Unsupervised Semantic and Syntactic Based Classification of Scientific Citations

Mohammad Abdullatif(✉), Yun Sing Koh, and Gillian Dobbie

Department of Computer Science,
The University of Auckland, Auckland, New Zealand
mabd065@aucklanduni.ac.nz,
{ykoh,gill}@cs.auckland.ac.nz

Abstract. In the recent years, the number of scientific publications has increased substantially. A way to measure the impact of a publication is to count the number of citations to the paper. Thus, citations are being used as a proxy for a researcher's contribution and influence in a field. Citation classification can provide context to the citations. To perform citation classification, supervised techniques are normally used. To the best of our knowledge there are no research that performs this task in a unsupervised manner. In this paper we present two techniques to cluster citations automatically without human intervention. This paper presents two novel techniques to cluster citations according to their contents (semantic) and the citation sentence styles (syntactic). The techniques are validated using external test sets from existing supervised citation classification studies.

1 Introduction

Authors cite papers for different reasons including identifying origin, providing background, giving credit, critiquing others' work and addressing the interestingness [8]. Citations are receiving more attention as the shear number of publications and subsequently the number of citations increase. This leads to the introduction of specialised research areas such as Bibliometrics concerned with utilising the existence of citations to evaluate the performance of researchers. One of the main citation assessment measures is the h-index. A scholar with an index of h means that the author has published h papers each of which has been cited at least h times. This measure presents both the productivity and impact of the researcher in his/her field using one number. A major drawback of this evaluation measure, as well as other similar ones, is that they focus on pure citation counts while ignoring the context or motivation of the citations. This leads to issues when comparing the work of researchers using the h-index. For example, consider two highly cited authors, one that has developed a new algorithm used by many in the field and the other published a paper with wrong, controversial claims, resulting in a high number of papers (and thus citations) disproving it. Using only measures that consider pure citation counts, both authors appear to have made the same contribution to their fields.

S. Madria and T. Hara (Eds.): DaWaK 2015, LNCS 9263, pp. 28–39, 2015.
DOI: 10.1007/978-3-319-22729-0_3

There has been some work in the area of citation analysis to address the pure citation count problem. A solution to this problem is to categorize citations based on their function or purpose. Such categorization will enable the creation of sophisticated evaluation models that are more accurate in reflecting the influence of a publication or the value of the work produced by a researcher.

Previous work tried to categorize citations by classifying them into predefined classes that are referred to as classification schemes [17]. Two approaches were used to perform citation classification. The first one uses manually created rules [18], and did not provide accurate classification. Moreover, the manually crafted rules were domain dependent and worked only for examples that have been accounted for by the domain expert crafting the rules. The second approach employed supervised machine learning techniques [4]. The classification accuracy achieved using this approach is reported to be much higher than the manually created rules. However, it was not possible to compare between the accuracies of the supervised machine learning techniques because they used different citation schemes for the classification. A supervised technique requires a manually coded dataset for training and testing. The difficulty lies in the fact that citation schemes proposed in the literature vary and thus a dataset created to train a classifier that uses one citation scheme cannot be used to train another classifier that uses a different citation scheme.

We are taking a different approach to citation classification. Instead of classifying citations to a pre-defined set of categories, we cluster a set of citations based on the reason behind them eliminating the need for a pre-defined classification scheme and the need for a training dataset, used in the other studies. Our unsupervised technique enables different levels of granularity when clustering, and does not require a scheme or a predefined training set.

In this paper we propose two novel techniques that can automatically cluster citations. As a brief overview, it is worth mentioning the most relevant inferences implemented aimed at solving the problem of citation clustering: First, syntactic inferences based on lexical distance measures. For instance, matching of consecutive subsequences, Jaro distance, Cosine distance, ISF specificity based on word frequencies extracted from corpora. Second, semantic inferences focused on semantic distances between concepts. These inferences implement several well-known WordNet-based similarity measures such as verb similarities between verbs.

The rest of the paper is organised as follows. Related work is outlined in Sect. 2. In Sect. 3 we describe our approaches to automatically cluster citation sentences. Section 4 presents our results and evaluation. We conclude and describe future work in Sect. 5.

2 Related Work

Drawing on classical bibliometrics papers, many researchers have focused their research on citation frequency and citation impact and applied it in different domains. Others have studied the association between the citation frequency of articles and various characteristics of journals, articles, and authors. Traditional citation analysis treats all citations equally. In reality, not all citations are equal. Some researchers consider location to be a factor affecting the relative importance of a citation. For example, the work by Herlach [11] found that a publication cited in the introduction or literature review section and re-mentioned in the methodology or discussion sections is more likely to make a greater contribution to the citing publication than others that have been mentioned only once. The stylistic aspects of a citation also matter. Bonzi [2] distinguished between a number of broad categories of citations. Examples of these include citations that are not specifically mentioned in the text and those barely mentioned or those with direct quotations.

Work on developing citation schemes goes back to the mid 1960's and was started by Garfield [9]. Garfield's work involved looking at the intentions of authors when citing and resulted in 15 different classes for citations. Other schemes have been developed since then and the number of classes vary between 3 and 35; some of which have mutually exclusive classes and others allow multiple classes to be attached to a single citation [21]. As an example, Moravcsik and Murugesan formed a scheme comprising four dimensions [17]. A citation can belong to one or more dimensions, which are presented in Table 1.

Table 1. Example of dimensions in a citation scheme

Dimension	Meaning
Conceptual	Concept or theory used
Operational	A tool or physical technique used
Organic	The reference is truly needed for the understanding of the paper
Perfunctory	The reference is just an acknowledgment of previous work
Evolutionary	The paper is a continuation of the cited work or the cited work is a foundation for the paper
Juxtapositional	The paper is an alternative to the work cited
Confirmative	The referenced paper is correct/the current paper agrees with the referenced paper
Negational	The paper contradicts or diminishes the referenced paper

All of the schemes described in the literature have been created by manually analyzing citation sentences and forming a scheme. The manually formed schemes are not compatible making it difficult to compare between the accuracy of classifiers trained using the different schemes.

3 Citation Clustering

In our approach we used a citation clustering approach to classify the citations. Ideally each cluster within the clustering results would represent a dimension/category in citation scheme. We looked at two different approaches to cluster citation sentences: Semantic-based models and Syntactic-based models. In this section we describe the measures used in each of these models.

3.1 Semantic-Based Model

The proposed semantic-based model aims to cluster citations by meaning. The introduced model consists of semantic-based term analysis and a semantic-based similarity measure. This work presents an ongoing project for semantic-based analysis of terms to enhance the quality of the citation clustering. This work introduces deep semantic analysis for concepts extracted from WordNet to enhance the quality of text clustering.

To automatically cluster a citation sentence, we concentrate on the verbs in each of the sentences. A verb is a word (part of speech) that in syntax conveys an action (bring, read, walk, run, learn), an occurrence (happen, become), or a state of being (be, exist, stand). We use verbs as the main indicator of the relationship between the citing and cited work. There are two approaches for selecting a verb from a sentence. The first one is using a supervised learning technique for verb selection such as Support Vector Machines [3]. The benefit of supervised learning is that it is easy to train given a training dataset. However, selecting the appropriate features for training is not always trivial. Also, a training dataset is not available. The second approach is using a systematically formed set of rules for deciding the most relevant verb to select. We opted for the second heuristic approach because it gives us human understandable reasons for why one verb was selected from a sentence as opposed to another. The rules also enable us to incrementally amend them as necessary to improve the performance of selecting the relevant verbs. Each citation sentence is processed through a semantic role labeler followed by our verb analysis process to extract the main verb relevant to the citations in the sentence. For example, passing the following sentence through our verb analysis process yields the verb “expand”.

> We **expand** on the work of James et al. [34] validated by the machine learning community.

The relevant verb extraction phase produces a set of relevant verbs $V = \{v_1, v_2, v_3, \ldots, v_k\}$ where k represents the total number of verbs found. The set of relevant verbs may not be unique. Each of the relevant verbs is derived from one citation sentence.Our citation relevant verb selection technique is described in [1].

Verb Similarity Computation. Given the set of relevant verbs V we need to sort the verbs in some methodological manner. To do so we measure the similarity between the verbs in the set. If we have k verbs, we need to compute the similarity between each verb and the rest of the verbs within the set. Thus we will have to make 2^k similarity comparisons. For example, if we have a set of verbs $\{v_1, v_2, v_3\}$, we will calculate the similarity for (v_1, v_2), (v_1, v_3), (v_2, v_3), (v_2, v_1), (v_3, v_1), (v_3, v_2). Each of the pairs will produce a similarity value, $s(v_i, v_j)$ where i and j are the index of two verbs where $i \neq j$. We then store the similarity values in a similarity matrix M of dimension $k \times k$.

We calculate the similarity between each pair of verbs. In our research the similarity measures we use are calculated using the graph structure of WordNet. A number of similarity measures that utilize WordNet for similarity computation have been proposed in the literature [13,19]. WordNet is a hand-crafted lexical database for the English language. The database groups words into synonym sets called synsets [6]. Synsets are linked together with many semantic links including the super-subordinate relation (also called hyperonymy, hyponymy or IS-A relation) and the part-whole relation (also called Meronymy). Synsets are organized in hierarchies where words become more abstract as you go up to the root node (called Entity for Nouns, and a simulated root node is created for verbs).WordNet has been used in a wide variety of applications including Information Retrieval [23] and Natural Language Processing tasks [16].

We used three verb similarity measures: Path Similarity, Wu-Palmer similarity [24], and Leacock Chodorow similarity [13] to measure the similarity between the pairwise verbs. We then present a comparison between them in order to select the most suitable for our application in Sect. 4. We note that all three similarity measures are symmetrical, thus the similarity value for $s(v_1, v_2)$ and $s(v_2, v_1)$ are the same. In this context the computational complexity of our similarity calculations is $\frac{2^k}{2}$.

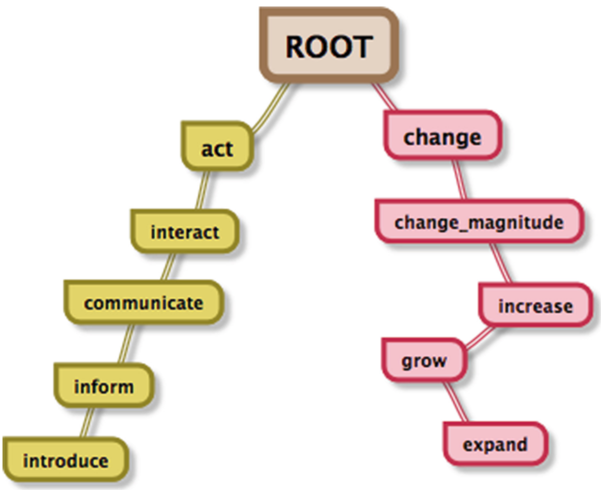

Fig. 1. WordNet hierarchy for verbs "expands" and "introduced"

The Path similarity measure is calculated as shown in Eq. (1) where $L(a, b)$ is the shortest path connecting verbs a and b in the IS-A (hypernym/hypnoym) taxonomy. The shortest path is based on the number of edges. Figure 1 shows the IS-A hierarchy for verbs "introduced" and "expands". In this example, the shortest path between the two verbs is 10. Using Eq. 1, the Path Similarity score is $1/(10 + 1) = 0.0909$.

$$PATH_{(a,b)} = \frac{1}{L_{(a,b)} + 1} \tag{1}$$

The Wu-Palmer similarity measure [24] is calculated using Eq. 2. $D(LCS(a, b))$ is the depth of the lowest common subsumer (deepest common ancestor/parent node of a and b) and $L(a, b)$ is the shortest path between nodes a and b. Using Eq. 2, the Wu-Palmer similarity score between verbs "introduced" and "expands" is as follows: The LCS of both verbs is the root node with a depth of 1. The shortest path is 10.

$$WUP_{(a,b)} = \max\Big[\frac{2 \times D(LCS_{(a,b)})}{L_{(a,b)} + 2 \times D(LCS_{(a,b)})}\Big] \tag{2}$$

Therefore the score is $(2 \times 1)/(10 + (2 \times 1) = 0.1667$.

The Leacock Chodorow similarity measure [13] is shown in Eq. 3 where $L(a, b)$ is the shortest path connecting a and b and $D_{\max}$ is the maximum depth from the root to the deepest leaf in the hierarchy in which the verbs occur. The distance for cases when a and b are the same verb will result in an infinite similarity score (log of zero). Therefore we always add 1 to the shortest path distance to avoid such a scenario. Using Eq. 3, the Leacock Chodorow similarity score between verbs "introduced" and "expands" is as follows: $L(a, b)$ is 10. $D_{\max}$ is 13. This is based on WordNet version 3.0. Therefore, the Leacock Chodorow similarity score is $-\log((10 + 1)/2 \times 13) = 1.0609$.

$$LCH_{(a,b)} = \max\Big[-log(\frac{L_{(a,b)}}{2D_{\max}})\Big] \tag{3}$$

Clustering of Verb Similarity Vectors. Calculating the similarity between all the pair of relevant verbs from the dataset of citation sentences results in a similarity matrix. Each row in the similarity matrix M represents the similarity between one verb and all the other verbs in the dataset. A row in the matrix is known as a similarity vector, SV_{v_k}, for its associated verb, v_k.

We cluster the vectors representing citations using the well-known clustering algorithm k-means [15]. We chose to use k-means because of its intuitive nature and its ability to allow us to specify the number of clusters we want. The ability to set a k value allows us to have control over the number of clusters which enables us to evaluate our technique using test sets from the supervised based citation classification work described in the literature. The end result of the clustering technique is a set of clusters $C = \{c_1, c_2, \ldots, c_k\}$. Each cluster c contains a set of verbs $\{v_j, v_{j+1}, \ldots, v_k\}$. Sentences associated with the verbs are clustered together. Each cluster represents one dimension or reason for citing.

3.2 Syntactic-Based Model

Instead of concentrating only on a particular verb, the entire citation sentence could be used to form the similarity matrix. There are various measures such as Jaro distance, Ratio, and Levenshtein distance that allow for character-based string similarity computation. We adapted these techniques to use words instead of characters as tokens. We then preprocess the citation sentences by performing stemming and stop words removal. Stemming is the process of reducing words to a base or root form. For example, words "advanced", "advancing", "advance" would all be reduced down to "advanc" after stemming is applied. Many stemming algorithms are available in the literature, we used Porter's stemmer [20] to stem words in our experiments. The intuition behind this is authors in the same field tend to write and cite literature in a similar fashion, thus looking at the syntactic nature of the sentences is a viable way of differentiating the types of citations.

Syntactic Similarity Computation. Current distance techniques measure similarity between the sentences based on the words used. However each of these techniques, produces similarity based on a slightly different approximation. The choice of the similarity distance can be dependent on the corpus used and allowable error estimates.

Here we used four different measures. The first measure is Levenshtein edit distance [14]:

$$Lev_{a,b}\ (i,j) = \begin{cases} \max(i,j) & \text{if} \min(i,j) = 0, \\ \min \begin{cases} \mathrm{lev}_{a,b}(i-1,j)+1 \\ \mathrm{lev}_{a,b}(i,j-1)+1 \\ \mathrm{lev}_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \tag{4}$$

where $lev_{a,b}(i-1,j)+1$ corresponds to a deletion operation, $lev_{a,b}(i,j-1)+1$ to an insertion and $lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)}$ to a match or mismatch. $1_{(a_i \neq b_j)}$ is equal to 0 when $a_i = b_j$ and equal to 1 otherwise, where a, b are citations within the sentences.

The second measure is Jaro Measure [12]:

$$Jaro(s,t) = \frac{1}{3} \cdot \left(\frac{|s'|}{|s|} + \frac{|t'|}{|t|} + \frac{|s'| - T_{(s',t')}}{|s'|} \right) \tag{5}$$

where s is a string with a_1 ... a_K characters, t is a string with b_1 ... b_L words. Here a character a_i in s is in common with t, if there is a $b_j = a_i$ in t such that $i-H \leq j \leq i+H$, where $H = \frac{min(|s|,|t|)}{2}$, s' has a'_1 ... a'_K words that are common between s and t in the same order they appear in s, t' has b'_1 ... b'_K words that are common between t and s in the same order they appear in t, $T_{(s',t')}$ is half the number of transpositions for s' and t'.

The third measure is Ratio:

$$Ratio(s,t) = \frac{2 \cdot M}{T} \tag{6}$$

Table 2. Similarity vectors

SV_{v_1}	$s(v_1, v_1)$	$s(v_1, v_2)$	$\ldots$	$s(v_1, v_k)$
SV_{v_2}	$s(v_2, v_1)$	$s(v_2, v_2)$	$\ldots$	$s(v_2, v_k)$
$\ldots$				
SV_{v_k}	$s(v_k, v_1)$	$s(v_k, v_2)$	$\ldots$	$s(v_k, v_k)$

where M is the number of matches and T is the number of elements in both s and t.

The final measure is Hamming distance, whereby we calculate the distance between two sentences by the number of positions at which the corresponding words are different [10]. On top of that we also used a TF-ISF based method that takes into account the word frequencies adjusted by the factor to account for very frequent words and computes the Cosine similarity between the resulting TF-ISF vectors.

Clustering of Syntactic Similarity Vectors. Similar to the clustering of verb similarity, the similarity between all the pairs of citation sentences in the dataset is stored in a similarity matrix. Each row in the similarity matrix M represents the similarity between one sentence and all the other sentences in the dataset. A row in the matrix is known as a similarity vector, SV_{s_k}, for its associated sentence, s_k. We then built a matrix similar to Table 2 and used k-means for clustering similar to the approach in Sect. 3.1.

4 Experiments, Results and Evaluation

To evaluate the viability of our technique we compared our technique to human annotated datasets. We used the weighted F-measure [7] to evaluate the performance of our techniques. The experiments were conducted on two different datasets used by Dong et al. [5] and Teufel et al. [22].

In the first set of experiments using the dataset from Dong et al. two different numbers of clusters were chosen: 3 and 6. These were chosen because in their schema they divide the citations into the 3 and 6 general categories. The corpus comes from the ACL (Association for Computational Linguistics) Anthology, a comprehensive collection of scientific conference and workshop papers in the area of computational linguistics and language technology. The authors randomly chose papers from proceedings of the ACL conference in 2007 and 2008. We compared the accuracy of our technique using the F-measure against the human annotated datasets. The results of the F-measures are shown in Figs. 2 and 3. Table 3 presents the name of each syntactic based experiment, the variation and the shorthand. We used both syntactic and semantic based approaches. The semantic based results are shown in the last three bars in the figures. In all the experiments we repeated k-means 200 times and averages are reported along with one standard deviation as an error bar with each average.

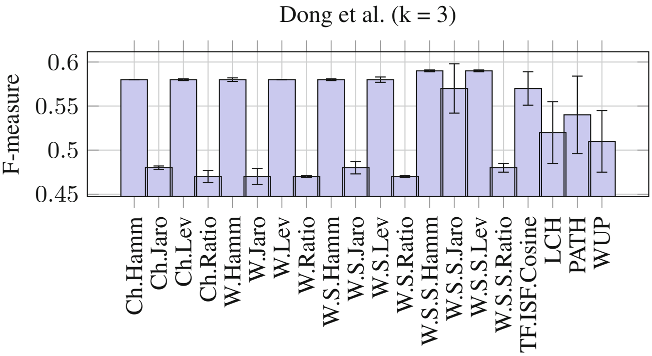

Fig. 2. F-measure on dataset from Dong et al. using k= 3

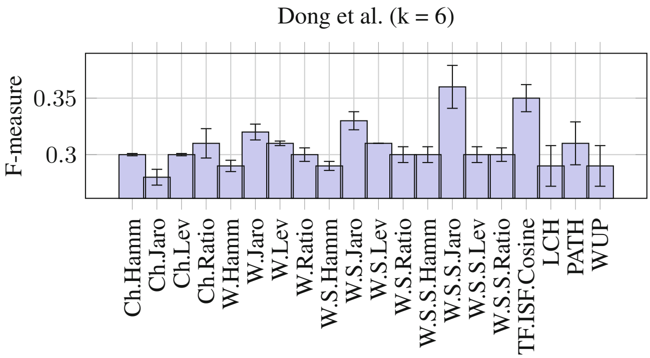

Fig. 3. F-measure on dataset from Dong et al. using k= 6

The number of independent experiments we ran for each test set is 20 where the difference between the experiments is the way the similarity vectors are formed. There are eight similarity measures that are used to form the similarity vectors. All of the measures except TF-ISF Cosine, WUP, PATH and LCH have four variations. The variations for each of the four measures include (1) characters: similarity between strings at the character level, (2) words: similarity between strings at the word level, (3) words stemmed: words are stemmed before the similarity between strings is computed, and (4) words stemmed and stop words removed: stop words are removed and the remaining words stemmed before the similarity is computed.

Table 3. Shorthands used to represent the experiments being run on each test set

Similarity measure	Hamming distance	Levenshtein distance	Jaro distance	Ratio
Characters	Ch. Hamm	Ch. Lev	Ch. Jaro	Ch. Ratio
Words	W. Hamm	W. Lev	W. Jaro	W. Ratio
Stemmed words	W.S. Hamm	W.S. Lev	W.S. Jaro	W.S. Ratio
Stemmed/Stopwords removed Words	W.S.S. Hamm	W.S.S. Lev	W.S.S. Jaro	W.S.S. Ratio

In the second set of experiments using the corpus of Teufel et al. three different numbers of clusters were chosen: 3, 4, and 12. This was chosen as in their schema they divide the citations into 3, 4, and 12 general categories. The dataset comes from a corpus of 360 conference articles in computational linguistics, drawn from the Computation and Language E-Print Archive. The results were compared against on the human annotated datasets and are shown in Figs. 4, 5 and 6. In total, 20000 clustering runs are performed (20 independent experiments × 200 runs × 2 Dong datasets × 3 Teufel datasets).

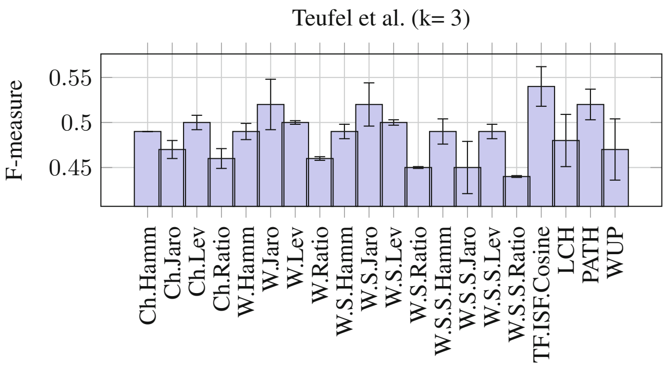

Fig. 4. F-measure on dataset from Teufel et al. using k= 3

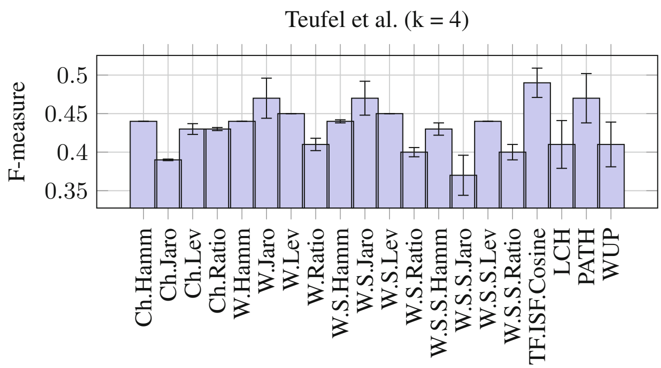

Fig. 5. F-measure on dataset from Teufel et al. using k= 4

From both sets of experiments we can see that the results are very consistent, whereby the top three performing measures were, Jaro, TF-ISF, and Path. Overall all the top performing measures performs well on the dataset across the experiments. This research is definitely a good starting point to how unsupervised automatic citation classification techniques can be built.

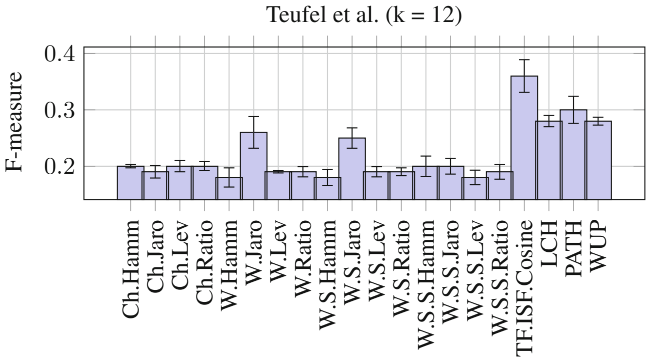

Fig. 6. F-measure on dataset from Teufel et al. using k = 12

5 Conclusion and Future Work

Citation Classification plays an important role in improving the current citation based research evaluation techniques such as the h-index. Most existing citation classification techniques perform the classification based supervised learning algorithms that require training data and the selection of a citation classification scheme. Comparison between the performance of the existing techniques is difficult because the techniques use different schemes and the supervised algorithms are therefore trained using different training sets. In this paper we proposed two novel techniques that can automatically cluster citations in an unsupervised manner based on semantic and syntactic inferences without the need for a citation scheme or a training corpus and we show that we can achieve reasonable results. Future work on the research includes combining both semantic and syntactic inference together and using an ensemble method to provide a better citation clustering and generate a citation scheme automatically.

References

1. Abdullatif, M., Koh, Y.S., Dobbie, G., Alam, S.: Verb selection using semantic role labeling for citation classification. In: Proceedings of the 2013 Workshop on Computational Scientometrics: Theory and Applications CompSci 2013, pp. 25–30. ACM, New York (2013). http://doi.acm.org/10.1145/2508497.2508502
2. Bonzi, S.: Characteristics of a literature as predictors of relatedness between cited and citing works. J. Am. Soc. Inf. Sci. **33**(4), 208–216 (1982). http://dx.doi.org/10.1002/asi.4630330404
3. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
4. Dong, C., Schäfer, U.: Ensemble-style self-training on citation classification. In: Proceedings of 5th International Joint Conference on Natural Language Processing, pp. 623–631. Asian Federation of Natural Language Processing, Chiang Mai, November 2011
5. Dong, C., Schäfer, U.: Ensemble-style self-training on citation classification. In: Proceedings of the 5th International Joint Conference on Natural Language Processing, pp. 623–631. Association for Computational Linguistics, November 2011

6. Fellbaum, C.: WordNet: An Electronic Lexical Database. Language, Speech, and Communication, MIT Press, Cambridge (1998)
7. Fung, B.C., Wang, K., Ester, M.: Hierarchical document clustering using frequent itemsets. In: SDM, vol. 3, pp. 59–70. SIAM (2003)
8. Garfield, E.: Journal impact factor: a brief review. Can. Med. Assoc. J. **161**(8), 979–980 (1999)
9. Garfield, E., Stevens, M.E., Giuliano, V.E., Heilprin, L.B.: Can citation indexing be automated? Essay Inf. Sci. **1**, 189–192 (1965)
10. Hamming, R.W.: Error detecting and error correcting codes. Bell Syst. Tech. J. **29**(2), 147–160 (1950)
11. Herlach, G.: Can retrieval of information from citation indexes be simplified? multiple mention of a reference as a characteristic of the link between cited and citing article. J. Am. Soc. Inf. Sci. **29**(6), 308–310 (1978). http://dx.doi.org/10.1002/asi.4630290608
12. Jaro, M.A.: Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida. J. Am. Stat. Assoc. **84**(406), 414–420 (1989)
13. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database, pp. 305–332. MIT Press (1998)
14. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions and reversals. Sov. Phys. Dokl. **10**, 707 (1966)
15. MacQueen, J., et al.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, p. 14, California (1967)
16. Morato, J., Marzal, M.A., Lloréns, J., Moreiro, J.: Wordnet applications. In: Proceedings of the 2nd Global Wordnet Conference, vol. 2004 (2004)
17. Moravcsik, M., Murugesan, P.: Some results on the function and quality of citations. Soc. Stud. Sci. **5**(1), 86 (1975)
18. Nanba, H., Okumura, M.: Towards multi-paper summarization using reference information. In: International Joint Conference on Artificial Intelligence, vol. 16, pp. 926–931. Lawrence Erlbaum Associates Ltd (1999)
19. Patwardhan, S., Pedersen, T.: Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In: Proceedings of the EACL 2006 Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together, pp. 1–8, Trento, April 2006
20. Porter, M.F.: An algorithm for suffix stripping. Prog. Electron. Lib. Inf. Syst. **14**(3), 130–137 (1980)
21. Radoulov, R.: Exploring automatic citation classification. University of Waterloo, Waterloo (2008)
22. Teufel, S., Siddharthan, A., Tidhar, D.: Automatic classification of citation function. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing EMNLP 2006, pp. 103–110. Association for Computational Linguistics, Stroudsburg, PA (2006). http://dl.acm.org/citation.cfm?id=1610075.1610091
23. Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E.G., Milios, E.E.: Semantic similarity methods in wordnet and their application to information retrieval on the web. In: Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management, pp. 10–16. ACM (2005)
24. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics ACL 1994, pp. 133–138. Association for Computational Linguistics, Stroudsburg (1994)