A Feature Selection Approach for Anchor Evaluation in Ontology Mapping

Frederik C. Schadd and Nico Roos

Maastricht University, The Netherlands, email: {frederik.schadd, roos}@maastrichtuniversity.nl

Abstract. Computing alignments between ontologies is a crucial task for the facilitation of information exchange between knowledge systems. An alignment is a mapping consisting of a set of correspondences, where each correspondence denotes two ontology concepts denoting the same information. In this domain, it can occur that a partial alignment is generated by a domain expert, which can then be exploited by specialized techniques. In order for these techniques to function as intended, it must be ensured that the given correspondences, also known as anchors, are indeed correct. We propose an approach to this problem by reformulating it as a feature selection task, where each feature represents an anchor. The feature space is populated with a set of reliably generated correspondences, which are compared with the anchors using a measure of alignment. We apply feature selection techniques to quantify how well the anchors align with this set of correspondences. The resulting scores are used as anchor reliability measures and combined with the anchor similarities.

We evaluate the approach by generating a set of partial alignments for the used dataset and weighting the concept similarities with anchor evaluation measure of our approach. Three different similarity metrics are used, a syntactic, structural and semantic metric, in order to demonstrate the effectiveness of our approach.

1 Introduction

The availability of semantically structured data via the semantic web [3] allows for a varied set of approaches exploiting this data, ranging from data-warehousing and web site creation and management [19, 27], to querying one or multiple knowledge sources [16, 6]. The semantic structure of this data is determined by an ontology written by a domain expert using expressive languages such as RDFS [4] or OWL [20]. A common issue in this field is that two ontologies describing the same information can be heterogeneous with regard to its terminology, structure, scope or granularity [8]. If the situation arises that one needs to transfer information between knowledge systems using heterogeneous ontologies, then one needs transform the data in such a way that it is in compliance with the new ontology. For every concept in the first ontology a corresponding concept in the second must be identified which is used to store the same information. The task of identifying the correspondences between ontologies is known as ontology mapping.

Creating a mapping between ontologies is a laborious task which would require a domain expert to inspect both ontologies and determine mappings, if it were done by hand. Doing so becomes prohibitively difficult and time consuming when faced with increasingly large ontologies, such that automatic approaches are necessary for large scale problems [13, 32]. This also becomes a problem with the rise of the Semantic Web [3], which envisions autonomous agents automatically querying multiple knowledge sources for information.

Approaches which autonomously map ontologies have been an active field of research in the past decade [2, 5, 7, 12, 17, 23]. These systems use a varied selection of similarity metrics to determine the similarity of ontology concepts and use these values to derive a complete mapping, also known as alignment, between the input ontologies. Typically, multiple similarity measures are used to ensure the robustness of the system in the case that certain meta-information is missing in the ontologies. These similarity measures utilize different types of meta-information of the ontology concepts with the intuition that corresponding concepts will have aspect of this information in common, e.g. similar names, properties or neighbouring concepts.

A special case of a mapping problem is where a partial mapping is already available from a domain expert. Specialized approaches can then utilize this mapping to derive further correspondences in order to produce a complete mapping [2, 24, 28, 30]. However, these techniques rely on the correctness of the correspondences in the partial mapping, known as anchors, in order to deliver additional high quality correspondences. While, evaluating these correspondences using similarity metrics can be used for this task, they require a substantially high similarity threshold to ensure a high likelihood of correctness. This results in a large quantity of correspondences being filtered out, which also negatively affects the specialized techniques.

In this paper, we present an approach for the task of evaluating the correspondences originating from a partial alignment. This approach does not compare the concepts of these correspondences directly, allowing true correspondences whose concepts do not share much meta-information to be classified more correctly. The approach utilizes feature-selection techniques stemming from the fields of data mining and machine learning. We create a feature space with each feature representing an anchor, which is populated by generating a set of reliably correct and incorrect correspondences. The values of this feature space are computed using a measure of dissonance, which should yield predictable results only if the given anchor is correct. This predictability is then exploited using different feature selection methods. We evaluate this approach on a real-world dataset [14] by randomly generating partial alignments containing both correct and incorrect anchors and evaluating how well our approach can complement similarity-based techniques.

The remainder of this paper is structured as follows. The mapping problem is formally introduced in section 2 and detail our approach in section 3. We present the empirical evaluation in section 4. Finally, section 5 concludes the paper and suggests future research.

2 Mapping with Partial Alignments

Formally, ontology mapping is defined as the process of identifying concepts pairs, also referred to as correspondences between two ontologies which denote the similar information [9]. We define a correspondence between two ontologies O_1 and O_2 as a

5-tuple $< id, e_1, e_2, r, s >$, where id denotes a unique identifier, e_1 and e_2 denote a reference to a concept originating from O_1 and O_2 respectively, r denotes the semantic relation that is asserted between e_1 and e_2 and s denotes a confidence value lying in the interval $[0, 1]$. The classic process of ontology mapping thus receives as input two ontologies O_1 and O_2, and produces a set of correspondences $A = \{c_1, c_2, \ldots, c_n\}$, referred to as an alignment.

A special case of this process occurs when a partial alignment PA is available as additional input for the mapping process. This is an alignment that has been produced by a pre-processing approach or assembled by hand by a domain expert. However, this alignment considered to be incomplete, such that it becomes necessary to compute addition correspondences in order to generate a complete mapping between O_1 and O_2. To achieve this, the correspondences in PA can be utilized by special techniques in order to determine the remaining correspondences [24, 28]. The resulting alignment A thus be considered as the merger of all correspondences given in PA, since these can be assumed to be correct, and all correspondences that have been computed.

2.1 Anchor Filtering

While the correspondences originating from a partial alignment, referred to as anchors, can be assumed to be correct, this is not always the case. In case of a generated partial alignment, there is no guarantee that the used approach has a precision of 100% for every mapping task. If the partial alignment is made by a domain expert, it can always occur that the expert makes a mistake. The presence of incorrect anchors can degrade the quality of the computed correspondences, with the degradation of quality being correlated to the quantity of incorrect anchors. In order to ensure that a mapping approach that utilizes partial alignments performs as designed, it becomes necessary to perform a pre-processing step that ensures that the provided anchors are of sufficient quality.

The procedure of pre-processing partial alignments can be described by two key steps: *anchor evaluation* and the application of a *filtering policy*. Given two ontologies O_1 and O_2, and a partial alignment PA consisting of n anchors $\{c_1, c_2, \ldots c_n\}$, the *anchor evaluation* step produces a set of n scores $S = \{s_1, s_2, \ldots, s_n\}$, with each score s_x indicating the quality of its anchor c_x. The *filtering* step uses these scores to discard any anchor which does not satisfy a given policy, creating a new partial alignment PA', such that $PA' \subseteq PA$. The entire process is illustrated in Figure 1.

Typically, the *evaluation* and *filtering* steps are achieved through the application of already existing approaches from the field of ontology mapping. The *filtering* step can be performed by simply applying a threshold to the score set S, with the threshold value set by a domain expert or learned using a training set. To evaluate the anchors, one can utilize any available concept similarity metric [31]. However, such metrics are unfortunately susceptible to concept heterogeneities, where a concept pair for which a human would immediately conclude that it denotes the same information would result in a low similarity values. Such heterogeneities can be mitigated through the combination of multiple similarity metrics, though the aggregation of several similarity values has its disadvantages. For example, given two concept pairs which respectively receive the similarity values $\{0, 1\}$ and $\{0.5, 0.5\}$ as determined by two metrics, one would be more inclined to accept the first pair than the second, since it can occur that the feature

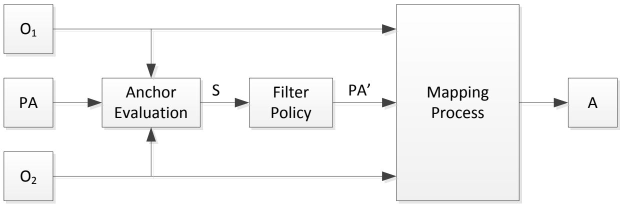


Fig. 1: Illustration of the anchor filtering process when mapping with partial alignments.

on which a similarity metric relies might be absent while at the same time the maximum score of a given metric is only rarely a false positive. Computing the aggregate of two similarities would thus obscure this information. The approach presented in this paper attempts to tackle this problem by proposing a new way in which a similarity metric can be used to evaluate anchors.

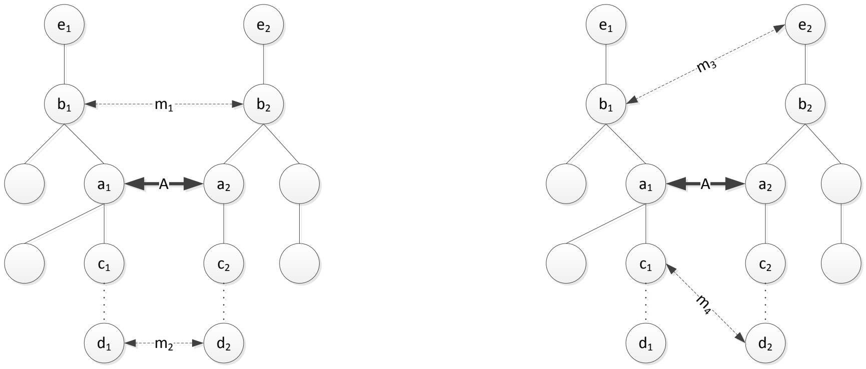
3 Proposed Approach

A similarity metric can produce a small set of reliable correspondences, given a sufficiently high similarity threshold. Furthermore, one can also reliably generate a set of incorrect correspondences, given a sufficiently low threshold. One can utilize correct and incorrect correspondences for the analysis of the anchors given in the input partial alignment. To achieve this, one needs a method for evaluating anchors against this set of reliable correspondences, which allows for the distinguishment of correct and incorrect anchors.

When comparing an anchor with a given correct correspondence, one would desire a measure which assigns this comparison a certain value, regardless of the proximity between the anchor and correspondence in the taxonomy. This aspect becomes especially important when dealing with poorly designed ontologies which may lack a thoroughly structure taxonomy. For example, comparing an anchor denoting the concept *vehicle* with two correct correspondences denoting the concepts *car* and *physical object* respectively, one would desire the same outcome despite *physical object* being less related to *vehicle* than *car*. The same would also be desired when comparing a correct anchor with incorrect correspondences, albeit with the output being different than the comparison with correct correspondences. One can interpret such a measure as expressing how well an anchor aligns with a correspondences, as opposed to measuring the semantic similarity between the anchor concepts. A correct anchor would thus be expected to be better aligned with regard to a reliably classified correspondence as opposed to an incorrect anchor. To minimize the effect of outliers and utilize all available reliably classified correspondences, one should measure the degree of alignment of an anchor and all given correspondences, and measure how well this measure correlates with the expected result. A way to measure how well an anchor aligns with a given correspondence would be to compute the concept similarities between the concepts in the anchor and the concepts of the given correspondence and express how these similarities differ. To measure this difference in similarity between the concepts of an anchor and the concepts of a given correspondence, we propose a measure of dissonance. Given a correspondence $\{c_1, c_2\}$, an anchor $\{a_1, a_2\}$ and a base similarity measure $sim(a, b) \in [0, 1]$, we define the dissonance d as follows:

$$d(\{c_1, c_2\}, \{a_1, a_2\}) = |sim(c_1, a_2) - sim(c_2, a_1)| \tag{1}$$

Using the measure of dissonance, the core of the approach consists of comparing the given anchor to a set of reliably generated correspondences, correct and incorrect, and quantifying to what extend the anchor aligns with the given correspondences. Based on this quantification, the set of anchors can then be filtered. For this research, we will investigate three different metrics when used as base similarity sim.

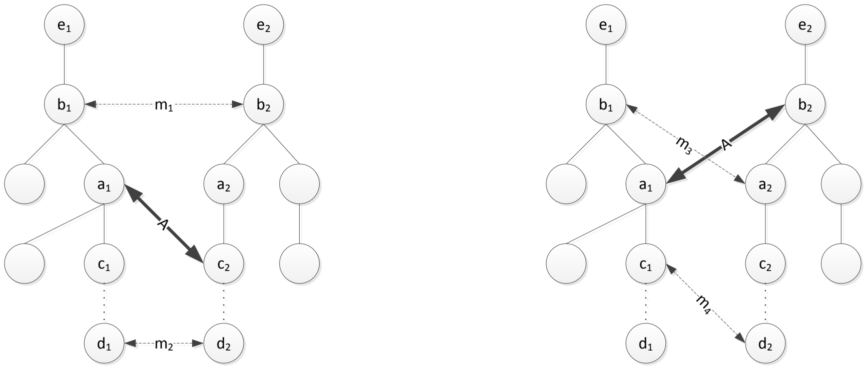


(a) Correct anchor A contrasted against two correct matches m_1 and m_2.

(b) Correct anchor A contrasted against two incorrect matches m_3 and m_4.

Fig. 2: Example scenarios of an anchor A being compared to either correct matches, illustrating the expected semantic difference between anchors and given correspondences.

To illustrate the principle behind the approach, consider the examples illustrated in Figures 2 and 3. Each example illustrates two ontologies, an anchor A and two correspondences linking two other concept pairs. Figure 2a depicts a correct anchor and two correct correspondences $m_1 = [b_1, b_2]$ and $m_2 = [d_1, d_2]$. m_1 is semantically more related to A than to m_2, thus it can be expected that when calculating $sim(a_1, b_2)$ and $sim(a_2, b_1)$ results in higher values than when computing $sim(a_1, d_2)$ and $sim(a_2, d_1)$. It is reasonable to presume that $sim(a_1, b_2)$ and $sim(a_2, b_1)$ will result in equally high, and $sim(a_1, d_2)$ and $sim(a_2, d_1)$ will result in equally low values, meaning that computing the dissonance $d(m_1, A)$ and $d(m_2, A)$ will result in equally low values, indicating a high degree of alignment.



(a) An incorrect anchor A contrasted against two correct matches m_1 and m_2.

(b) An incorrect anchor A contrasted against two incorrect matches m_3 and m_4.

Fig. 3: Four example scenarios of an anchor A being compared to incorrect matches, illustrating the irregularity in the expected semantic difference between anchors and given correspondences.

Comparing a correct anchor to dissimilar correspondences is expected to not exhibit this behaviour. Figure 2b illustrates a correct anchor A, consisting of the concepts a_1 and a_2, and two incorrect matches m_3 and m_4, which link the concepts b_1 with e_2 and c_1 with d_2 respectively. In this situation, a similarity calculation between a_2 and b_1 is likely to result in a higher value than the similarity between a_1 and e_2. Similarly, the concept similarity between the concepts of A and m_3 are also likely to differ, despite m_4 being semantically more apart from A than m_3.

When given an incorrect anchor, the similarity difference between the concepts of A and the concepts of either correct or incorrect matches are less likely to be predictable, as illustrated in Figure 3a and 3b. Figure 3a depicts an incorrect anchor A being compared to two correct correspondences. Here, both correspondences contain one concept, b_1 and d_2 respectively, which are semantically closer to A than their other concept. Thus, computing a similarity measure between the concepts of a correct correspondence and the concepts of an incorrect anchor will likely produce unequal results, regardless of the semantic distance of the correspondence to the anchor. However, to which degree these similarity will differ is not predictable, since this depends on how semantically related the concepts of the incorrect anchor are. If one were to compare an incorrect anchor to an incorrect correspondences, then the expected difference in concept similarities is not predictable at all, as illustrated in Figure 3b. The comparison of A with m_3 is likely to produce a low difference in similarity when comparing a_1 with a_2 and b_1 with b_2. On the other hand, the similarity difference between an anchor can be very large, as illustrated with m_4.

3.1 Filtering using Feature Selection

Having identified a measurement which leads to predictable behaviour for correct anchors and less predictable behaviour for incorrect anchors, one now needs to find a method for quantifying this predictability. As previously stated, in order for the dissonance to behave in a predictable way one must use correspondences of which their truth value is known with a high degree of certainty. The correct and incorrect comparison correspondences need to be generated reliably, such that labelling them as true and false respectively results in only few incorrect labels. Assuming that these generated correspondences have indeed their corresponding labels, one can interpret the different dissonance measures as separate samples over a feature space. Given a set of n input anchors $A = \{a_1, a_2, \ldots, a_n\}$ and the set of generated correspondences $C = \{c_1, c_2, \ldots, c_m\}$ with their respective labels $Y = \{y_1, y_2, \ldots, y_m\}$, containing both reliably correct and incorrect correspondences, each correspondence c_x would thus consist of n dissonance measurements $d_{x,i}(i = 1, \ldots n)$ and its label y_x. If an anchor a_x is correct, then evaluating the dissonances over C would lead to discernible differences for correct and incorrect correspondences, making the variable representing a_x in the feature space a good predictor of the labels Y.

To determine how well each dimension can serve as a predictor, one can utilize established feature selection techniques [15], which have become part of a set of important pre-processing techniques facilitating the use of machine learning and data-mining techniques on high-dimensional datasets. These techniques quantify how much a feature can contribute to the classification of a given labelled dataset. Their scores are then used in order to dismiss features which do not hold information that is relevant for classifying the data, allowing for the reduction of the feature space and the quicker training and execution of classifiers.

For this research, we will use the computed feature scores as evaluation metric for their corresponding anchors. Based on these values, a filtering policy can then dismiss anchors which are unlikely to be correct. Feature selection methods can utilize different underlying principles, for instance using correlation measures or information theory approaches. In order to not bias our approach to a single method, we will evaluate six different feature evaluation measures.

Pearson Correlation Coefficient A fundamental method in the field of mathematical analysis, the *Pearson Correlation Coefficient* [22] measures the linear correlation between two variables. Having the sample set X and Y of two variables, the *Pearson Correlation Coefficient* is defined as:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{2}$$

Spearman Rank Correlation The *Spearman Rank Correlation* [22] is a method which utilizes the method of computing the *Pearson Correlation Coefficient*. However, the sample sets X and Y are transformed into the ranking sets x and y. The correlation between x and y is then computed as:

$$p = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

Gain Ratio Information theoretical approaches have also been employed as measures of feature quality. Information gain techniques compute how much impurity is left in each split after a given attribute has been employed as the root node of a classification tree [26]. To measure this impurity, the measure of entropy is commonly employed. The entropy of a variable X is defined as:

$$H(X) = -\sum_{x_i} p(x_i) log_2 p(x_i) \tag{4}$$

The entropy after observing another variable is defined as:

$$H(X|Y) = -\sum_{y_j} p(y_j) \sum_{x_i} p(x_i|y_j) log_2 p(x_i|y_j) \tag{5}$$

The information gain of X is defined as the additional amount of information left after partitioning for all values of Y:

$$IG(X|Y) = H(X) - H(X|Y) \tag{6}$$

The *Gain Ratio* is defined as the normalized information gain:

$$GainRatio(X|Y) = IG(X|Y)/H(X) \tag{7}$$

Symmetrical Uncertainty The *Symmetrical Uncertainty* [11] is a measure that is similar to the *Gain Ratio*. It however employs a different normalization principle to counteract the bias towards larger attribute sets. Using equations 4 and 6, the *Symmetrical Uncertainty* $SU(X)$ can be computed as follows:

$$SU(X) = 2 \left[\frac{IG(X|Y)}{H(X) + H(Y)} \right] \tag{8}$$

Thornton's Separability Index Instead of using a correlation measure, *Thornton's Separability Index*[33] expresses separability between the classes in a dataset. Specifically, it is defined as the fraction of data-points whose nearest neighbour shares the same classification label. It is computed as follows:

$$TSI = \frac{\sum_{i=1}^{n} (f(x_i) + f(x_i') + 1) \mod 2}{n} \tag{9}$$

where f is a binary value function returning 0 or 1, depending on which class label is associated with value x_i. x_i' is defined as the nearest neighbour of x_i.

Fisher's Linear Discriminant *Fisher's Linear Discriminant* [10] evaluates the discriminatory quality of a set of features by calculating the difference of means of the features and normalizing this distance by a measure of the within-class scatter. The dataset is transformed into a linear space using the projection w which optimizes the output of the value function. The discriminant of two features can be computed as follows:

$$J(w) = \frac{|\mu_{y_1} - \mu_{y_2}|}{s_{y_1}^2 + s_{y_2}^2} \tag{10}$$

where μ_y and s_y^2 denote the means and variance of class y.

Using these feature evaluation methods one can evaluate the given anchors of a partial alignments with regards to their discriminatory qualities over the dissonance feature space. Based on the evaluation values, a filtering policy can then decide which anchors to discard before continuing the mapping process. The computation of these measures has been facilitated using the Java-ML framework [1].

4 Evaluation

To evaluate the proposed technique of filtering anchors, we utilized the *conference* dataset originating from the 2013 Ontology Alignment Evaluation Initiative [14]. This dataset contains matching tasks, including reference alignments, of real-world ontologies describing the domain of scientific conferences. While this dataset does not contain predefined partial alignments as additional input, it is possible to simply generate partial alignments from the supplied reference alignments. For this domain, it is preferable that the partial alignment also contains incorrect anchors such that the capability of filtering these incorrect anchors can be adequately tested. For each mapping task, PA is generated randomly such that it exhibits a *precision* and *recall* of 0.5 with respect to the reference alignment. Since we assume that a similarity metric can produce limited set reliable correspondences given a high threshold, as mentioned in Section 3, we limit the set of correct correspondences in the partial alignment to correspondences which do no exhibit a high pairwise similarity. The experiments thus provide an insight to what extent we can reliably evaluate anchors for situations where a basic similarity-based evaluation produces unreliable results.

Each task is repeated 100 times and the results aggregated in order to minimize random fluctuations. For each task, the given approach evaluates the given anchors, such that from the resulting scores a ordered ranking is created. While in a real-world application a given filtering approach would discard a series anchors based on a given policy, for instance by applying a threshold, for an experimental set-up it is more appropriate to perform a precision vs. recall analysis. Such an analysis allows for a comparison of performances without having to limit oneself to a set filtering policies.

To evaluate the dissonance between an anchor and a comparison correspondence, as stated in Section 3, a base similarity metric sim is required. We investigate three different categories of base similarity metrics:

Syntactic A comparison between concept names and labels using a specific algorithm. The Jaro [18] similarity was applied for this purpose.

Structural A comparison between concepts which also includes information of related concepts in its computation. As an example of a structural similarity, a profile similarity [25] has been evaluated. A profile similarity gathers syntactical information, e.g. concept names, labels and comments, from a given concept and its related concepts into a collection, which is referred to as profile. The similarity of two profiles determines the similarity of the corresponding concepts.

Semantic A similarity of this type aims to identify the meanings of concept senses within a lexical resource. The senses of the lexical resource are related with each other using semantic relations, e.g. *'is-a-kind-of'* relations, forming a taxonomy of senses. Concept similarities are determined by identifying the correct concept senses and determining the distance of these senses within the lexical taxonomy. This distance is then transformed into a similarity metric. For this evaluation a semantic similarity using WordNet as a lexical resource has been evaluated [29].

The final score of each anchor is determined by computing the pairwise similarity of the anchor concepts, also computed using sim, and multiplying this similarity with the anchor consistency score as determined using the proposed approach, using one of the tested feature evaluation methods. We will compare the rankings of our approach with a *baseline*, which is obtained by computing the pairwise similarities of the anchor concepts using the base similarity sim, while omitting the evaluation of the anchors using our approach. The comparison with the *baseline* allows us to establish how much our approach contributes to the evaluation of the given anchors.

The presented approach requires a method of generating the set of correspondences C which serve as individuals of the feature space. In order to apply feature selection techniques on a dataset, the class labels y of each individual must be known, and ideally also correct. Since a single similarity metric can produce a reliable set or correct correspondences, albeit limited in size, one can use this set as the part of C which represent true correspondences. In order to generate reliably incorrect correspondences, one can simply select two concepts at random while ensure that their pairwise similarity is below a threshold. For the experiments, the quantity of incorrect correspondences is set to be equal to the quantity of reliably correct correspondences. To generate C the *Jaro* similarity with thresholds 0.75 and 0.3 was utilized to ensure that the correspondences had a sufficiently high or low similarity.

4.1 Syntactic Similarity

In the first performed experiment the *Jaro* similarity was evaluated when applied as sim in order to evaluate a syntactical similarity. The generated anchors are evaluated and ranked according to their evaluation scores. We evaluate these rankings by computing their aggregated interpolated precision vs. recall values, displayed in Figure 4.

From the results depicted in Figure 4 several observations can be made. The most striking observation to be made is that all six tested feature evaluation methods produced a better ranking than the un-weighted baseline. At low recall levels this resulted in an increased precision of up to .057. At the higher recall levels we observe an increase in precision of up to .035.

With regard to the individual feature evaluation metrics a few trends are observable. First of all, we can see that the information theoretical approaches, meaning the *Gain-Ratio* and the *Symmetrical Uncertainty* improve the precision fairly consistently across all recall levels. On average, these measure improve the precision by approximately 0.3. The *Spearman rank correlation* and *Fisher's discriminant* only display a marginal improvement for lower recall levels, however show a more significant improvement for higher recall levels. The most significant improvements for the lower recall levels are observed when applying *Thornton's separability index* and *Pearson's correlation coefficient*.

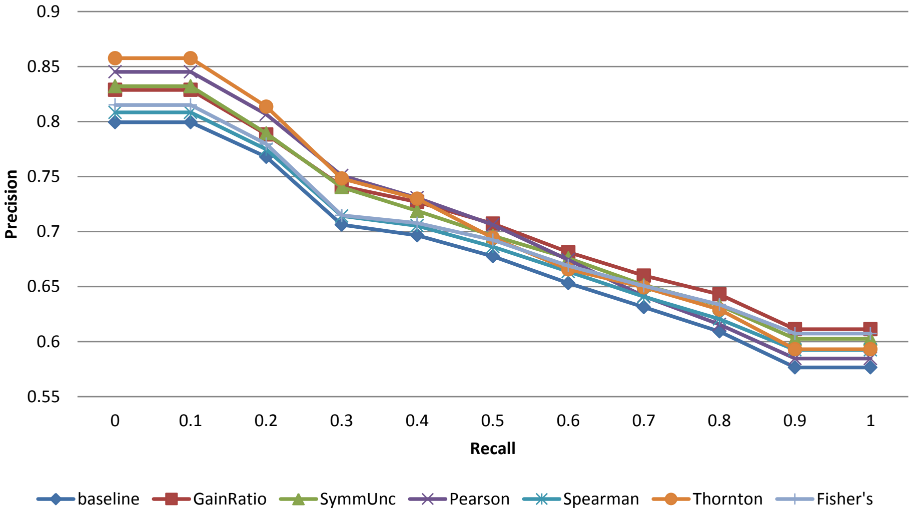


Fig. 4: Precision vs. recall of the rankings created using a syntactic similarity weighted by the evaluated feature selection methods. The un-weighted variant of the syntactic similarity is used as baseline.

4.2 Structural Similarity

For the second evaluation of our approach, we replaced the *Jaro* similarity with a *profile* similarity for sim. The *profile* similarity [25] compiles meta-information, primarily the name, comments and annotations, of a given concept and concepts that are linked to the given concept using relations such as *'is-a'* and *'domain-of'*. A profile similarity can be classified as a structural similarity due the utilization of information originating from related concepts. The gathered meta-information is represented as a weighted document-vector, also referred to as a profile. The similarity between two concepts is determined by computing the cosine similarity of their corresponding document vectors. The results of evaluating our approach using a profile similarity as sim can be seen in Figure 5.

From Figure 5 we can observe a more mixed result compared to the previous evaluation. The information-theoretical methods, namely *Gain Ratio* and *Symmetrical Uncertainty* outperform the baseline at lower recall levels, maintaining a near-perfect precision of 0.99 for one additional recall level and outperforming the baseline by a margin of roughly .022 at a recall of 0.3. However, for higher recall levels this margin drops until both measures perform roughly on par with the baseline at the highest recall levels. *Thornton's Separability Index* outperforms the baseline only at lower recall levels, while *Pearson's correlation coefficient* performs lower than the baseline. The most promising measures in this experiment were *Fisher's linear discriminant* and the *Spearman rank correlation*, which performed higher than the baseline for all recall levels. Contrary to the baseline, both measures produce a near perfect ranking of 0.99 at a recall of 0.2. The *Spearman rank correlation* produces rankings which have an increased precision

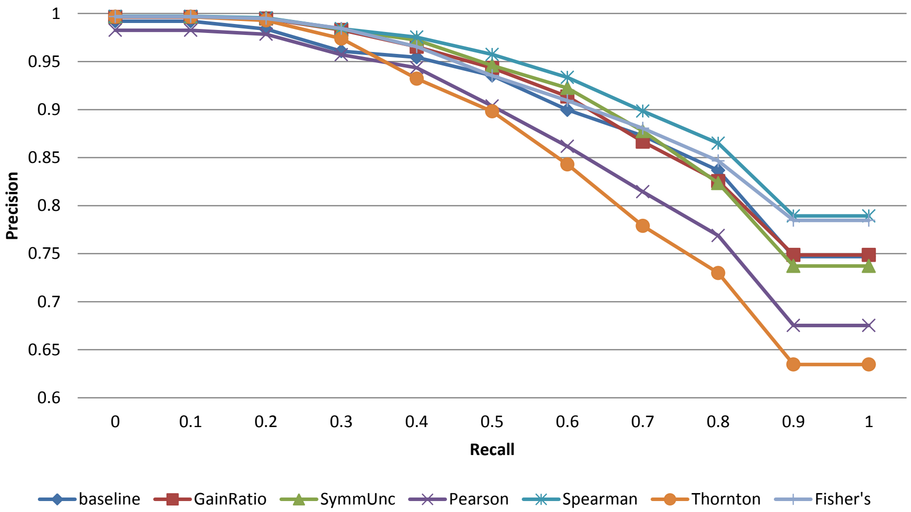


Fig. 5: Precision vs. recall of the rankings created using a structural similarity weighted by the evaluated feature selection methods. The un-weighted variant of the structural similarity is used as baseline.

of roughly .025 for most recall levels, while for the highest recall levels this difference is widened to roughly .045.

4.3 Semantic Similarity

In the third performed evaluation, we evaluated our approach when utilizing a semantic similarity as sim. A semantic similarity derives a similarity between two concepts by identifying their intended senses within a corpus and computing the semantic or taxonomic distance between the senses. The resulting distance value is then transformed into a similarity measure. For a semantic similarity to functions it is necessary that the given corpus also models the domains of the two input ontologies. To ensure this, WordNet [21] has been utilized as corpus, which aims at modelling the entire English language. The result of utilizing a semantic similarity as sim can be seen in Figure 6.

From Figure 6 several key observations can be made. First of all, the baseline displays a distinctively constant precision of .82 up to a recall level of .5. For the lower recall levels, our approach outperforms the baseline by a significant margin using any of the tested feature evaluation methods. Most measures produced an interpolated precision and recall of approximately .9, indicating an improvement of .08. When increasing the recall levels, the performance of these measures slowly approaches the performance of the baseline, while still staying above it. The exception is *Pearson's correlation coefficient*, which performs lower than the baseline at higher recall levels.

The clearly best performing measure is *Thornton's separability index*, which produced a precision higher than both the baseline and the other measures for all recall levels. At recall levels of .3 and higher Thornton's separability index improved upon

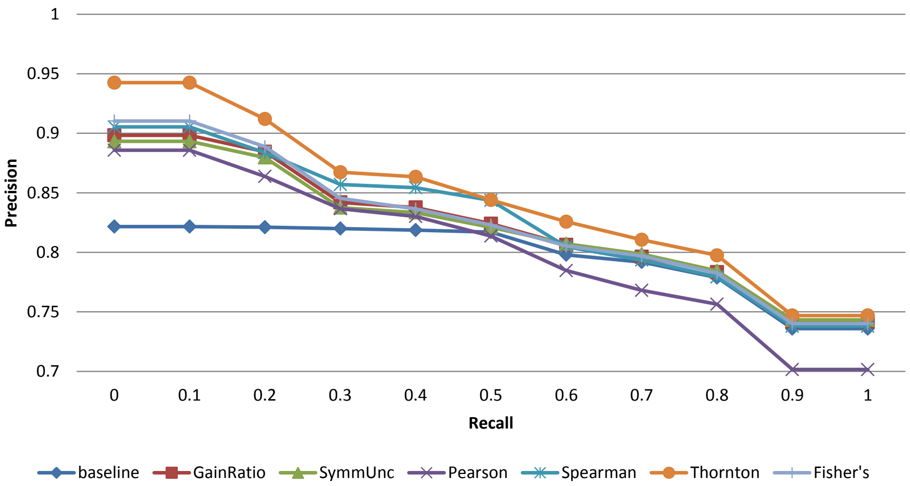


Fig. 6: Precision vs. recall of the rankings created using a semantic similarity weighted by the evaluated feature selection methods. The un-weighted variant of the semantic similarity is used as baseline.

the baseline by up to .047. At recall levels of .0 and .1 Thornton's separability index produced rankings with a precision of approximately .94, an improvement of .12 compared to the baseline. At a recall level of .2 it still produced rankings with a commendable precision of .91, which is .09 higher than the baseline.

Improvements of this magnitude are particularly important for the utilization of partial alignments, since they allow a significantly larger amount of anchors to be utilized while maintaining a degree of certainty that the anchors are correct. An approach which utilizes partial alignments relies on the quantity and quality of the anchors, but is likely biased towards the quality of the anchors. Thus in order to perform well, such an approach is likely to enforce stringent criteria on the given anchors instead of risking wrong anchors to be included in its computation. In the case of using a semantic similarity to achieve this, our approach would lead to a significantly higher amount of correct anchors being retained.

5 Conclusion and Future Research

In this paper we presented an approach of filtering correspondences of partial alignments, known as anchors, using feature selection techniques. By defining a measure of dissonance, with which one can compare anchors to correct or incorrect correspondences, and observing that this measure only renders predictable results if the anchors represent correct correspondences, we can formulate the task of evaluating anchors as a feature selection problem. Since this approach does not utilize the direct comparison of the anchor concepts, its resulting scores should display better results for anchors whose concept similarities do not allow for a reliable positive classification.

A variety of feature evaluation methods are empirically evaluated. A syntactical, structural and semantic similarity are evaluated as base similarity of our approach and compared to a baseline rankings obtained from computing the pairwise similarity of the anchors. We observe improvements for all tested base similarity measures, with the most significant improvements observed when utilizing the syntactic and semantic similarity. For the syntactic similarity all tested feature evaluation methods outperformed the baseline with regard to precision for all recall levels by approximately .057. For the semantic similarity we observed a particularly significant increase in precision up to .12 for lower recall levels, and a considerate increase for the remaining recall levels up to .047.

Overall, we conclude that our proposed approach displayed a promising start as a novel approach for evaluating anchors when mapping with partial alignments.

References

1. Abeel, T., Van de Peer, Y., Saeys, Y.: Java-ml: A machine learning library. Journal of Machine Learning Research 10, 931–934 (2009)
2. Aumueller, D., Do, H.H., Massmann, S., Rahm, E.: Schema and ontology matching with coma++. In: Proceedings of the 2005 ACM SIGMOD international conference on Management of data. pp. 906–908. ACM (2005)
3. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American 284(5), 34–43 (2001)
4. Brickley, D., Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema. http://www.w3.org/TR/rdf-schema (Feb 2004)
5. Cheatham, M.: Mapsss results for oaei 2011. In: Proceedings of the ISWC 2011 workshop on ontology matching. pp. 184–190 (2011)
6. Cruz, I., Xiao, H.: Ontology driven data integration in heterogeneous networks. Complex Systems in Knowledge-based Environments: Theory, Models and Applications pp. 75–98 (2009)
7. David, J., Guillet, F., Briand, H.: Matching directories and owl ontologies with aroma. In: Proc. of the 15th ACM international conference on Information and knowledge management. pp. 830–831 (2006)
8. Euzenat, J.: Towards a principled approach to semantic interoperability. In: Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing. pp. 19–25 (2001)
9. Euzenat, J., Shvaiko, P.: Ontology Matching, vol. 18. Springer Berlin (2007)
10. Fisher, R.A.: The use of multiple measurements in taxinomic problems. Annals of Eugenics 7(2), 179–188 (1936)
11. Flannery, B.P., Press, W.H., Teukolsky, S.A., Vetterling, W.: Numerical recipes in c. Press Syndicate of the University of Cambridge, New York (1992)
12. Giunchiglia, F., Autayeu, A., Pane, J.: S-match: an open source framework for matching lightweight ontologies. Semantic Web 3(3), 307–317 (2012)
13. Giunchiglia, F., Yatskevich, M., Avesani, P., Shvaiko, P.: A large dataset for the evaluation of ontology matching. The Knowledge Engineering Review Journal 24, 137–157 (June 2009)
14. Grau, B.C., Dragisic, Z., Eckert, K., Euzenat, J., Ferrara, A., Granada, R., Ivanova, V., Jiménez-Ruiz, E., Kempf, A.O., Lambrix, P., et al.: Results of the ontology alignment evaluation initiative 2013. In: Proc. 8th ISWC workshop on ontology matching (OM). pp. 61–100 (2013)
15. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (2003)

16. He, B., Chang, K.C.C.: Automatic complex schema matching across web query interfaces: A correlation mining approach. ACM Transactions on Database Systems (TODS) 31(1), 346–395 (2006)
17. Hu, W., Qu, Y.: Falcon-ao: A practical ontology matching system. Web Semantics: Science, Services and Agents on the World Wide Web 6(3), 237–239 (2008)
18. Jaro, M.: Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. Journal of the American Statistical Association 84(406), 414–420 (1989)
19. Kim, W., Seo, J.: Classifying schematic and data heterogeneity in multidatabase systems. Computer 24(12), 12–18 (1991)
20. McGuinness, D., van Harmelen, F.: OWL web ontology language overview. W3C recommendation, W3C (Feb 2004)
21. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38, 39–41 (November 1995)
22. Myers, J.L., Well, A.D., Lorch Jr, R.F.: Research design and statistical analysis. Routledge (2010)
23. Ngo, D., Bellahsene, Z., Coletta, R.: Yam++-a combination of graph matching and machine learning approach to ontology alignment task. Journal of Web Semantic (2012)
24. Noy, N.F., Musen, M.A.: Anchor-prompt: Using non-local context for semantic matching. In: Proceedings of the ICJAI workshop on ontologies and information sharing. pp. 63–70 (2001)
25. Qu, Y., Hu, W., Cheng, G.: Constructing virtual documents for ontology matching. In: Proceedings of the 15th international conference on World Wide Web. pp. 23–31. WWW '06, ACM, New York, NY, USA (2006)
26. Quinlan, J.R.: Induction of decision trees. Machine learning 1(1), 81–106 (1986)
27. Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. the VLDB Journal 10(4), 334–350 (2001)
28. Schadd, F.C., Roos, N.: Anchor-profiles for ontology mapping with partial alignments. In: Proceedings of the 12th Scandinavian AI Conference (SCAI 2013). pp. 235–244 (2013)
29. Schadd, F., Roos, N.: Coupling of wordnet entries for ontology mapping using virtual documents. In: Proceedings of The Seventh International Workshop on Ontology Matching (OM-2012) collocated with the 11th International Semantic Web Conference (ISWC-2012). pp. 25–36 (2012)
30. Seddiqui, M.H., Aono, M.: An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size. Journal of Web Semantics 7(4), 344–356 (2009)
31. Shvaiko, P., Euzenat, J.: A survey of schema-based matching approaches. In: Journal on Data Semantics IV, pp. 146–171. Springer (2005)
32. Shvaiko, P., Euzenat, J.: Ten challenges for ontology matching. Proceedings of ODBASE 2008 pp. 1164–1182 (2008)
33. Thornton, C.: Separability is a learners best friend. In: 4th Neural Computation and Psychology Workshop, London, 9–11 April 1997. pp. 40–46. Springer (1998)