Automatic identification of general and specific sentences by leveraging discourse annotations

Anne Louis

Ani Nenkova

Iulia Banghea

Introduction

Sentences in text differ by how much specific context they have.

There are two types distinguished in this paper:

- general
- specific

The aim of this paper is to introduce an approach to automatically distinguish between the two.

Introduction - an example

The Booker prize has, in its 26-year history, always provoked controversy.

The novel, a story of Scottish low-life narrated largely in Glaswegian dialect, is unlikely to prove a popular choice with booksellers who have damned all six books shortlisted for the prize as boring, elitist and - worst of all - unsaleable.

Applications

- **Prediction of writing quality**

  Prescriptive books on writing advise that sentences that make use of vague and abstract words should be avoided or else immediately followed by specific clarifications

- **Text generation systems**

  Used to control the type of content produced

- **Information extraction systems**

  Used to extract different types of information

Classification

Supervised classifier

Training Data: Penn Discourse Treebank (PDTB)

Classification

Supervised classifier

Training Data: Penn Discourse Treebank (PDTB)

- problem not addressed in prior work
- no existing corpus annotated for specificity

Classification

Supervised classifier

Training Data: Penn Discourse Treebank (PDTB)

- problem not addressed in prior work
- no existing corpus annotated for specificity

Solution?

- Exploit indirect annotations of discourse relation distinctions

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification **Instantiation**

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	**Instantiation**
Applies when Arg2 describes the situation in Arg1 in more detail	

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	**Instantiation**
Applies when Arg2 describes the situation in Arg1 in more detail Typical connectives: “specifically”, “indeed”, “in fact”	

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	**Instantiation**
Applies when Arg2 describes the situation in Arg1 in more detail Typical connectives: “specifically”, “indeed”, “in fact”	Arg1 evokes a set and Arg2 describes it in further detail

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	Instantiation
Applies when Arg2 describes the situation in Arg1 in more detail	Arg1 evokes a set and Arg2 describes it in further detail
Typical connectives: “specifically”, “indeed”, “in fact”	Typical connectives: “for example”, “for instance”, “in particular”

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	Instantiation
Applies when Arg2 describes the situation in Arg1 in more detail	Arg1 evokes a set and Arg2 describes it in further detail
Typical connectives: “specifically”, “indeed”, “in fact”	Typical connectives: “for example”, “for instance”, “in particular”
Alan Spoon, recently named Newsweek president said Newsweek’s ad rates would increase 5% in January. ***(<u>Implicit</u> = SPECIFICALLY)*** A full, four-color page in Newsweek will cost $100,980.	

Training Data

Use certain types of discourse relations annotated in the Penn Discourse Treebank (PDTB)

Specification	**Instantiation**
Applies when Arg2 describes the situation in Arg1 in more detail	Arg1 evokes a set and Arg2 describes it in further detail
Typical connectives: "specifically", "indeed", "in fact"	Typical connectives: "for example", "for instance", "in particular"
Alan Spoon, recently named Newsweek president said Newsweek's ad rates would increase 5% in January. **(Implicit = *SPECIFICALLY*)** A full, four-color page in Newsweek will cost $100,980.	*Despite recent declines in yields, investors continue to pour cash into money funds.* **(Implicit = *FOR INSTANCE*)** Assets of the 400 taxable funds grew by $1.5 billion during the last week.

Features

Sentence Length

General sentences are usually shorter than the specific ones

Features: number of words and nouns

Features

Sentence Length

General sentences are usually shorter than the specific ones

Features: number of words and nouns

Polarity

Sentences with strong opinion are typically in the general category

Features: count of positive / negative / polar words

each of the above normalized by sentence length

Features

Sentence Length

General sentences are usually shorter than the specific ones

Features: number of words and nouns

Polarity

Sentences with strong opinion are typically in the general category

Features: count of positive / negative / polar words
each of the above normalized by sentence length

Specificity

Specific sentences are more likely to contain specific words and details

Features: Hypernym Relations from WordNet
Inverse Document Frequency (idf) for a word

Features

NE+CD

Specific sentences often contain numbers and dollar amounts while general sentences have more plural nouns

Features: count of numbers, proper names and dollar signs
number of plural nouns

Features

NE+CD

Specific sentences often contain numbers and dollar amounts while general sentences have more plural nouns

Features: count of numbers, proper names and dollar signs
number of plural nouns

Language Model

General sentences often contain unexpected, catchy words or phrases

Features: Using the unigram, bigram and trigram language model, we obtain the log probability and perplexity of the sentences to use as features

Features

Syntax

General sentences show frequent usage of qualitative words such as adjectives and adverbs

Features: counts of adjectives, adverbs, adjective / adverbial phrases
number of verb phrases and their average length in words
number of prepositional phrases

Features

Syntax

General sentences show frequent usage of qualitative words such as adjectives and adverbs

Features: counts of adjectives, adverbs, adjective / adverbial phrases
number of verb phrases and their average length in words
number of prepositional phrases

Words

Features: count of each word in the sentence as a feature, excluding words not seen in training

Results

Two classifiers for distinguishing general and specific sentences:

- one trained on sentences from **Instantiation** relations
- one trained on sentences from **Specification** relations

Logistic Regression classifier was trained on the features

- a probability measure is more appropriate to associate with each sentence rather than hard classification into the two classes

Predictions are evaluated using 10-fold cross-validation

Results

Features	**Instantiations**	**Specifications**
NE+CD	68.6	56.1
language models	65.8	55.7
specificity	63.6	57.2
syntax	63.3	57.3
polarity	63.0	53.4
sentence length	54.0	57.2
all non-lexical	75.0	62.0
lexical (words)	74.8	59.1
all features	75.9	59.5

The Instantiation classifier yields better results than the Specifications classifier

Highest accuracy comes from combining all features, reaching 75%

The individually best class features are words

Feature Analysis

Top word features for the two types of sentences, which appear in at least 25 training examples:

- **General:** number, but, also, however, officials, some, what, prices, made
- **Specific:** one, a, to, co, i, called, we, could, get, and, first, inc

Feature Analysis

Top word features for the two types of sentences, which appear in at least 25 training examples:

- **General:** number, but, also, however, officials, some, what, prices, made
- **Specific:** one, a, to, co, i, called, we, could, get, and, first, inc

Feature intuitions were mostly true:

- Numbers and names are predictive of specific sentences
- Plural nouns are a property of general sentences
- Dollar signs (expected to be more likely with specific sentences) turned out to be more frequent with the other category
- For language model features, general sentences tended to have a lower probability and higher perplexity than specific ones

Testing on new sentences

Selection of articles from:

- Three WSJ articles from the PDTB corpus
- Six Associated Press articles from the AQUAINT corpus
- Two Financial Times articles from the AQUAINT corpus

Testing on new sentences

Selection of articles from:

- Three WSJ articles from the PDTB corpus
- Six Associated Press articles from the AQUAINT corpus
- Two Financial Times articles from the AQUAINT corpus

Annotations were obtained manually using Amazon's Mechanical Turk (MTURK)

Testing on new sentences

Selection of articles from:

- Three **WSJ** articles from the **PDTB** corpus
- Six **Associated Press** articles from the **AQUAINT** corpus
- Two **Financial Times** articles from the **AQUAINT** corpus

Annotations were obtained manually using Amazon's Mechanical Turk (MTURK)

Users were presented with one sentence at random and three options for classifying it

- general
- specific
- can't decide

Testing on new sentences

	WSJ articles			AP articles		
Agree	**total**	**gen**	**spec**	**total**	**gen**	**spec**
5	96	51	45	108	33	75
4	102	57	45	91	35	56
3	95	52	43	88	49	39
undecided	1			5		
Total	294	160	133	292	117	170

Annotator agreement

Results

Examples	Size	All features	Nonlexical	Words
		WSJ sentences		
Agreement 5	96	90.6	96.8	84.3
Agreement $4+5$	198	80.8	88.8	77.7
Agreement $3+4+5$	293	73.7	76.7	71.6

Examples	Size	All features	Nonlexical	Words
		AP sentences		
Agreement 5	108	69.4	94.4	78.7
Agreement $4+5$	199	65.8	89.9	74.8
Agreement $3+4+5$	287	59.2	81.1	67.5

Non-lexical features give the best performance on both sets of articles

The **word features** now give more than 10% lower accuracy than non-lexical features

Lexical features probably do not cover all example types

Non-lexical features provide better abstraction and portability across corpora

Accuracy increases on examples with **higher agreement** (over 90% for sentences with full agreement)

Task based evaluation

User created general or specific summaries for a set of articles

- Conveying only the general ideas
- Providing specific details about the topic

Data

- Summaries and source texts from the Document Understanding Conference (DUC)
- Each input consists of 25 to 50 news articles on a common topic
- The input texts and topic statements were given to trained NIST assessors for writing summaries
- Resulted in a roughly equal distribution of general (146) and specific (154) summaries

Results

Text	General category	Specific category
Summaries	0.55 (0.15)	0.63 (0.14)
Inputs	0.63 (0.06)	0.65 (0.04)

Mean value (and standard deviation) of specificity levels for inputs and summaries

Used the classifier from the Instantiation relations and extra annotations with a combination of all features

A specificity level was assigned for each summary

For **specific summaries**, the mean specificity is 0.63
For **general summaries**, the mean specificity is only 0.55

For a **two sided t-test** the difference in values has a p-value of 1.5e-06

This result shows that the predictions can distinguish the two types of summaries

Conclusions

- A new task—identification of general and specific sentences
- Discourse relations can be used as training data for the task
- Features such as polarity, word specificity, language models, entity-related and lexical features
- High classification performance, 25% absolute increase over the baseline
- The classifier also provides a graded score for specificity and can distinguish general and specific summaries written by people