Zone identification based on features with high semantic richness and combining results of separate classifiers

Kambiz Badie a, Nasrin Asadi b and Maryam Tayefeh Mahmoudic c

aKnowledge Management & e-Organization Group, IT Research Faculty, ICT Research Institute, Tehran, Iran; bIT Research Faculty, ICT Research Institute, Tehran, Iran; cMultimedia Research Group, IT Research Faculty, ICT Research Institute, Tehran, Iran

ABSTRACT

In this paper, we propose a new approach to zone identification which is based on considering features with high semantic richness. Out of the scenarios of selecting features for identifying a zone based on classifying the sentences in a text, we came to notice that in the scenario where specialized names belonging to a text’s domain and mode of the verbs together with reduced versions of conventional features, including history, are taken into account, an accuracy rate of 61% (resp. 81%) is obtained which is higher than that belonging to Liakata’s (resp. Fisas’s) approach. Also, to have a genuine comparison, both Liakata’s and Fisas’s corpora are used in our experiments. Such accuracy is obtained at the place where less computational cost for extracting the features was decreased. In order to improve the accuracy of zone identification, a decision-level fusion process based on combining the results of separate classifiers, was considered. With regard to this, two fusion techniques of ‘majority voting’ and ‘average of probabilities’ were used. Experimentations show the fact that ensemble of ‘Logistic Regression’, ‘Support Vector Machine’ and ‘Neural Network’ as classifiers yields the best performance. Also ‘majority voting’ was shown to perform a bit better than ‘average of probabilities’.

Introduction

The number of journal papers has grown up dramatically during recent years and analysing the content and the structure of scientific texts can be of great assistance to researchers to access their needed information more easily. Within the past years, zone identification has therefore been elaborated as a major research concern within the areas of text mining in general and text summarization in particular (Hong, 2007; Liakata, Saha, Dobnik, Batchelor, & Rebholz-Schuhmann, 2012; Teufel & Moens, 2002). The major purpose behind this issue is to identify those zones in a text which tackle a certain concept issue, topic or subject from the reader’s point of view. Examples can be mentioned for the approaches trying to find out which parts (comprising a number of...
sentences) in a text refer to significant sections in a paper like ‘background’, ‘proposed approach’, ‘experimentation’, ‘approach validation’ or ‘conclusion’ as the important perspectives which are to be followed (pursued) in a paper, or for instance the approaches which try to figure out whether or not a certain scientific concept, subject or issue has been addressed somewhere in a scientific paper. Obviously, the more concrete the concept of a zone class with well-defined elements, a higher possibility would exist to identify the desired zone meaningfully with less emphasis on complicated features. Also, the higher the abstraction level of a zone class, the more effort would be necessary to take into account higher order linguistic features to identify the desired zones. This is mainly because, in comparison with a simple phrase, a subject or an issue usually calls for sophisticated relations between varieties of simple concepts to have itself characterized meaningfully.

In earlier works by various authors, zone identification has been examined under various titles (with little difference in meaning). The available schemas include Argumentative Zoning (Guo, Korhonen, & Poibeau, 2011; Teufel & Kan, 2011; Teufel, 2000), Discursive Structure (Fisas, Saggion, & Ronzano, 2015), Conceptual Zone categories (Liakata et al., 2012) and Information Structure (Guo et al., 2010; Guo, Reichart, & Korhonen, 2015). However, many efforts are still in progress to improve the existing methods in this direction and the not so far-reaching goal is to facilitate the search and study of scientific papers. The current paper should be considered to follow the same lines of thought.

This paper is an improved and extended version of our previous work (Badie, Asadi, & Mahmoudi, 2017). In the aforementioned conference paper, we demonstrated how through considering features with high semantic richness such as specialized names (belong specifically to a text’s domain of interest) and besides that mode of verbs, a higher classification accuracy can be attained for zone identification at the place where features with less computational cost compared to the conventional features are being used. Two well-known datasets with different granularity level have been selected for our experimentations: ART and Dr Inventor (DRI) corpora. ART corpus contains more detailed zone categories in the field of biochemistry, while DRI includes more general zone categories in the field of computer graphics.

In the current paper, four common classifiers are applied to evaluate the proposed features either individually or by combining a set of them. Two decision-level fusion techniques including majority voting and average of probabilities are considered. In this regard, we also use feature selection to decrease the training and testing time. In addition, supplementary experiments are conducted to show the importance of the proposed features on the classification accuracy.

**Related works**

Recently, automatic identification of zone categories existing within the scope of articles has become quite important and many researchers have tried to analyse the content of scientific texts from various points of view. Some of them focus on categorizing the sentences in the abstracts of articles (Bonn & Swales, 2007; McKnight & Srinivasan, 2003) while many others have worked on full-text articles (Groza, Hassanzadeh, & Hunter, 2013; Teufel & Kan, 2011; Mizuta & Collier, 2004). It should be mentioned that these zone identification
approaches may be different in classification method, selected features, annotation scheme and domain of the dataset to be used.

Most of the existing zone identification approaches make use of classification techniques such as ‘Support Vector Machine (SVM)’, ‘Naïve Bayes’, ‘Logistic Regression (LR)’ and ‘Conditional Random Fields (CRF)’ to classify the sentences (Fisas et al., 2015; Kilicoglu, 2017). Seaghdha and Teufel (2014) proposed BOILERPLATE-LDA, an unsupervised model, that elicits some aspects of rhetorical structure from an unannotated text and uses them as features to classify the sentences in zone categories. In another study, using association rule mining on the dependency structure of the sentences, Groza (2013) detected structural differences between zone categories. Guo et al. (2015) used topic models to extract the latent topics in the papers and tried to recognize information structure of the sentences by applying unsupervised machine learning techniques such as Graph Clustering and Generalized Expectation.

With regard to annotation schemes, several alternatives have also been created for various fields of sciences, some of which include only a few number of categories (Agarwal & Yu, 2009) while some others are finer-grained (Liakata et al., 2012) that capture the content and the conceptual structure of scientific articles. These schemes have been applied to articles in various domains such as biochemistry, chemistry, graphic computer, etc.

Soldatova and Liakata (2007) introduced a sentence-based 3-layer scheme, called CoreSC, which recognizes the key points of articles and consists 11 categories. Liakata et al. (2012) made use of SVM and CRF techniques to classify the related sentences in the biochemistry texts. In this regard, a classification rate of 51.6% was obtained by applying multi-class classification through using SVM. Fisas et al. (2015) used both SVM and LR to make classification on sentences within the area of computer graphics, and a rate of 80% was obtained at most by using regression. The features used to classify the sentences take into account different aspects of a sentence, ranging from its location in the article and the headline of the corresponding section within the document, to features which relate to the components of a sentence such as verbs, n-grams and the relation between them, like Grammatical triples (Fisas et al., 2015; Liakata, 2010). In the aforementioned works, n-grams have been introduced as the main features so that removing them could lead to a decrease in the precision of the classifiers. However, examining the significant n-grams show they have not potential to exhibit the required semantic richness. Examples can be mentioned for n-grams like ‘chem’, ‘along with’, ‘micropore and’, ‘bulk’, ‘available and’, ‘al.’, ‘often be’.

The current paper comes up from an effort to improve the aforementioned results of Liakata and Fisas. The advantage of our work is that we exploit features with high semantic richness that in the meantime are of less computational costs. Instead of all n-grams, we extracted only specialized noun phrases from the corpus. Also, due to the great role played by the verbs in discriminating the zone categories, we try to classify and use them as complementary features. Furthermore, in order to improve the classification accuracy, we make use of the fusion techniques, on the one hand, to combine decisions made by the classifiers, and feature selection as a pre-processing approach, on the other hand, to decrease the classification time. Applying these considerations provides us with a higher classification accuracy.
The proposed approach

Basic idea

The main point in our approach is to see how far, through considering features with semantic richness such as mode of verbs in a sentence, one can attain a better perception toward the zone class to which a sentence belongs to. In the meantime, a right perception toward the status of specialized nouns (either general or specific) in a sentence may have the potential to help zone identification be performed in a more meaningful way with less amount of computational cost. Status of verbs is important since the identity of a zone class in many cases depends on the way its specialized nouns are verbalized. Meanwhile, relative’s position of a sentence in the text for which a variety of parameters are to be considered is to be characterized with reasonable amount of information to avoid extra computational cost.

Another point is that mapping correctly a sentence onto zone classes which share some similarities is a difficult issue, which calls for further features preferably with deeper linguistic sense. Such an issue which is quite hard to be tackled from natural language processing viewpoint. With regard to Liakata’s approach, examples can be mentioned for a sentence belonging to zones such as ‘result’, ‘observation’ and ‘conclusion’, or those belonging to zones like ‘experiment’ and ‘validation’. Zones such as ‘model’ and ‘method’ and also ‘goal’ and ‘objective’ have equally such a characteristic as well.

Taking into account the aforementioned points, in this paper we propose a structure for mapping from a sentence onto a zone class using four different classifiers, and then compare these classifiers with each other. In this regard, four common and powerful classifiers including SVM, Neural Networks, LR and Bayesian Network (BN) are considered. In addition, we make use of decision-level fusion techniques to combine the decisions achieved by the aforementioned classifiers. In this regard, two well-known fusion techniques are taken into account: ‘average of probabilities (Avg)’ and ‘majority voting (Maj)’ (Mangai, Samanta, Das, & Chowdhury, 2010).

As far as, ‘Avg’ is considered, the average of the output probabilities of the classifiers is calculated and the class with the highest average is then selected as a final decision, while for ‘Maj’, the most frequently zone class which is predicted by the individual classifiers, is selected as a final decision.

Features like ‘position of a sentence in text’, ‘tense of verbs’, ‘class of previous sentence’, ‘both general and specific specialized names’, ‘highly frequent verbs’, ‘particular modes of verbs’, etc. are also taken into account. It should be noted that the huge number of features may lead to a complicated computational process and, on the other hand, irrelevant and redundant features may have a negative effect on the classification accuracy. Therefore, feature selection methods, which aim at selecting an optimized subset of features before starting the learning process, could be useful to overcome these deficits (Dasgupta, Drineas, Harb, Josifovski, & Mahoney, 2007). In the meantime, we apply Information Gain (IG) as a powerful and simple technique for filter-based feature selection. Indeed, the IG of a particular feature $t$, which might be calculated via the following formula, has the ability to measures the amount of information the presence or absence of $t$ may reflect about the
class of a sentence.

\[ IG(t) = - \sum_{i=1}^{m} P(C_i) \log P(C_i) + P(t) \sum_{i=1}^{m} P(C_i|t) \log P(C_i|t) + P(\overline{t}) \sum_{i=1}^{m} P(C_i|\overline{t}) \log P(C_i|\overline{t}), \]

where \( m \) is the number of classes, \( P(C_i) \) is the probability of the class \( C_i \), and \( P(t) \) and \( P(\overline{t}) \) denote respectively the probability of presence and absence of the feature \( t \) (Uysal, 2016).

**Features used in the suggested approach**

As discussed before, our main objective in this paper is to show how a fine classification accuracy can be obtained for zone identification through considering more significant features with less amount of computational cost. Here, features used by Liakata in her approach to zone identification, are considered as the ground for our trial. Our intention is to see whether we can replace some of these features by some other features with less computational cost but meaningful enough from some other perspectives. In the meantime, we are curious to see how through adding features with semantic richness we may compensate for a possible drop in accuracy which is resulted due to this replacement. Below, we present some details regarding these features.

- **Location**: Dividing the whole paper into 11 unequal parts and deciding to which part the given sentence belongs to. (This resembles the so-called ‘Loc’ applied by Teufel and Moens (2002); however, we refine it here by dividing the fifth part into two equal parts)
- **Heading types**: The heading of the section within which a particular sentence exists. There are 8 types of heading called Introduction, Related Works, Proposed Approach, Experimental Results, Conclusion, Abstract, Specific and None.
- **Citations, figures and tables**: The number of citations, figures and tables in a sentence.
- **Verb tense**: The tense of the main verbs in the sentence, including present, past, present perfect, past perfect and future.
- **Passiveness or activeness**: Status of passiveness or activeness of the main verbs in the sentence.
- **Adjective**: The number of superlative or comparative adjectives in a sentence.
- **First-person pronoun**: Presence of first-person pronouns in the sentence.
- **History**: The zone class of the sentence previous to a current sentence (Liakata et al., 2012).
- **Frequent verb class**: The ratio of the number of verbs in each zone to the number of sentences in this zone. Fifty highly frequent verbs in each zone have been considered in this respect.
- **Mode of verbs**: Verbs which are frequently used in the whole corpus which are manually divided into two main classes: Description verbs like ‘describe’, ‘explain’, ‘introduce’ and ‘suggest’, and Evaluation verbs like ‘evaluate’, ‘measure’, ‘increase’ and ‘test’.
- **Specialized Names**: A noun phrase is said to be a ‘general specialized name (GSN)’ if it addresses a general aspect in that domain. It is called a ‘specific specialized name (SSN)’ in case it addresses some specific aspect of an issue or a subject, such as the name of tools and methods (Asadi, Badie, & Mahmoudi, 2016). For instance, ‘hydrogen’ and ‘temperature’ are GSN and ‘fluorescence spectrum’ and ‘hydrogen bonding’ are SSN.
in the chemical field. More specifically, we extract the noun phrases from the training data and classify them into three categories by using both domain specific and domain-general ontologies. Here, we exploit ChEBI\(^1\) and Gene\(^2\) ontology as chemistry ontologies and WordNet\(^3\) as a general ontology. This is subject to the following rules in Figure 1.

**Experimental results**

**Dataset used in simulations**

In an attempt, to show the effectiveness of the suggested approach, we decided to compare it with Liakata’s approach for zone identification which has for the first time been applied to identify a variety of significant zones such as *Motivation*, *Observation*, *Method* and *Conclusion* in scientific papers like chemical tests (Soldatova & Liakata, 2007). Regarding this, ART Corpus\(^4\) was decided to be a ground for such a comparison.

The ART corpus consists of 225 papers in the field of chemistry and biochemistry and has become annotated by 20 expert chemists. It is based on CoreSC scheme that comprises the following categories: *Background* (Bac), *Goal* (Goa), *Object* (Obj), *Motivation* (Mot), *Hypothesis* (Hyp), *Method* (Met), *Model* (Mod), *Experiment* (Exp), *Observation* (Obs), *Result* (Res) and *Conclusion* (Con). Table 1 illustrates the statistics of the ART corpus.

In another attempt, we decided to have our approach compared with Fisas’s approach in the scope of computer graphics. The related corpus which is called Dr. Invertor corpus\(^5\) (DRI corpus) consists of 40 papers in the area of computer graphics and has been

---

**Algorithm 1**

```
if NP is in the acronym form then
    NP_class ← SSN

if number of tokens of NP == 1 then
    if NP is found in domain-ontology then
        NP_class ← GSN
    else if NP is found in general ontology then
        NP_class ← general
    else if at least one word in NP is found in domain-ontology then
        NP_class ← SSN
    else
        NP_class ← GSN
```

**Figure 1.** Main rules for classifying noun phrases (NP).

**Table 1.** Statistics of the ART corpus.

| Zone class | Bac | Goa | Obj | Mot | Hyp | Met | Mod | Exp | Obs | Res | Con | Total |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| Number of sentences | 6656 | 507 | 1022 | 466 | 654 | 3747 | 3456 | 2841 | 4659 | 7370 | 3077 | 34,455 |
| Percentage | 19 | 1.4 | 2.9 | 1.3 | 1.8 | 10.8 | 10 | 8.2 | 13.5 | 21.3 | 8.9 |
annotated by 3 computationally oriented linguists. The whole dataset has been divided into four subgroups each of which contains 10 papers and concerns a specific field in computer Graphics; these include ‘Skinning’, ‘Motion’, ‘Fluid simulation’ and ‘Cloth simulation’.

The scientific annotation schema includes five top-level categories and three sub-categories. Namely, the former includes Background, Challenge, Approach, Outcome and Future Work while Contribution is served as a sub-category of Outcome; moreover, Hypothesis and Goal are referred to as sub-categories of Challenge (Fisas et al., 2015; Ronzano & Saggion, 2016). Table 2 illustrates the statistics of DRI corpus.

### Analysis of simulation results

Our main goal in simulation was to show how tending to features with semantic richness as well as considering ‘highly-frequent verbs’ for each zone class (instead of co–occurrence and status of grammatical triple between verbs) can lead to a reasonable separation between the related zone classes. With regard to semantic richness, ‘specialized noun phrases’ (both general and specific) which take part in different types of zone, as well as verbs which have a particular mode like those standing for ‘description’ and ‘evaluation’, are taken into account. Following scenarios were considered for simulation:

**Scenario 1:** In order to show how far the modified features, which are more cost–effective compared to those in Liakata’s approach, can behave successfully, classification was performed with this feature but excluding ‘history’ as a feature and considering ‘specialized names’ (instead of ‘n-gram’ in Liakata’s approach) instead. The motive for such a simulation was that, extracting ‘history’ calls for a pre-tagging on the papers, which in turn is in need of intensive experience.

In this scenario, a classification accuracy of 48.8% was obtained for Art corpus; this is about 3% lower than the one obtained by Liakata. Furthermore, the classification rate for DRI corpus turned out to be 72.3% which, compared to the results of Fisas, indicates nearly 4% decrease in the accuracy.

**Scenario 2:** to show how features with semantic richness such as ‘specialized names’ and ‘mode of verbs’ can increase the classification accuracy, simulations were done with these features but not considering ‘history’ as a feature, and a classification accuracy of 50.5% was obtained. This rate is quite close to the one obtained by Liakata and its message is that features with semantic nature is good alternatives for replacing ‘history’ as a feature.

**Scenario 3:** To show to what extent ‘history’ is significant, simulations were done taking into account this feature, but avoiding ‘mode of verbs’ as a feature. A classification accuracy of 60.4% was then considering the sentences in a near neighbourhood of

### Table 2. Statistics of the DRI corpus.

| Zone class          | Number of sentences | Percentage |
|---------------------|---------------------|------------|
| Background          | 1591                | 20         |
| Challenge           | 405                 | 5          |
| Approach            | 4477                | 56         |
| Outcome             | 1259                | 16         |
| Future work         | 127                 | 1.6        |
| Total               | 7859                |            |

In this scenario, a classification accuracy of 48.8% was obtained for Art corpus; this is about 3% lower than the one obtained by Liakata. Furthermore, the classification rate for DRI corpus turned out to be 72.3% which, compared to the results of Fisas, indicates nearly 4% decrease in the accuracy.

**Scenario 2:** to show how features with semantic richness such as ‘specialized names’ and ‘mode of verbs’ can increase the classification accuracy, simulations were done with these features but not considering ‘history’ as a feature, and a classification accuracy of 50.5% was obtained. This rate is quite close to the one obtained by Liakata and its message is that features with semantic nature is good alternatives for replacing ‘history’ as a feature.

**Scenario 3:** To show to what extent ‘history’ is significant, simulations were done taking into account this feature, but avoiding ‘mode of verbs’ as a feature. A classification accuracy of 60.4% was then considering the sentences in a near neighbourhood of
a particular sentence may result in a more precise identification of zones, due to the fact that consecutive sentences lie very often in a neighbourhood of a particular sentence.

Scenario 4: Results of previous scenarios show that both 'history' and features with semantic richness play a high role in increasing the classification accuracy. This persuades us to see how co–presence of these features may lead us to a higher classification accuracy. According to this, simulations were performed and, concerning ART corpus, a classification rate of 61% was obtained which is quite remarkable compared to Liakata’s. For DRI corpus, the classification rate is in the meantime higher than that of Fisas which does not exceed 76% using SVM.

We performed a 9-fold cross-validation of LibSVM (a library for SVM) with a linear kernel (Chang & Lin, 2011). Figure 2 demonstrates the corresponding results together.

The precision, recall and F-measure of each zone class of ART corpus and DRI corpus associated to scenario 4 are, respectively, shown in Tables 3 and 4.

Analysing the results of Table 3 stands for the fact that Exp (0.8), Bac (0.69) and Mod (0.68) are of the highest F-measure, while the lowest F-measure is assigned to Mot (0.16) and Goa (0.24). Moreover, the large gap between the F-measures has possibly been raised by the unbalanced distribution of the train data assigned to every zone, besides the noises imposed by manual annotation. Despite suitable number of instances of the training data for the particular categories Obs, Res and Con, they do not correspond to a high F-measure. This may be construed in view of the fact that the sentences belonging to these three categories are close in the meaning and probably some more features of high semantic richness are needed for a better result.

Comparing the F-measures obtained from our approach to those by Liakata settles that the proposed features in this paper are substantially effective to discriminating the zone categories. Nonetheless, there exist two exceptions here: Goa and Obj. Moreover, the highest increase in the F-measures has possessed by Met, Con and Mod.

As it comes up from Table 4, Approach (0.88) and Background (0.84) are of the highest F-measure, whereas Challenge (0.48) and Future Work (0.49) are of the lowest F-measure. This happens because Approach and Background categories have high percentage of training instances (more than 76% together) while Challenge and Future Work have a very small
portion of training instances (less than 7% together). However, the number of instances in the Future Work category is less than that of Challenge and, in the meantime, it attains a higher F-measure than that of Challenge.

This exception is due to some features like ‘verb tense’ and ‘frequent verb class’ which are powerfully distinguishing between Future Work category and the other zone categories. This brings us to the fact that more strong semantic features could help characterize the zone classes significantly even though the number of instances in the training data is not enough.

It is worth noting that the absolute superiority of our approach, over that of Fisas, becomes evident by a straight-forward comparison of the F-measures for all of the predefined zones. More narrowly, the use of SVM and Simple Logistic Regression (SLR) shows that, in average, the F-measures have improved, approximately by 4% and 1%, respectively. Note that there already exist enough evidences to deduce, as Fisas does, the slightly better behaviour of SLR in comparison to that of SVM.

In Fisas’s and Liakata’s approaches, all ‘unigrams’, ‘bigrams’ and ‘trigrams’ with a frequency greater than or equal to 4 have been included in the feature vector. It should, however, be noticed that the number of aforementioned features in ART corpus is 10515, 42438 and 11854, respectively. Thus the length of the feature vector has increased substantially and this, in turn, has led to a high computational cost. However, in the suggested approach, instead of working with all these ‘n-grams’, we focused only on ‘specialized noun phrases’ (GSN and SSN) which particularly causes a sharp decrease in the length of the feature vector. Thus reducing significantly the computational cost.

### Table 3. Precision, recall and F-measure of scenario 4 and Liakata’s features for ART corpus, LIBSVM, 9-fold cross-validation, (Bold numbers indicate the highest value of F-measures).

| Zone class | Our proposed features | Liakata’s features |
|------------|-----------------------|--------------------|
|            | Precision | Recall | F-measure | Precision | Recall | F-measure |
| Obj        | 0.38      | 0.28   | 0.32      | 0.43      | 0.29   | 0.34      |
| Met        | 0.54      | 0.55   | 0.55      | 0.33      | 0.25   | 0.29      |
| Mod        | 0.68      | 0.69   | 0.68      | 0.54      | 0.52   | 0.53      |
| Mot        | 0.27      | 0.36   | 0.16      | 0.25      | 0.06   | 0.10      |
| Hyp        | 0.4       | 0.26   | 0.32      | 0.32      | 0.13   | 0.19      |
| Obs        | 0.58      | 0.58   | 0.58      | 0.53      | 0.47   | 0.50      |
| Res        | 0.55      | 0.61   | 0.58      | 0.46      | 0.57   | 0.51      |
| Exp        | 0.81      | 0.8    | 0.8       | 0.72      | 0.78   | 0.75      |
| Con        | 0.6       | 0.52   | 0.56      | 0.50      | 0.41   | 0.45      |
| Bac        | 0.66      | 0.71   | 0.69      | 0.56      | 0.68   | 0.62      |
| Goa        | 0.32      | 0.19   | 0.24      | 0.37      | 0.20   | 0.26      |
| Average    | 0.6       | 0.61   | 0.6       | 0.52      | 0.50   | 0.51      |

### Table 4. F-measure of scenario 4 and Fisa’s features for DRI corpus, LIBSVM and SLR, 9-fold cross-validation (Bold numbers indicate the highest value of F-measures)

| Zone class | Our features (LIBSVM) | Fisas’s features (LIBSVM) | Our features (SLR) | Fisas’s features (SLR) |
|------------|-----------------------|---------------------------|--------------------|------------------------|
| Background | 0.8                   | 0.73                      | 0.8                | 0.77                   |
| Challenge  | 0.48                  | 0.43                      | 0.47               | 0.46                   |
| Approach   | 0.88                  | 0.85                      | 0.88               | 0.87                   |
| Outcome    | 0.68                  | 0.62                      | 0.68               | 0.67                   |
| Future work| 0.59                  | 0.49                      | 0.63               | 0.67                   |
| Average    | 0.80                  | 0.76                      | 0.81               | 0.80                   |
essential to extract the features. Let say we just used 1300 features thus requiring only 28 minutes to train ART corpus, and 16 minutes to test a single fold.

**Experiments for feature contribution**

In order to evaluate the importance of each feature, we use a variety of feature configurations including ‘Leave-out-one-feature (LOOF)’ method and ‘Single feature classification (SFC)’. LOOF method recognizes the importance of a feature in terms of the accuracy decrease that emerges by removing it. Indeed, more important features turn out to be those features whose dropping from the computations leads to less classification accuracy. On the other hand, in SFC, the effectiveness of a particular feature, when applied singly depends on its individual role in classification accuracy.

Tables 5 and 6 show the details of experiments on ART corpus using SFC and LOOF, respectively. We perform SLR via 3-fold cross-validation.

The results show in Tables 5 and 6 demonstrate that, as mentioned before, ‘history’ has a meaningful impact on the classification accuracy. This is observed from the fact that its usage, as a single feature, results in an accuracy of 55%. Moreover, the feature of ‘heading’ proves useful to identify Exp and Mod, and ‘citation’ seems to be indispensable for identification of Obs and Bac. Also, the main role in discriminating Obj, Mot and Hyp is played by ‘verbs’ while ‘GSNs’ are fruitful to a more accurate recognition of Met, Mod and Mot.

### Table 5: SFC on ART corpus, SLR, 3-fold cross-validation (the most important features are highlighted).

| Features       | Obj. | Met. | Mod. | Mot. | Hyp. | Obs. | Res. | Exp. | Con. | Bac. | Goa. | Acc. |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| All            | 0.34 | 0.54 | 0.69 | 0.18 | 0.32 | 0.57 | 0.58 | 0.81 | 0.55 | 0.69 | 0.2  | 60.9 |
| Heading        | 0    | 0    | 0.36 | 0    | 0    | 0.45 | 0.65 | 0.34 | 0.48 | 0    |      |      |
| Verb tense     | 0    | 0    | 0    | 0    | 0    | 0.3  | 0.36 | 0.32 | 0    | 0.17 | 0    | 25   |
| History        | 0.26 | 0.52 | 0.67 | 0    | 0.28 | 0.45 | 0.52 | 0.76 | 0.5  | 0.65 | 0.07 | 55   |
| First-person pro| 0.01 | 0    | 0.03 | 0    | 0    | 0.35 | 0    | 0    | 0    | 0    |      | 21.3 |
| Mode of verb   | 0    | 0    | 0    | 0    | 0.25 | 0.31 | 0    | 0.03 | 0.3  | 0    | 0    | 23   |
| Citation/Fig/Table | 0    | 0    | 0    | 0    | 0    | 0.41 | 0.35 | 0    | 0    | 0.49 | 0    | 30.7 |
| Frequent verbs | 0.02 | 0.11 | 0.16 | 0.01 | 0.01 | 0.25 | 0.33 | 0.31 | 0.01 | 0.32 | 0.01 | 26.6 |
| SSNs           | 0.02 | 0.05 | 0.1  | 0.01 | 0    | 0.07 | 0.35 | 0.09 | 0.05 | 0.01 | 0.01 | 22.8 |
| GSNs           | 0.04 | 0.14 | 0.28 | 0.02 | 0.01 | 0.2  | 0.36 | 0.31 | 0.3  | 0.1  | 0.01 | 28   |
| Verbs          | 0.17 | 0.24 | 0.3  | 0.01 | 0.06 | 0.34 | 0.38 | 0.15 | 0.33 | 0.3  | 0.1  | 32   |

### Table 6. LOOF on ART corpus, SLR, 3-fold cross-validation (the most important features are highlighted).

| Features       | Obj. | Met. | Mod. | Mot. | Hyp. | Obs. | Res. | Exp. | Con. | Bac. | Goa. | Acc. |
|----------------|------|------|------|------|------|------|------|------|------|------|------|------|
| All            | 0.34 | 0.54 | 0.69 | 0.18 | 0.32 | 0.57 | 0.58 | 0.81 | 0.55 | 0.69 | 0.2  | 60.9 |
| -Heading       | 0.33 | 0.55 | 0.67 | 0.15 | 0.31 | 0.55 | 0.57 | 0.78 | 0.53 | 0.68 | 0.17 | 59.78|
| -Verb tense    | 0.33 | 0.55 | 0.68 | 0.16 | 0.31 | 0.57 | 0.58 | 0.79 | 0.54 | 0.69 | 0.19 | 60.6 |
| -History       | 0.33 | 0.29 | 0.47 | 0.13 | 0.17 | 0.5  | 0.48 | 0.72 | 0.41 | 0.59 | 0.18 | 49.29|
| -First-person pro| 0.3 | 0.55 | 0.69 | 0.16 | 0.31 | 0.56 | 0.58 | 0.8  | 0.54 | 0.68 | 0.16 | 60.6 |
| -Mode of verb  | 0.33 | 0.54 | 0.69 | 0.19 | 0.32 | 0.57 | 0.58 | 0.8  | 0.55 | 0.69 | 0.2  | 60.8 |
| -Citation/Fig/Table | 0.33 | 0.55 | 0.69 | 0.16 | 0.3 | 0.52 | 0.57 | 0.8  | 0.54 | 0.67 | 0.18 | 59.8 |
| -Frequent verbs| 0.31 | 0.54 | 0.68 | 0.16 | 0.31 | 0.55 | 0.57 | 0.8  | 0.54 | 0.69 | 0.17 | 60.32|
| -SSNs          | 0.33 | 0.55 | 0.69 | 0.16 | 0.32 | 0.56 | 0.58 | 0.8  | 0.55 | 0.69 | 0.18 | 60.8 |
| -GSNs          | 0.34 | 0.54 | 0.69 | 0.18 | 0.31 | 0.56 | 0.58 | 0.8  | 0.55 | 0.69 | 0.2  | 60.75|
| -Verbs         | 0.27 | 0.54 | 0.68 | 0.14 | 0.29 | 0.55 | 0.56 | 0.8  | 0.52 | 0.68 | 0.18 | 59.8 |
It is seen that the proposed features including ‘mode of verbs’, ‘frequent verbs’, ‘GSNs’ and ‘SSNs’ have enough potential to increase remarkably both accuracy and F-measure. It should be noticed that this improvement takes place at the place where we decrease the size of the feature vectors compared to the earlier work of Liakata et al. (2012).

The F-measure results of each zone categories obtained by applying SLR on DRI corpus are shown in Table 7 and Figure 3. They indicate that removing ‘history’ drops the F-measure of all zone categories, and its usage as a single feature results in an accuracy of 77.5%. Hence, one may conclude that consecutive sentences likely belong to the same zone category and this happens even more visibly in a corpus with a small number of categories.

Further analysis of Table 7 and Figure 3 reveals the usefulness of ‘SSNs’, ‘GSNs’ and ‘verbs’ in discriminating Background, Outcome and Future Work. Besides, one may also

| Features         | Background | Challenge | Outcome | Approach | Future Work | Acc    |
|------------------|------------|-----------|---------|----------|-------------|--------|
| All              | 0.79       | 0.46      | 0.68    | 0.88     | 0.63        | 81.16  |
| -Heading         | 0.79       | 0.48      | 0.65    | 0.87     | 0.6         | 80.5   |
| -Verb tense      | 0.87       | 0.44      | 0.67    | 0.88     | 0.61        | 80.8   |
| -History         | 0.66       | 0.23      | 0.56    | 0.82     | 0.24        | 72.5   |
| -First-person pro| 0.74       | 0.45      | 0.67    | 0.86     | 0.59        | 79     |
| -Mode of verb    | 0.8        | 0.47      | 0.68    | 0.88     | 0.64        | 81.39  |
| -Citation/fig/table| 0.77      | 0.45      | 0.68    | 0.88     | 0.64        | 80.7   |
| -Frequent verbs  | 0.79       | 0.45      | 0.68    | 0.88     | 0.64        | 81     |
| -SSNs            | 0.74       | 0.45      | 0.67    | 0.86     | 0.58        | 79     |
| -GSNs            | 0.74       | 0.45      | 0.66    | 0.86     | 0.63        | 78.5   |
| -Verbs           | 0.74       | 0.45      | 0.67    | 0.86     | 0.58        | 78.6   |
| -Loc             | 0.79       | 0.47      | 0.66    | 0.87     | 0.6         | 80.5   |

Figure 3. SFC on DRI corpus, SLR, 3-fold cross-validation.
notice that removing ‘First-person pronoun’ will decrease the F-measures of Background and Future Work.

In general, in comparison with ‘SSNs’, ‘GSNs’ seems to act more effectively in increasing the F-measures of the zone categories. This might be due to the fact that ‘SSNs’ are not capable of individually identifying the zone category of a particular sentence, and in many cases could be even replaced by any other special phrase without any change in the zone class of that sentence. For instance, in the sentence ‘We used the LibSVM implementation of SVM, coded in C++’, whose zone class is determined to be Exp, the underlined terms could be replaced by any other tool or method and the category of the sentence is thus subject to no change.

Regarding the above discussion, it becomes clear that our proposed features seem to work more suitably for DRI corpus. This can be construed by the problematic and long-term process of zone identification over the finer-grained ART corpus and in addition to that, somewhat different annotations made by the experts with regard to some of the zones such as Goa and Obj, or Obs and Res.

Another relevant factor that affects the improper presentations of the roles of ‘SSNs’ and ‘GSNs’ is lack of some ontological structures that can fully contain the terms in the domain of corpora. Indeed, despite their specialty, some of the noun phrases could not be found in our selected ontologies.

**Experiments for evaluation of fusion performance**

To evaluate the impact of fusion on the accuracy of the sentence classification, four common classifiers have been selected. In this regard, four Weka libraries including LibSVM with linear kernel, ‘multi-layer perception (MLP)’ with four hidden layer, SLR and ‘BN’ have been chosen. Table 8 shows the results of 3-fold cross-validation for ART and DRI corpora on the classifiers mentioned above.

SLR and LibSVM have turned out to be of the highest accuracy and F-measure, respectively. SLR is pretty robust to noise and avoids overfitting, and SVM with linear kernel behaves rather in the same manner. The results of MLP classifier are quite weak; this is due to the high number of features that, accordingly, will increase the number of input and hidden neurons. This leads to a complicated neural network that, in turn, may increase the possibility of overfitting effect and might get stuck in a local minima. Even though BN has turned out to be of a lower accuracy in comparison with SVM and SLR, its coverage velocity is higher than that of the other classifiers and, taking into account its simplicity, its performance seems fairly well.

In WEKA, there are a variety of voting approaches to be applied as fusion techniques. Amongst all, we choose two well-known combining rules including ‘Maj’ and ‘Avg’ that,

| Classifier | ART corpus | DRI corpus |
|------------|------------|------------|
|            | Precision  | Recall     | F-measure | Accuracy | Precision | Recall | F-measure | Accuracy |
| SVM        | 60         | 60.5       | 60.1      | 60.47    | 80.4      | 81     | 80.5      | 80.96    |
| MLP        | 20.1       | 34.5       | 22.5      | 34.5     | 72.7      | 76.5   | 74.4      | 76.5     |
| SLR        | 60.1       | 60.9       | 60.2      | 60.9     | 80.5      | 81.2   | 80.6      | 81.16    |
| BN         | 58.3       | 58.8       | 58.3      | 58.7     | 78.9      | 78.1   | 78.4      | 78.1     |
compared to the other approaches, usually report better results. **Table 9** shows the results made by ‘Maj’ and ‘Avg’ for all potential 3-combinations of the aforementioned classifiers.

As it is shown in **Table 9**, the best combination of the classifiers for ART corpus has been achieved by fusing SVM, SLR and BN using majority voting algorithm. Here, the upshot is an accuracy of 61.7, which yields a slight improvement in the one achieved by the most accurate individual classifier, SLR. On the other hand, the combination of the decisions of SVM, SLR and MLP, using ‘Maj’, achieves the accuracy of 81.9 for DRI corpus which is a little bit more than the accuracy obtained by SLR individually (81.16). As a summary, ‘Maj’ and ‘Avg’ behave almost in the same way; any of which could reflect slightly better results on a certain combination of the classifiers. Nonetheless, comparing the results shown in **Tables 8 and 9**, may convince us that there still exist situations where both of the aforementioned techniques may fail to improve the classification accuracy.

**Experiments for evaluation of feature selection**

Although in the present study the number of features has significantly been reduced from more than 40,000 (used in Liakata’s approach) to 1300, this is still problematic with regard to classification time, thus leaving the online zone identification still as a serious problem. Such a fact motivates the use of IG as a feature selection method to choose the top 100 features. Amongst the aforementioned top features, there exist some particularly important ones like ‘history’, ‘Loc’, ‘verb tense’, ‘heading type’, ‘citation’, ‘passive’, ‘figure’, ‘first-person pronoun’, etc. The most important ‘GSNs’, ‘SSNs’ and ‘verbs’ that are considered as part of the 100-top features of ART and DRI corpora are listed in **Tables 10 and 11**, respectively, which are more conceptual than Liakata’s n-grams.

Note that the experiments described in the previous subsection have been done on this reduced feature set for which the results are shown in **Table 12**.

**Table 9.** The accuracy of the 3-combinations of the classifiers.

| 3-combinations of classifiers | ART corpus | DRI corpus |
|------------------------------|------------|------------|
|                              | Maj        | Avg        | Maj        | Avg        |
| SVM + MLP + SLR              | 60.5       | 60.4       | 81.9       | 81.55      |
| SVM + MLP + BN               | 60.8       | 60         | 80.7       | 81.25      |
| MLP + SLR + BN               | 59.9       | 60.9       | 80.8       | 81         |
| SVM + SLR + BN               | 61.7       | 60.7       | 81.55      | 81.55      |

**Table 10.** The most important GSNs, SSNs, and verbs on ART corpus.

| Verbs         | Use, show, observe, suggest, expect, investigate, perform, study, indicate, assume, see, report, prepare, conclude, consider, define, increase, decrease |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GSNs          | Eqn, result, band, interaction, term, mechanism, co-workers, state, increase, structure, formation, fact, system, molecule                                                                                           |
| SSNs          | Metal ion, matrix element, basis set, reaction mixture, wave function                                                                                                                                   |

**Table 11.** The most important GSNs, SSNs and verbs on DRI corpus.

| Verbs         | Show, present, demonstrate, define, have, use, propose, improve, compose, be, introduce, coordinate                                                                                                    |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GSNs          | Point, matrix, position, vector, edge, second, application, velocity, result, skinning, behaviour, term                                                                                                |
| SSNs          | Shape interpolation, skin deformation, graph walk, projection method, space deformation, SSD, SBS, cloth capture                                                                                   |
The results exhibited in Table 8 and Table 12 settle that the use of feature selection leads, on the one hand, to a slight improvement in accuracy and, on the other hand, to a meaningful decrease in the evaluation time; for instance we require only three minutes to train ART corpus using LibSVM and two minutes to evaluate a single fold.

Moreover, these results illuminate that the high number of features might cause overfitting effect and, despite the long-term learning and testing process, one may not be able to acquire a satisfactory precision. Finally, a sharp improvement in the accuracy of MLP proves our assumption about overfitting occurrences.

Table 13 shows the results came up by using voting approaches for all possible 3-combinations of the aforementioned classifiers after applying feature selection.

Concluding remarks and future prospects

In this paper, we demonstrated that how, through considering features with high semantic richness such as specialized names and mode of verbs, one may attain a higher classification accuracy (with regard to zone identification) for the sentences in a text at the place where features with less computational cost are being used. This seems to be mainly because features with high semantic richness have principally the ability to participate effectively in classification with less need for involving highly syntactical features which in turns call for high computational cost. Taking this point into account a deeper investigation on the linguistic features with high semantic richness is expected to eventually lead to higher performance. In order to perform zone classification, we made use of SVM, Neural Networks, LR and BN classifiers amongst which, LR and SVM have turned out to be of the highest accuracy. To improve the zone identification accuracy, the classifiers were then fused to combine the decisions made by individual classifiers.

**Table 12.** The accuracy of the four classifiers after attribute selection (Bold letters indicate the most accurate classifiers)

| Classifier | ART corpus | DRI corpus |
|------------|------------|------------|
| SVM        | 60.64      | 81.46      |
| MLP        | 58.5       | 80.28      |
| SLR        | 61         | 81.28      |
| BN         | 58.5       | 78.1       |

**Table 13.** Accuracy of the 3-combinations of the classifiers after attribute selection.

| 3-combination of the classifiers | ART corpus | DRI corpus |
|----------------------------------|------------|------------|
|                                   | Maj | Avg | Maj | Avg |
| SVM + MLP + SLR                  | 61.5 | 60.84 | 81.67 | 81.54 |
| SVM + MLP + BN                   | 61  | 61.23 | 81.65 | 81.52 |
| MLP + SLR + BN                   | 60.5 | 60.7  | 81.4 | 81.5  |
| SVM + SLR + BN                   | 60.54 | 60.9  | 81.6 | 81.6  |
particular, combining SVM, LR and Neural Network, besides using the ‘majority voting’ technique, was shown to yield an increase in the classification accuracy. In the meantime it was shown that, through selecting features based on IG, we were able to reduce reasonably both the training and testing time with no affect on the classification accuracy.

Since a zone identity manifests highly in a set of neighbouring sentences, it would therefore be more reasonable to perform classification on the ground of fusion between the local decisions belonging to the neighbouring sentences. Realizing such an objective can be viewed as an essential research work in future.

Notes

1. https://www.ebi.ac.uk/chebi/
2. http://www.geneontology.org/
3. https://wordnet.princeton.edu/
4. https://www.aber.ac.uk/en/cs/research/cb/projects/art/art-corpus/
5. http://sempub.taln.upf.edu/dricorpus

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Nasrin Asadi is currently a Ph.D student at ICT Research Institute, Tehran, Iran. She received her B.Sc. in the field of software engineering from Amirkabir University in 2007 and her M.Sc. in software engineering from Shiraz University in 2010. Her current research interests include text mining, text summarization and natural language processing.

Kambiz Badie graduated from Alborz high school in Tehran and received all his degrees from Tokyo Institute of Technology, Japan, majoring in pattern recognition.

Within the past years, he has been actively involved in doing research in a variety of issues, such as machine learning, cognitive modeling, and knowledge processing and creation in general and analogical knowledge processing, experience modeling interpretation process in particular, with emphasis on creating new ideas, techniques and contents.

Out of the frameworks developed by Dr Badie, ‘interpretative approach to analogical reasoning’, ‘viewpoint oriented manipulation of concepts’, ‘semantic transformation for text interpretation purposes’ and ‘schema satisfaction reasoning’, are particularly mentionable as novel approaches to creative idea generation, which in tum have a variety of application in developing novel scientific frameworks as well as creating potential pedagogical and research support contents.

Dr Badie is one of the active researchers in the areas of interdisciplinary and transdisciplinary studies in Iran, and has a high motivation for applying intelligent/modeling methodology to the human issues. At present, he is deputy for research affairs in ICT Research Institutes, an affiliated professor at Faculty of Engineering Science in the University of Tehran, and in the meantime, the editor-in-chief of International journal of information & Communication Technology Research (IJICTR) being published periodically by this institute.

Maryam Tayefeh Mahmoudi holds a B.Sc. in Computer Engineering-Software, a B.A. in Business Management, an M.Sc. in Computer Engineering-Software from Iran University of Science & Technology, and a Ph.D degree in Artificial Intelligence from the University of Tehran with emphasis on intelligent organization of educational contents.
Within the past years, she has been involved in a variety of research works at Knowledge Management & e-Organization and Multimedia Systems Research Groups of IT Research Faculty in ICT Research Institute (ex ITRC), working on issues like automatic generation and personalization of ideas and contents, decision support and recommendation systems for research & education purposes, augmented reality for added-value multimedia content generation, as well as making conceptual models for IT research projects with emphasis on using ontological structures. She is a co-author of many research papers in different international journals and proceedings of conferences.

Dr Mahmoudi is an active researcher in the areas of content management & creation and augmented reality and has a high motivation for applying the related techniques to human issues such as human–computer interaction and e-pedagogy/e-learning as well. At present, she is an Assistant Professor at ICT Research Institutes, a senior member of IEEE, IEEE WIE Committee chair, and in the meantime, the editor-in-chief of Iran e-learning Association’s Newsletter.

**ORCID**

Kambiz Badie [http://orcid.org/0000-0003-1468-7010]

Nasrin Asadi [http://orcid.org/0000-0002-5048-4125]

**References**

Agarwal, S., & Yu, H. (2009). Automatically classifying sentences in full-text biomedical articles into introduction, methods, results and discussion. *Bioinformatics (Oxford, England)*, 25, 3174–3180.

Asadi, N., Badie, K., & Mahmoudi, M. T. (2016). Identifying categories of zones in scientific papers based on lexical and syntactical features. Second International Conference on Web Research (ICWR), April 2016, (pp. 177–182), Tehran, Iran.

Badie, K., Asadi, N., & Mahmoudi, M. T. (2017). A new approach to zone identification based on considering features with high semantic richness. IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA), Jun 2017, (pp. 443–448), Gdynia, Poland.

Bonn, S. V., & Swales, J. M. (2007). English and French journal abstracts in the language sciences: Three exploratory studies. *Journal of English for Academic Purposes*, 6(2), 93–108.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, 27.

Dasgupta, A., Drineas, P., Harb, B., Josifovski, V., & Mahoney, M. W. (2007). Feature selection methods for text classification. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (pp. 230–239), San Jose, California, USA.

Fisas, B., Saggion, H., & Ronzano, F. (2015). On the discoursive structure of computer graphics research papers. *LAW@ NAAACL-HLT*, (pp. 42-51).

Groza, T. (2013). Using typed dependencies to study and recognise conceptualisation zones in biomedical literature. *PloS one*, 8, e79570.

Groza, T., Hassananzadeh, H., & Hunter, J. (2013). Recognizing scientific artifacts in biomedical literature. *Biomedical Informatics Insights*, 6, 15.

Guo, Y., Korhonen, A., Liakata, M., Karolinska, I. S., Sun, L., & Stenius, U. (2010). Identifying the information structure of scientific abstracts: An investigation of three different schemes. Proceedings of the 2010 Workshop on Biomedical Natural Language Processing, (pp. 99–107), Uppsala, Sweden.

Guo, Y., Korhonen, A., & Poibeau, T. (2011). A weakly-supervised approach to argumentative zoning of scientific documents. Proceedings of the Conference on Empirical Methods in Natural Language Processing, (pp. 273–283), Edinburgh, United Kingdom.

Guo, Y., Reichart, R., & Korhonen, A. (2015). Unsupervised declarative knowledge induction for constraint-based learning of information structure in scientific documents. *Transactions of the Association for Computational Linguistics*, 3, 131–143.
Hong, e. a. (2007). Development, implementation, and a cognitive evaluation of a definitional question answering system for physicians. *Journal of Biomedical Informatics, 40*, 236–251.

Kilicoglu, Halil. (2017). Biomedical text mining for research rigor and integrity: Tasks, challenges, directions. *Briefings in Bioinformatics, 20*, 1–20. http://dx.doi.org/10.1093/bib/bbx057. preprint

Liakata, M. (2010). Zones of conceptualisation in scientific papers: A window to negative and speculative statements. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, (pp. 1–4).

Liakata, M., Saha, S., Dobnik, S., Batchelor, C., & Rebholz-Schuhmann, D. (2012). Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics (Oxford, England), 28*, 991–1000.

Mangai, U. G., Samanta, S., Das, S., & Chowdhury, P. R. (2010). A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Technical Review, 27*(4), 293–307.

McKnight, L., & Srinivasan, P. (2003). Categorization of sentence types in medical abstracts. *AMIA Annu Symp*, (pp. 440–444).

Mizuta, Y., & Collier, N. (2004). Zone identification in biology articles as a basis for information extraction. Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, (pp. 29–35), Geneva, Switzerland.

Ronzano, F., & Saggion, H. (2016). *Knowledge extraction and modeling from scientific publications*. International Workshop on Semantic, Analytics, Visualization, (pp. 11–25).

Séaghdha, D., Ó., & Teufel, S. (2014). *Unsupervised learning of rhetorical structure with un-topic models*. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, (pp. 2-13), Dublin, Ireland.

Soldatova, L., & Liakata, M. (2007). *An ontology methodology and CISP-the proposed core information about scientific papers*. JISC Project Report.

Teufel, S., & others. (2000). *Argumentative zoning: Information extraction from scientific text*. Ph.D. dissertation, University of Edinburgh.

Teufel, S., & Kan, M. Y. (2011). Robust Argumentative Zoning for Sensemaking in Scholarly Documents. In R. Bernardi, S. Chambers, B. Gottfried, F. Segond, & I. Zaihrayeu (Eds.), *Advanced Language Technologies for Digital Libraries. Lecture Notes in Computer Science* (Vol. 6699). Berlin, Heidelberg: Springer.

Teufel, S., & Moens, M. (2002). Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics, 28*(4), 409–445.

Uysal, A. K. (2016). An improved global features election scheme for text classification. *Expert Systems With Applications, 43*, 82–92.