in P_+ \wedge u \in U\}$ related to a profile of a user u (positively evaluated by her/him), and a set D_u related with the semantic descriptions of these items, our first goal is to evaluate the semantic similarity between the set I_u and each of the other items in the dataset not evaluated by the user u. The main objective of our work is to define a function $f : D \times D \to [0, 1]$ that calculates the semantic similarity between two items by considering their semantic descriptions.

In order to validate or proposal, we are going to adopt the proposed metric in a content-based recommender system, by defining a function $g : I_u \to R_u$ that, given a user profile $I_u = \{i \in I | \exists (u, i, v) \in P_+ \wedge u \in U\}$, returns a set of items $R_u = \{u \in U \wedge R \subseteq I\}$ to recommend to u, using the proposed weighted ontology, improving the accuracy of the canonical approach based on a non-weighted WordNet ontology.

IV. WWP: a Weighted Graph-based Similarity Metric

In this section we present our weighted semantic similarity metric. The steps performed by the algorithm that implements our proposal are the following:

1) **Text Preprocessing**: processing of the textual information (description, title, etc.) present in all the items, in order to remove the useless elements for the subsequent operation of synset retrieving;
2) **Weighted Ontology Definition**: creation of a Weighted Ontology (WO) built by using the distinct synsets extracted from the textual description of all the items in the dataset, after the text has been preprocessed, according to the operations described in Section IV-A. A weight w that expresses the importance of each synset of the ontology in the dataset is assigned to it: the rarer it is, the higher the assigned weight will be, with $w \in [0, 1]$.
3) **Weighted Graph Construction**: definition of a *WordNet Weighted Graph* (WWG), where, unlike WordNet, each synset (vertex) is mapped with a weighted edge calculated following an *Inverse Synset Frequency* (ISF) criterion, which gives more weight to the less frequent synsets, and a lower one to the most frequent synsets;
4) **Weighted Wu and Palmer Metric Formalization**: definition of a new metric named *Weighted Wu and Palmer* (WWP), based on the canonical *Wu and Palmer* (WP) approach, but able to exploit the information of weight reported in the WWG graph previously defined;

In the following, we will describe in detail how each step works.

A. Text Preprocessing

Before extracting the WordNet synsets from the text that describes each item, we need to follow several preprocessing steps. The first step is to detect the correct *Part-Of-Speech* (POS) for each word in the text; in order to perform this task, we have used the *Stanford Log-linear Part-Of-Speech Tagger* [27]. In the second step we remove punctuation marks and *stop-words*, which represent noise in the semantic analysis. Several *stop-words* lists can be found in the Internet, and in this work we have used a list of 429 *stop-words* made available with the *Onix Text Retrieval Toolkit*[3]. In the third step, after we have determined the lemma of each word using the Java API

[3] http://www.lextek.com/manuals/onix/stopwords.html

implementation for WordNet Searching JAWS[4], we perform the so-called word sense disambiguation, a process where the correct sense of each word is determined, which permits us to individuate the appropriate synset in a precise way. The best sense of each word in a sentence was found using the Java implementation of the adapted Lesk algorithm provided by the *Denmark Technical University* similarity application [28]. All the collected synsets form the set $S = \{s_1, \ldots, s_W\}$ defined in Section II. The output of this step is the semantic disambiguation of the textual description of each item $i \in I$.

B. Weighted Ontology Definition

Based on the previous stage of preprocessing of the text descriptions of the items, we create an ontology of distinct synsets, mapping each of them with a value that reports how many times the synset is present in the used dataset.

C. Weighted Graph Construction

As we can see, comparing the Fig. 1, which shows the canonical organization of the synsets into the WordNet structure, with the Fig. 2, which instead shows the organization of a weighted graph, the main difference is the absence of the weight information on the edges of the WordNet structure. We enrich the WordNet structure by introducing the weights in its edges before each vertex, which represent the weight of synsets.

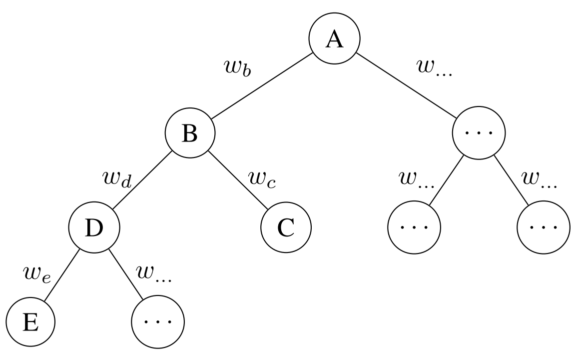

Fig. 3: Weighted WordNet Structure

It should be noted that the values w on the edges are the weights of the synsets (i.e., the weight w_e is the weight of the synset E). In Formula 3, used to calculate these weights, $num(s)$ represents the occurrences of the synset s in the entire dataset ds, while $max(ds)$ is the maximum value of synset occurrences measured in this one. As we can observe, the weight is inversely proportional to the frequency of synsets in the dataset

$$w = \frac{num(s)}{max(ds)} \tag{3}$$

[4]http://lyle.smu.edu/ tspell/jaws/index.html

D. Weighted Wu and Palmer Metric Formalization

We exploit the new information of weight introduced in the WordNet graph to modify the canonical formalization of the *Wu and Palmer* metric, into a new metric named *Weighted Wu and Palmer*, which gives more weight to the less frequent synsets, and a lower one to the most frequent ones (according to the *Inverse Synset Frequency* criterion previously enunciated). The new metric is shown in Equation 4, where we calculate the semantic similarity between the synset s_1 and the synset s_2, adopting the same notation introduced in Section II-C.

$$\begin{aligned} & sim_{WWP}(s_1, s_2) = \tfrac{2 \cdot A}{B \cdot w_1 + C \cdot w_2 + (2 \cdot A)} \\ & with: \\ & w_1 = \tfrac{num(s_1)}{max(ds)},\ w_2 = \tfrac{num(s_2)}{max(ds)} \end{aligned} \tag{4}$$

It should be noted that the values w_1 and w_2 in Equation 4 are the weights of the two synsets s_1 and s_2, i.e., the values reported in the edges in input to the vertices of the synsets, as shown in Section IV-C. Replicating the same example made in Equation 2, adopting the new WWP metric, and the values $max(ds) = 500$, $num(s_1) = 200$, and $num(s_2) = 30$, we obtain the result reported in the Equation 5.

$$sim_{WWP}(C, E) = \frac{2 \cdot 1}{2 \cdot \frac{200}{500} + 1 \cdot \frac{30}{500} + (2 \cdot 1)} = 0.69 \tag{5}$$

The result is different (0.69 instead of 0.40), because WWP takes in account the weight of the two synsets involved in the computation (the more rare are the synsets, the lower is their weight, and higher is the similarity value), and this produces a substantial change in the ranking of the items, made during the recommendation process (i.e., a different performance of the recommender system).

V. Benchmark Domain of Application: Semantic Item Recommendation

A possible application scenario that can be used as a benchmark to evaluate our similarity metric is a content-based recommender system. These systems recommend items to a user if their content is similar to those that she/he previously evaluated [5]. Therefore, we measure the semantic similarity according to the canonical *Wu and Palmer* metric, and according to the proposed *Weighted Wu and Palmer* new metric, producing a set of recommendations. For each approach, we sort the not evaluated items by their similarity with the user profile, and recommend to the user a subset of those with the highest values of similarity. The last step consists in comparing the performance of the two different approaches.

VI. Experiments

A. Experimental Setup

To conduct the experiments we adopted the Java language, with the support of Java API implementation for WordNet

Searching (JAWS) to perform the semantic analysis. The real-world dataset used during the experiments is the Yahoo! Webscope Movie (R4)[5] dataset, which represents a quite standard benchmark in the context of the recommender systems. The performed experiments want to answer the following research question: *is our approach able to improve the recommendation process, with respect to a canonical approach based on a WordNet metric that does not take in account the number of synset occurrences in the dataset, i.e., that does not exploit a weighted ontology of synsets?* In order to provide an answer, the experiments are organized as follows:

- in the first experiment, shown in Section VI-E1, we analyze the synset occurrences in the dataset. The aim is the detection of a of the maximum number of occurrences for a synset in the entire dataset, which is used by our metric to weigh the edges;
- in the second experiment, presented in Section VI-E2, we measure the ability of our approach to produce the correct recommendations for a user, thus the capacity to rank the items that a user has not evaluated yet, according to her/his user profile;
- in the last experiment, whose results are shown in Section VI-E3, we want study the relation between the number of user ratings in the test set, and the average value of the Jaccard Index, calculated in the corresponding range of values.

We use the *Jaccard index* metric in order to evaluate the recommendations generated through our strategy based on a weighted ontology and on the new WWP metric, and by the canonical strategy based on a non-weighted ontology and on the standard *Wu and Palmer* metric. We compare these results with the data present in the test set provided by the Yahoo! Webscope Movie (R4) dataset.

B. Datasets and Data Preprocessing

Yahoo! Webscope (R4). This dataset contains a large amount of data related to users preferences expressed by the Yahoo! Movies community that are rated on the base of two different scales, from 1 to 13 and from 1 to 5 (we use the latter). The training data is composed by $7,642$ users ($|U|$), $11,915$ movies/items ($|I|$), and $211,231$ ratings ($|R|$). The average user rating ($\overline{R_u} = \frac{\sum_u \overline{r_u}}{|U|}$, *macro-averaged*) is 3.70 and the average item rating (*macro-averaged*) is 3.58. The average number of ratings per user is 27.64 and the average number of ratings per item is 17.73. All users have rated at least 10 items and all items are rated by at least one user. The density ratio ($\delta = \frac{|R|}{|U|*|I|}$) is 0.0023, meaning that only 0.23% of entries in the user-item matrix are filled. The test data is composed by $2,309$ users, $2,380$ items, and $10,136$ ratings. There are no test users/items that do not also appear in the training data. The average number of ratings/user is 4.39 and the average number of ratings/item is 4.26. All users have rated at least one item and all items have been rated by at least one user.

[5]http://www.cs.umd.edu/~mount/pubs.html

C. Ontology Definition

Following the operation described in the Section IV-B, we obtain an ontology composed by $20,698$ synsets, with a maximum occurrences for synset of $2,939$, and an average value of 21.56 occurrences for synset. The result is a *Weighted Ontology* (WO) that in the next step allows us to define a weighed version of the WordNet graph.

D. Metric

In this section, we present the metric used to evaluate our approach.

1) Jaccard Index: The performance measure used during the experiments is the *Jaccard* index. We chose to adopt this metric, because the most common metrics usually adopted in these contexts, the *Recall* and *Precision* metrics (show in Equation (6)), coincide with the Jaccard metric, as in our experiments the number of predicted items $|P|$ coincides with that of the real ones $|R|$ (we generate and test the same number of items).

$$recall(P,R) = \frac{|R \cap P|}{|R|} \quad precision(P,R) = \frac{|R \cap P|}{|P|} \tag{6}$$

It is showed in the Equation (7), where as mentioned earlier, P denotes the set of predicted items (those recommended by the used approach), and R the set of items in the test set (i.e., the real preferences expressed by the users).

$$Jaccard(P,R) = \frac{|R \cap P|}{|R \cup P|} \tag{7}$$

E. Experimental Results

Here, we report the results of the experiments presented in the *Experimental Setup* (Section VI-A).

1) **Synset Occurrences:** Fig. 4 reports the synset occurrences in the used dataset. We need this information in order to individuate the maximum number of occurrences, since it represents $max(ds)$, one of the parameters to use in the synset weight calculation, as described in the Section IV-D. As we can see in Fig. 4, in our dataset $max(ds) = 2939$.

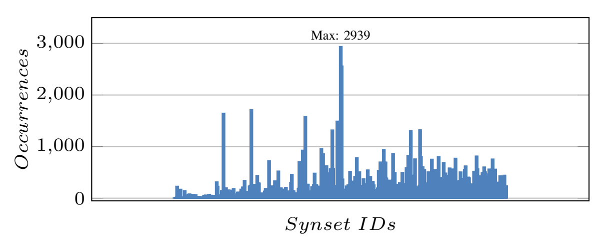

Fig. 4: *Synsets Occurrences*

A secondary but important aspect, which emerges by the graph in Fig. 4, is that big differences in the occurrence values exist. This evidence supports the idea behind this work, i.e., that a different weight to each synset in the adopted ontology should be assigned, in order to better interpret the user preferences, distinguishing the common concepts (less characterizing) from the rare ones (thus more characterizing).

*2) **Rating Prediction:*** In the second experiment we want to evaluate the ability of the proposed approach to detect the user preferences. We proceed by dividing the user preferences in the test set in two sets, I_u^+ and I_u^-, respectively containing the items positively evaluated, and those negatively evaluated by a user $u \in U$. For each user $u \in U$ present in the test set, we compared the set P, consisting of the first n recommendations of items, generated by our approach (the n items with the best rating, calculated using the WWP similarity), with the set R, consisting of the first n recommendations of items, generated by the canonical approach (the n items with the best rating, calculated using the WP similarity), with $n = 2, 4, \ldots, 10$. During the evaluation process, we ignored the users which do not have at least 5 ratings in the test set. The best *Jaccard Index* denotes the best strategy of recommendation, because it indicates the approach able to detect the highest number of correct recommendations. In Fig. 5, the y-axis reports the average value of the *Jaccard Index*, calculated for all users u in the test set (with a number of ratings ≥ 3), and for each n value tested ($top-2, top-4, \ldots, top-10$). As we can see, despite the low values measured, attributable to the wide range of values taken into account, our approach overcomes the canonical approach of recommendation based on the non-weighted Wu and Palmer metric.

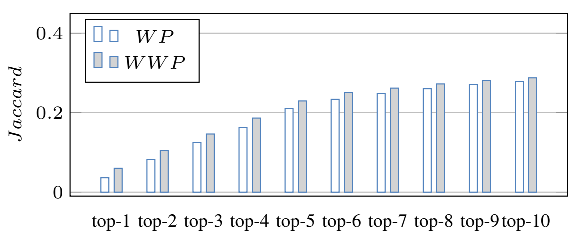

Fig. 5: Ratings Prediction

*3) **Ratings Influence:*** This last experiment has the aim to investigate about the relation existing between the number of positive ratings ($rating \geq 3$) present in the test set, and the performance of the approaches of recommendation (i.e., the canonical approach of recommendation based on the WP metric, and the proposed approach based on the WWP metric).

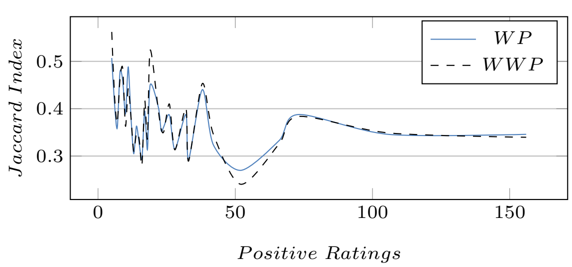

Fig. 6: $Ratings\ Influence$

The result of the performed experiment shows two interesting aspects:

1) firstly, the proposed WWP approach overlaps many times with the standard approach based on the WP metric, as we can observe in the first part of the graph. Considering that we adopted the Jaccard metric to evaluate the results, this means that we have been able to infer more correct recommendations than those performed through the canonical approach based on the non-weighted ontology of synsets;
2) secondly, most of the obtained improvements are correlated with a small number of recommendations, as a matter of fact, above the 20 recommendations, the results obtained by our novel WWP approach, and those obtained by the canonical WP approach, are almost overlapped.

This is an interesting result, because for a recommendation system is more difficult to make correct predictions with a few recommendations, rather than with many of these.

F. Discussion

The result of the first experiment, described in the Section VI-E1, shows the distribution of the synset occurrences, and at the same time detects the maximum number of occurrences for a single synset (since it is an essential parameter of the WWP metric). The obtained result also confirms the importance to attribute a different weight to each synset of the adopted ontology, because it is clearly evident that there are synsets which are less characterizing than others (i.e., synsets with a big number of occurrences in the ontology).

In the second experiment of Section VI-E2, we evaluated the ability of WWP approach to recommend items to the users, comparing to the ability of the canonical approach based on WP metric. The results show that our approach is able to produce better recommendations, compared to the canonical strategy of computation of the semantic similarity, which does not take into account the weight of the synsets (the standard *Wu and Palmer* formulation).

The last experiment, presented in Section VI-E3, confirms the results of the previous experiment, and shows two other important aspects correlated to each other. On the one hand, the proposed approach overcomes the standard strategy of recommendation, and on the other hand the major improvements are related to a small number of recommendations. This result proves the effectiveness of the proposed strategy, because in the context of recommender systems, this result represent an important improvement, considering the difficulty to make correct predictions, generating few recommendations.

In conclusion, it should also be noted that in this work we compared our approach to the Wu and Palmer metric, but the same strategy is able to enrich any other metric of the semantic similarity evaluation, based on the WordNet structure, and more generally, on any structure based on graphs.

VII. Conclusions and Future Work

In this paper we proposed a novel approach to to evaluate the semantic similarity between words (and sentences), which exploits a weighted ontology based on the inverse synset frequency. We have proved the effectiveness of the proposed strategy, comparing our results with those generated by a canonical approach of recommendation based on a non-weighted ontology, obtaining an improvement in terms of accuracy of the recommendations, especially when we generated a little number of these.

In future work, we will apply our approach in the context of different datasets and metrics based on a graph structure, in order to evaluate its effectiveness in others operating environments, expanding the possible area of application.

ACKNOWLEDGMENTS

This work is partially funded by Regione Sardegna under project SocialGlue, through PIA - Pacchetti Integrati di Agevolazione "Industria Artigianato e Servizi" (annualità 2010), and by MIUR PRIN 2010-11 under project "Security Horizons".

REFERENCES

[1] T. Slimani, "Description and evaluation of semantic similarity measures approaches," *CoRR*, vol. abs/1310.8059, 2013. [Online]. Available: http://arxiv.org/abs/1310.8059

[2] T. Berners-Lee, J. Hendler, O. Lassila *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.

[3] N. F. Noy, "Semantic integration: A survey of ontology-based approaches," *SIGMOD Record*, vol. 33, no. 4, pp. 65–70, 2004. [Online]. Available: http://doi.acm.org/10.1145/1041410.1041421

[4] E. G. M. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou, "X-similarity: Computing semantic similarity between concepts from different ontologies," *JDIM*, vol. 4, no. 4, pp. 233–237, 2006. [Online]. Available: http://www.dirf.org/jdim/abstractv4i4.htm

[5] P. Lops, "Semantics-aware content-based recommender systems," in *Proceedings of the 1st Workshop on New Trends in Content-based Recommender Systems co-located with the 8th ACM Conference on Recommender Systems, CBRecSys@RecSys 2014, Foster City, Silicon Valley, California, USA, October 6, 2014*, ser. CEUR Workshop Proceedings, T. Bogers, M. Koolen, and I. Cantador, Eds., vol. 1245. CEUR-WS.org, 2014, p. 1. [Online]. Available: http://ceur-ws.org/Vol-1245

[6] S. B. Johnson, "Research paper: A semantic lexicon for medical language processing," *JAMIA*, vol. 6, no. 3, pp. 205–218, 1999. [Online]. Available: http://dx.doi.org/10.1136/jamia.1999.0060205

[7] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," *CoRR*, vol. cmp-lg/9709008, 1997. [Online]. Available: http://arxiv.org/abs/cmp-lg/9709008

[8] C. Leacock and M. Chodorow, "Combining local context and wordnet similarity for word sense identification," *WordNet: An electronic lexical database*, vol. 49, no. 2, pp. 265–283, 1998.

[9] D. Lin, "An information-theoretic definition of similarity," in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998), Madison, Wisconsin, USA, July 24-27, 1998*, J. W. Shavlik, Ed. Morgan Kaufmann, 1998, pp. 296–304.

[10] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *CoRR*, vol. abs/cmp-lg/9511007, 1995. [Online]. Available: http://arxiv.org/abs/cmp-lg/9511007

[11] Z. Wu and M. S. Palmer, "Verb semantics and lexical selection," in *32nd Annual Meeting of the Association for Computational Linguistics, 27-30 June 1994, New Mexico State University, Las Cruces, New Mexico, USA, Proceedings*, J. Pustejovsky, Ed. Morgan Kaufmann Publishers / ACL, 1994, pp. 133–138. [Online]. Available: http://aclweb.org/anthology-new/P/P94/

[12] P.-O. Fjällström, "Algorithms for graph partitioning: A survey," *Linköping electronic articles in computer and information science*, vol. 3, no. 10, 1998.

[13] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer, 2011, pp. 1–35.

[14] A. Dennai and S. M. Benslimane, "Toward an update of a similarity measurement for a better calculation of the semantic distance between ontology concepts," in *The Second International Conference on Informatics Engineering & Information Science (ICIEIS2013)*. The Society of Digital Information and Wireless Communication, 2013, pp. 197–207.

[15] M. Capelle, F. Hogenboom, A. Hogenboom, and F. Frasincar, "Semantic news recommendation using wordnet and bing similarities," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ser. SAC '13. New York, NY, USA: ACM, 2013, pp. 296–302. [Online]. Available: http://doi.acm.org/10.1145/2480362.2480426

[16] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manage.*, vol. 24, no. 5, pp. 513–523, Aug. 1988. [Online]. Available: http://dx.doi.org/10.1016/0306-4573(88)90021-0

[17] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[18] G. Guizzardi, "On ontology, ontologies, conceptualizations, modeling languages, and (meta)models," in *Databases and Information Systems IV - Selected Papers from the Seventh International Baltic Conference, DB&IS 2006, July 3-6, 2006, Vilnius, Lithuania*, ser. Frontiers in Artificial Intelligence and Applications, O. Vasilecas, J. Eder, and A. Caplinskas, Eds., vol. 155. IOS Press, 2006, pp. 18–39. [Online]. Available: http://www.booksonline.iospress.nl/Content/View.aspx?piid=5421

[19] A. C. Costa, R. S. G. G. Guizzardi, and J. G. P. Filho, "Cores: Context-aware, ontology-based recommender system for service recommendation," in *Proceedings of the 19th international conference on advanced information systems engineering (CAiSE'07)*. Citeseer, 2007.

[20] H. Wache, T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner, "Ontology-based integration of information-a survey of existing approaches," in *IJCAI-01 workshop: ontologies and information sharing*, vol. 2001. Citeseer, 2001, pp. 108–117.

[21] M. Grüninger and M. S. Fox, "Methodology for the design and evaluation of ontologies," 1995.

[22] M. Uschold, "Building ontologies: Towards a unified methodology," *Technical Report - University of Edinburgh Artificial Intelligence Applications Institute AIAI TR*, 1996.

[23] M. S. Fox and M. Grüninger, "Ontologies for enterprise modelling," in *Enterprise Engineering and Integration*. Springer, 1997, pp. 190–200.

[24] A. Gomez-Perez, M. Fernández-López, and O. Corcho-Garcia, "Ontological engineering," *Computing Reviews*, vol. 45, no. 8, pp. 478–479, 2004.

[25] N. Noy and D. L. McGuinness, "Ontology development 101," *Knowledge Systems Laboratory, Stanford University*, 2001.

[26] C. Berge, "La theorie des graphes," *Paris, France*, 1958.

[27] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, ser. NAACL '03. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003, pp. 173–180. [Online]. Available: http://dx.doi.org/10.3115/1073445.1073478

[28] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.