

Paper Presentation: CS671

A Weakly-Supervised Approach to Argumentative Zoning of Scientific Documents

Yufan Guo

Anna Korhonen

Thierry Poibeau

1

Review By:
Pranjal Singh
10511

Overview

- The paper talks about annotation of documents according to certain broad classes such as objective, methodology, results obtained, etc.
- The paper basically investigates the performance of weakly-supervised learning for Argumentative Zoning(AZ) of scientific abstracts.
- Uses weakly-supervised learning which is less expensive as compared to fully supervised approach.

What is Argumentative Zoning??

- AZ is an approach to information structure which provides an analysis of the rhetorical progression of the scientific argument in a document (Teufel and Moens, 2002).
- Because of the utility of this method in several domains (such as summarization, computational linguistics), weakly-supervised approach is much more practical.

Architecture

Data

- Guo et al. (2010) provide a corpus of 1000 biomedical abstracts (consisting of 7985 sentences and 225785 words) annotated according to three schemes of information structure(section names, AZ and Core Scientific Concepts)
- The paper uses Argumentative Zoning (Mizuta et al., 2006)
- According to Cohens kappa (Cohen, 1960) the inter-annotator agreement is relatively high: $\text{kappa}(k) = 0.85$.

Architecture

Methods

- Comparison between supervised classifier (SVM and CRF) with four weakly supervised classifiers:
- Two based on semi-supervised learning (transductive SVM and semi-supervised CRF) and
- Two on active learning (Active SVM alone and in combination with self-training).

Zones in AZ

Category	Abbr.	Definition
Background	BKG	The circumstances pertaining to the current work, situation, or its causes, history, etc.
Objective	OBJ	A thing aimed at or sought, a target or goal
Method	METH	A way of doing research, esp. according to a defined and regular plan; a special form of procedure or characteristic set of procedures employed in a field of study as a mode of investigation and inquiry
Result	RES	The effect, consequence, issue or outcome of an experiment; the quantity, formula, etc. obtained by calculation
Conclusion	CON	A judgment or statement arrived at by any reasoning process; an inference, deduction, induction; a proposition deduced by reasoning from other propositions; the result of a discussion, or examination of a question, final determination, decision, resolution, final arrangement or agreement
Related work	REL	A comparison between the current work and the related work
Future work	FUT	The work that needs to be done in the future

An example of an annotated abstract

Butadiene (BD) metabolism shows gender, species and concentration dependency, making the extrapolation of animal results to humans complex. BD is metabolized mainly by cytochrome P450 2E1 to three epoxides, 1,2-epoxy-3-butene (EB), 1,2;3,4-diepoxybutane (DEB) and 1,2-epoxy-butanediol (EB-diol). For accurate risk assessment it is important to elucidate species differences in the internal formation of the individual epoxides in order to assign the relative risks associated with their different mutagenic potencies. Analysis of N-terminal globin adducts is a common approach for monitoring the internal formation of BD. **Background** es. Our long term strategy is to develop an LC-MS/MS method for simultaneous detection of all three BD hemoglobin adducts. This approach is modeled after the recently reported immunoaffinity LC-MS/MS method for the cyclic N,N-(2,3-dihydroxy-1,4-butanediol)-valine (pyr-Val, derived from DEB). We report herein the analysis of the EB-derived 2-hydroxyl-3-butenyl-valine peptide. **Objective** 1). The procedure utilizes trypsin hydrolysis of globin and immunoaffinity (IA) purification of alkylated heptapeptides. Quantitation is based on LC-MS/MS monitoring of the transition from the singly charged molecular ion of HB-Val (1-7) to the a(1) fragment. Human HB-Val (1-11) was synthesized and used for antibody production. As internal standard, the labeled rat-[(13)C(5)(15)N]-Val (1-11) was prepared through direct alkylation of the corresponding peptide with EB. Standards were characterized and quantified by LC-MS/MS and LC-UV. The method was validated with different amounts of human HB-Val standard. The recovery was >75% and coefficient of variation <25%. The LOQ was set to 100 fmol/injection. For a proof of principal experiment, globin samples from male and female rats exposed to 1000 ppm BD for 90 days were analyzed. **Method** ed. The amounts of HB-Val present were 268.2+/-56 and 350+/-70 pmol/g (mean+/-S.D.) for males and females, respectively. No HB-Val was detected in control samples. **Results**. These data are much lower compared to previously reported values measured by GC-MS/MS. **Related work** MS. The difference may be due higher specificity of the LC-MS/MS method to the N-terminal peptide from the alpha-chain versus derivatization of both alpha- and beta-chain by Edman degradation, and possible instability of HB-Val adducts during long term storage (about 10 years) between samples. **Conclusion** es. These differences will be resolved by examining recently collected samples, using the same internal standard for parallel analysis by GC-MS/MS. **Future work** MS. Based on our experience with pyr-Val adduct assay we anticipate that this assay will be suitable for evaluation of HB-Val in multiple species.

Methodology

Feature Extraction

- Documents were first identified by their features which included Zones (e.g. abstracts were divided into ten parts where zones typically occur)
- Words, Bi-grams, Verbs, Verb Class, POS, Grammatical Relation, Subject & Object, Voice of Verbs.
- These were extracted using tools such as tokenizer(detects boundaries in a sentence), C&C Tool(POS tagging, lemmatization and parsing), unsupervised spectral clustering method(to acquire verb classes).
- The lemma output was used to create Word, Bi-gram and Verb features.

SVM(Support Vector Machine)

- Supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis
- The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier.

CRF(Conditional Random Field)

- A form of discriminative modelling
 - Has been used successfully in various domains such as part of speech tagging and other Natural Language Processing tasks
- Processes evidence bottom-up
 - Combines multiple features of the data
 - Builds the probability $P(\text{sequence} \mid \text{data})$

Results

- Only one method(ASVM) identifies six out of the seven possible categories. Other methods identify five categories.
- RESULTS has the highest amount of data and OBJECTIVE has the minimum.
- The LOCATION feature is found to be the most important feature for ASSVM.
- Voice, Verb class and GR contribute to general performance.
- Least helpful features are Word, Bi-gram and Verb because they suffer from sparse data problems.

Results

	Acc. F-score								
		MF	BKG	OBJ	METH	RES	CON	REL	FUT
SVM	.77	.74	.84	.68	.71	.82	.64	-	-
CRF	.70	.65	.75	.46	.48	.78	.76	-	-
ASVM	.80	.75	.88	.56	.68	.87	.78	.33	
ASSVM	.81	.76	.86	.56	.76	.88	.76	-	-
TSVM	.76	.73	.84	.61	.71	.79	.71	-	-
SSCRF	.73	.67	.76	.48	.52	.81	.78	-	-

References

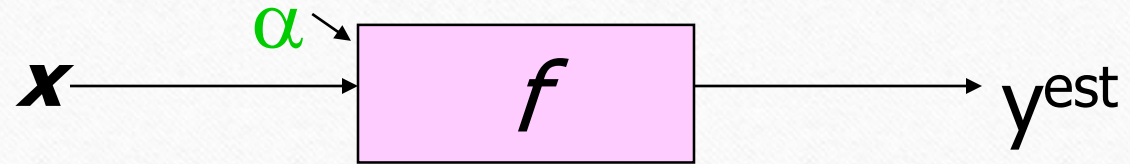
- Yufan Guo, Anna Korhonen, and Thierry Poibeau. **A Weakly-Supervised Approach to Argumentative Zoning of Scientific Documents.** In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 273–283, 2011.
- Simone Teufel and Marc Moens. **Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status.** *Comput. Linguist.*, 28(4):409–445, December 2002.

**Thank
You!!!**

Future Work

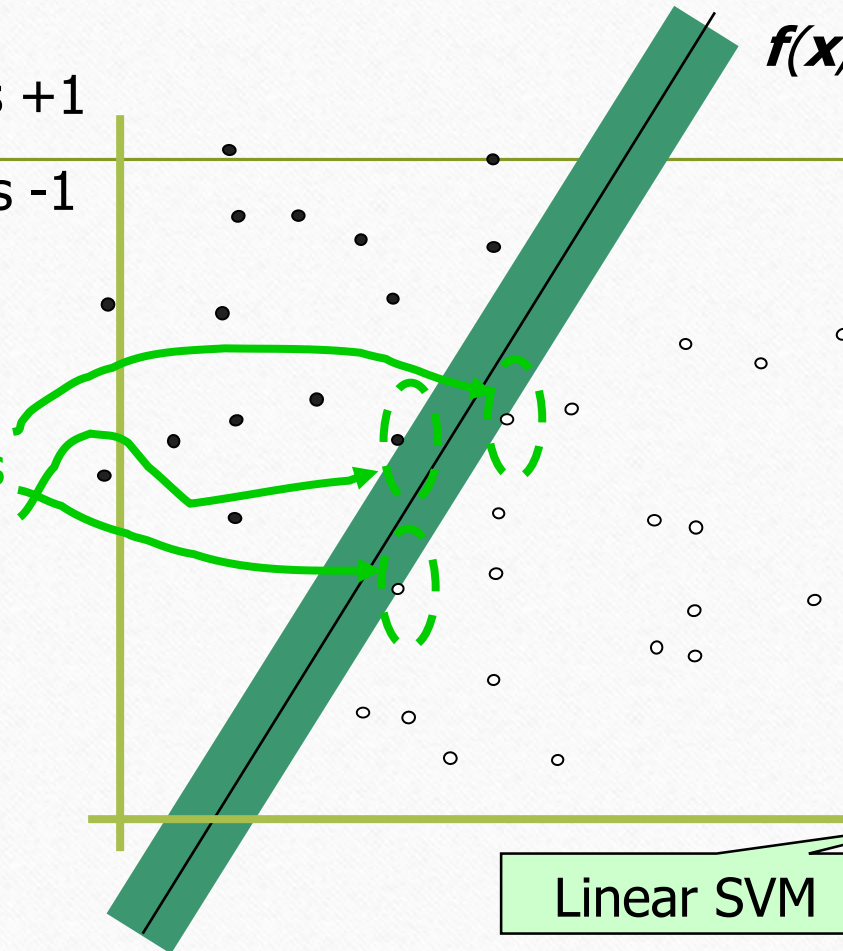
- The approach to active learning could be improved in various ways by experimenting with more complex query strategies such as margin sampling algorithm by (Scheffer et al., 2001) and query-by-committee algorithm by (Seung et al., 1992).
- Also, looking for other optimal features could improve the result a lot.
- It could also be very interesting to evaluate the usefulness of weakly-supervised identification of information structure for NLP tasks such as summarization and information extraction and for practical tasks such as manual review of scientific papers for research purposes.

Maximum Margin



- denotes +1
- denotes -1

Support Vectors
are those
datapoints that
the margin
pushes up
against



$$f(\mathbf{x}, \mathbf{w}, b) = \text{sign}(\mathbf{w} \cdot \mathbf{x} - b)$$

The **maximum margin linear classifier** is the linear classifier with the, um, maximum margin.

This is the simplest kind of SVM (Called an LSVM)

Copyright © 2001,
2003, Andrew W.

Moore

Linear SVM

CRF(Conditional Random Field)

- Each attribute of the data we are trying to model fits into a *feature function* that associates the attribute and a possible label
 - A positive value if the attribute appears in the data
 - A zero value if the attribute is not in the data
- Each feature function carries a *weight* that gives the strength of that feature function for the proposed label
 - High positive weights indicate a good association between the feature and the proposed label
 - High negative weights indicate a negative association between the feature and the proposed label
 - Weights close to zero indicate the feature has little or no impact on the identity of the label

Discussion

- Almost all the methods performed as they have been performing in other domains and on work done by other authors.
- But TSVM did not perform better than SVM with the same amount of labelled data. This could be due to higher dimensional data in this work as compared to other works.
- SSCRF did not perform as expected on the data may be due to less number of labelled and unlabelled instances.

Methodology

Machine Learning Methods

- SVM and CRF were used as Supervised methods on the data obtained after feature extraction.
- Active SVM: Starts with a small amount of labelled data, and iteratively chooses a proportion of unlabelled data for which SVM has less confidence to be labelled and used in the next round of learning,
- Active SVM with self-training
- Transductive SVM: Takes advantage of both labelled and unlabelled data
- Semi-supervised CRF were used as Weakly-Supervised Methods.

Results

- With 10% training data, ASSVM performs best with 81% accuracy and macro F-score of .76.
- ASVM performs with accuracy of 80% and F-score of .75. Both of them outperform supervised SVM.
- TSVM is the worst performing SVM-based method with an accuracy of 76% and F-score of .73 which is less than supervised SVM.
- But it outperforms both CRF-based methods.