

Extracting Case Law Sentences for Interpretation of Terms from Statutory Law

Jaromir Savelka
Kevin D. Ashley

Intelligent Systems Program
University of Pittsburgh

jas438@pitt.edu

ISP Seminar, University of Pittsburgh
April 08, 2015

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

Suppose there is a Thai restaurant at Shadyside and an Indian restaurant in Oakland having a single owner.

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

Suppose there is a Thai restaurant at Shadyside and an Indian restaurant in Oakland having a single owner.

Are these restaurants an “enterprise” within the meaning of the definition?

Motivation

29 U.S. Code 203 - Definitions

“Enterprise” means the **related activities** performed (either through **unified operation** or **common control**) by any person or persons for a **common business purpose**, and includes all such activities whether performed in one or more **establishments** or by one or more corporate or other **organizational units** including departments of an establishment operated through leasing arrangements, but shall not include the **related activities** performed for such enterprise by an **independent contractor**. [...]

Suppose there is a Thai restaurant at Shadyside and an Indian restaurant in Oakland having a single owner.

Are these restaurants an “enterprise” within the meaning of the definition?

Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.

Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



Motivation

Example Rule

No **vehicles** in the park.

Abstract rules in statutory provisions must account for diverse situations (even those not yet encountered).

⇒

Legislators use vague, open textured terms, abstract standards, principles, and values.



When there are doubts about the meaning of the provision they may be removed by interpretation.

Motivation

Interpretation involves an investigation of how the term has been referred to, explained, interpreted or applied in the past.

Example Uses of the Term

- i. Any mechanical device used for transportation of people or goods is a **vehicle**.
- ii. A golf cart is to be considered a **vehicle**.
- iii. To secure a tranquil environment in the park no **vehicles** are allowed.
- iv. The park where no **vehicles** are allowed was closed during the last month.
- v. The rule states: “No **vehicles** in the park.”

Going through the sentences is labor intensive because many sentences are useless and there is a large redundancy.

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

Task

Ultimately we would like to generate the set of the useful sentences automatically.

Task Definition

Given the term of interest t , the statutory provision it comes from (sp), and a corpus of the available documents DB generate a set of sentences S ($s_i \in DB$) of the size n that provides the most informative insight of how t is used.

Task

Ultimately we would like to generate the set of the useful sentences automatically.

Task Definition

Given the term of interest t , the statutory provision it comes from (sp), and a corpus of the available documents DB generate a set of sentences S ($s_i \in DB$) of the size n that provides the most informative insight of how t is used.

No **vehicles** in the park.

Task

Ultimately we would like to generate the set of the useful sentences automatically.

Task Definition

Given the term of interest t , the statutory provision it comes from (sp), and a corpus of the available documents DB generate a set of sentences S ($s_i \in DB$) of the size n that provides the most informative insight of how t is used.

No **vehicles** in the park.

Any mechanical device used for transportation of people or goods is a **vehicle**.

A golf cart is to be considered a **vehicle**.

To secure a tranquil environment in the park no **vehicles** are allowed.

Hypotheses

(H0: A sentence may be reliably evaluated in terms of its usefulness for an interpretation of the term from a specific statutory provision.)

Hypotheses

(H0: A sentence may be reliably evaluated in terms of its usefulness for an interpretation of the term from a specific statutory provision.)

H1: By using an appropriate list of linguistic features about/in the sentence it is possible to **automatically** evaluate how useful the sentence is for an interpretation of the term.

Hypotheses

(H0: A sentence may be reliably evaluated in terms of its usefulness for an interpretation of the term from a specific statutory provision.)

H1: By using an appropriate list of linguistic features about/in the sentence it is possible to **automatically** evaluate how useful the sentence is for an interpretation of the term.

(H2: By using the information about the interpretive usefulness of a sentence we can outperform existing systems, that do not use the information, in the task of retrieving the n best sentences for the interpretation of the term.)

Query-focused Summarization of Multiple Documents (as described in Gupta 2010)

- ▶ system based on supervised sentence ranking (Fisher and Roark 2006)
- ▶ handling a large pool of retrieved documents (Daumé and Marcu 2006)
- ▶ understanding the problem as QA (Schiffman and McKeown 2007)
- ▶ interactive component (Lin et al. 2010)

Related Applications in Different Domains

- ▶ automatic generation of Wikipedia articles (Sauper and Barzilay 2009)
- ▶ extractive summarization system for clinical QA (Demner-Fushman and Lin 2006)
- ▶ system for recommending relevant information to the users of Internet forums and blogs (Wang et al. 2010)
- ▶ mining of important product aspects from online consumer reviews (Yu et al. 2011)

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

Statutory Term Interpretation Data Set

Court decisions are an ideal source of sentences interpreting statutory terms.

For our corpus we selected three terms from different provisions of the United States Code:

1. “independent economic value” (18 U.S. Code § 1839(3)(B))
2. “identifying particular” (5 U.S. Code § 552a(a)(4))
3. “common business purpose” (29 U.S. Code § 203(r)(1))

For each term we have collected a small set of sentences by extracting all the sentences mentioning the term from the top 20 court decisions retrieved from Court Listener.¹

In total we assembled a small corpus of 243 sentences.

1. <https://www.courtlistener.com/>

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

1. **high value** – sentence intended to define or elaborate on the meaning of the term

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

1. **high value** – sentence intended to define or elaborate on the meaning of the term
2. **certain value** – sentence that provides grounds to elaborate on the term's meaning

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

1. **high value** – sentence intended to define or elaborate on the meaning of the term
2. **certain value** – sentence that provides grounds to elaborate on the term's meaning
3. **potential value** – sentence that provides additional information beyond what is known from the provision the term comes from

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

1. **high value** – sentence intended to define or elaborate on the meaning of the term
2. **certain value** – sentence that provides grounds to elaborate on the term's meaning
3. **potential value** – sentence that provides additional information beyond what is known from the provision the term comes from
4. **no value** – no additional information over what is known from the provision

Statutory Term Interpretation Data Set

Two expert annotators classified the sentences into four categories according to their usefulness for the interpretation:

1. **high value** – sentence intended to define or elaborate on the meaning of the term
2. **certain value** – sentence that provides grounds to elaborate on the term's meaning
3. **potential value** – sentence that provides additional information beyond what is known from the provision the term comes from
4. **no value** – no additional information over what is known from the provision

inter-annotator agreement: .746

weighted kappa: .66

Statutory Term Interpretation Data Set

Summary statistics about the annotated corpus:

Term	# HV	# CV	# PV	# NV	# Total
Ind. economic val.	2	5	40	5	52
Identifying part.	6	8	40	17	71
C. business purp.	20	26	51	23	120
Total	28	39	131	45	243

HV high value

CV certain value

PV potential value

NV no value

	high	certain	potential	no
high	19	1	1	0
certain	15	12	9	1
potential	2	27	105	11
no	0	0	4	36

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

The processing performed by the framework for each interpretation query can be divided into five rather self-contained stages:

1. sentence retrieval
2. sentence classification/annotation
3. sentence interpretive value analysis
4. sentence clustering
5. sentence cluster selection/ranking

The input to the process is the 3-tuple $\langle t, sp, DB \rangle$ where t is the term of interest, sp the provision the term comes from, and DB the document base.

Sentence Retrieval

At this stage we retrieve all the documents matching the query and from them extract the sentences mentioning the term of interest.

We used simple key-word matching for both, the document retrieval and the sentence extraction.

In future we would like to:

- ▶ propose a mechanism for dealing with multi-word terms in a more sophisticated way
- ▶ take synonymity into account
- ▶ use anaphora resolution
- ▶ use more meaningful segmentation

Sentence Classification

At this stage we would like to assign the sentences with labels/annotations that could later help in assessing the usefulness of the sentences for the interpretation of the term of interest.

We assign the labels in the following eight categories:

1. source
2. similarity
3. syntactic importance
4. assignment or contrast
5. feature assignment
6. structural placement
7. rhetorical role
8. attribution

Sentence Classification: Source

In this category a sentence can be assigned one of the following labels:

- ▶ same provision
- ▶ same section
- ▶ different section
- ▶ different jurisdiction
- ▶ unknown

Example

The full text of §1839(3)(B) is: “[...]” . [...] Every firm other than the original equipment manufacturer and RAPCO had to pay dearly to devise, test, and win approval of similar parts; the details unknown to the rivals, and not discoverable with tape measures, had considerable “independent economic value ... from not being generally known” .

Sentence Classification: Similarity

In this category a sentence can be assigned one of the following labels:

- ▶ same
- ▶ similar
- ▶ related
- ▶ different

18 U.S. Code §1839

[...] the information derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable through proper means by, the **public**;

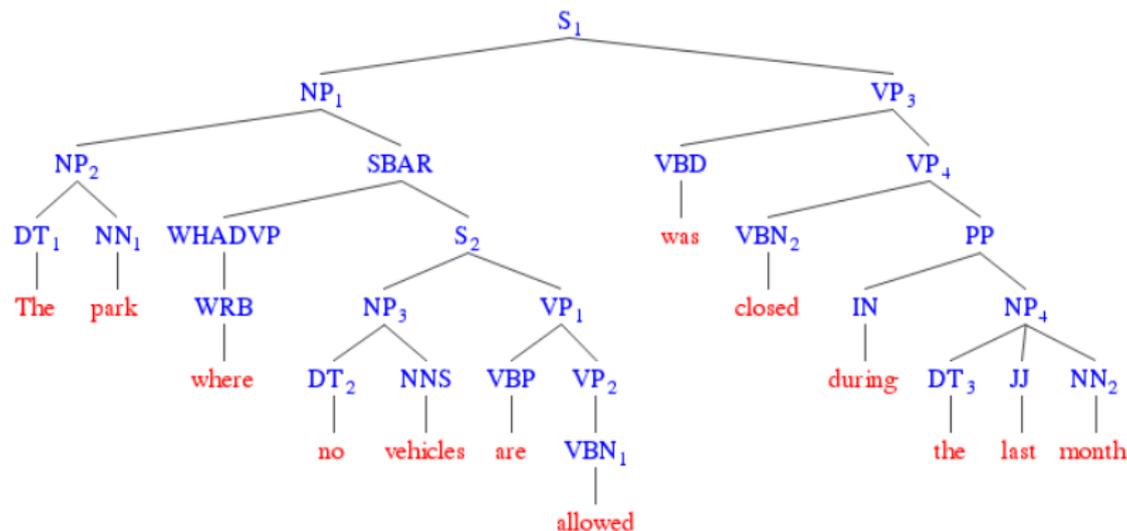
17 U.S. Code §116

[...] posted in the establishment in a prominent position where it can be readily examined by the **public**;

Sentence Classification: Syntactic Importance

In this category a sentence can be assigned one of the following labels:

- ▶ dominant
- ▶ important
- ▶ not important



Sentence Classification: Assignment or Contrast

In this category a sentence can be assigned one of the following labels:

- ▶ another term is a specific case of the term of interest
- ▶ the term of interest is a specific case of another term
- ▶ the term of interest is the same as another term
- ▶ the term of interest is not another term
- ▶ no assignment

Example: Another term a specific case

The Fifth Circuit has held that the **profit motive** is a **common business purpose** if shared.

Sentence Classification: Feature Assignment

In this category a sentence can be assigned one of the following labels:

- ▶ the term of interest is a feature of another term
- ▶ another term is a feature of the term of interest
- ▶ no feature assignment

Example: Another term is a feature

However, Reiser concedes in its brief that the **process** has **independent economic value**.

Sentence Classification: Structural Placement

In this category a sentence can be assigned one of the following labels:

- ▶ standard sentence
- ▶ citation
- ▶ quoted expression
- ▶ heading
- ▶ footnote

Example: Heading

A. Related Activities and **Common Business Purpose**.

Example: Footnote

[5] [...] However, in view of the '**common business purpose**' requirement of the Act, we think [...]

Sentence Classification: Rhetorical Role

In this category a sentence can be assigned one of the following labels:

- ▶ application of law to factual context
- ▶ applicability assessment
- ▶ statement of fact
- ▶ statement of law
- ▶ interpretation of law
- ▶ general explanation or elaboration
- ▶ reasoning statement
- ▶ holding
- ▶ other

Sentence Classification: Attribution

In this category a sentence can be assigned one of the following labels:

- ▶ judge
- ▶ legislator
- ▶ party to the dispute
- ▶ witness
- ▶ expert
- ▶ other

Example: Party to the dispute

In support of his contention that Gold Star Chili and Caruso's Ristorante constitute an enterprise, **plaintiff alleges** that Caruso's Ristorante and Gold Star Chili were engaged in the related business activity [...].

Sentence Interpretive Value Analysis

At this stage we would like to assign each sentence with one of the four labels:

1. high value
2. certain value
3. potential value
4. no value

For each sentence we would like to generate an interpretive value score based on the annotations and the content of the sentence.

Eventually we would like the score to be a value from a continuous interval. However, for the purpose of the evaluation we use the discretized version.

Sentence Clustering

At this stage we would like to cluster together sentences that:

- ▶ are exact duplicates of each other
- ▶ differ only in negligible aspects
- ▶ are semantically very close

... because the system should avoid presenting a user with several s_j that are too similar.

In addition, it is important to distinguish between an isolated statement expressed in a single s_j and an established doctrine repeated many times.

For each group G_i a representative s_j is picked.

We pass the clusters to the next processing stage.

Sentence Cluster Selection/Ranking

At this stage we would like to select the clusters that will be presented to the user and decide about their order.

The selection of n clusters $G_i \in G$ is an optimization problem where we wish to select a subset G_S of G of the size n such that the following criteria are maximized:

- ▶ **sum of interpretive value scores** of the representative sentences of each $G_i \in G_S$
- ▶ **sum of the sizes** of all $G_i \in G_S$
- ▶ **joint informativeness** of the representative sentences of each $G_i \in G_S$
- ▶ **sum of the relevance scores** of the documents the sentences in $G_i \in G_S$ come from

After this processing stage the system outputs G_S . A user may be presented with a list of representative sentences of each $G_i \in G_S$.

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

Experiment

We conducted an experiment to confirm H1, i.e., we investigated if the interpretive value of a sentence can be predicted automatically.

The goal is to assign each sentence with one of the four labels:

1. high value
2. certain value
3. potential value
4. no value

As features we used the eight linguistic categories:

1. source
2. similarity
3. syntactic importance
4. assignment or contrast
5. feature assignment
6. structural placement
7. rhetorical role
8. attribution

Experiment

We randomly divided the sentences into the training set (2/3) and the test set (1/3).

As classification models we used:

1. Most frequent class (baseline)
2. Naïve Bayes
3. SVM
4. Random Forest

Because the dataset is small we repeated the experiment 100 times.

In each run we evaluated the performance on the test set as well as performed a 10-fold cross validation on the training set.

Results

Mean results from 100 runs of a classification experiment:

Classifier	CV	STD	TEST	STD	SIG
Most frequent	.545	.025	.531	.049	–
Naïve Bayes	.544	.037	.611	.066	no
SVM	.633	.044	.657	.066	no
Random Forest	.677	.033	.696	.042	yes

CV 10-fold cross validation on the training set
STD standard deviation
TEST validation on the test set
SIG statistical significance

	high	certain	potential	no
high	2	0	0	0
certain	2	7	14	1
potential	1	5	33	2
no	0	0	1	13

Results: Features

Mean results of classification experiment where each line reports the performance when the respective feature was removed:

Features	CV	STD	TEST	STD
all	.677	.033	.696	.042
-source	.519	.05	.586	.046
-semantic relationship	.675	.031	.694	.049
-syntactic importance	.532	.028	.521	.047
-structural placement	.695	.033	.708	.047
-rhetorical role	.687	.033	.695	.049
-attribution	.657	.034	.671	.048
-assignment/contrast	.668	.032	.669	.045
-feature assignment	.662	.032	.684	.047

Presentation Overview

Motivation

Task

Statutory Interpretation Data Set

Framework

- Sentence Retrieval

- Sentence Classification

- Interpretive Value Analysis

- Sentence Clustering

- Sentence Cluster Selection

Predicting Usefulness

Conclusion

Future Work

The ultimate aim is to develop and test a fully functional and well described framework supporting interpretation of statutory terms.

We would like to further develop the component for predicting interpretive value of a sentence.

We would also like to focus on the other constituents of the processing pipeline.

As the next step we would like to add new data to the corpus to have 1,000–2,000 sentences.

Then, we would like to test H2 (outperform the existing systems in the retrieval of the n best sentences).

Conclusion

We have shown that:

- ▶ a sentence may be reliably evaluated in terms of its usefulness for an interpretation of a selected statutory term.
(0.746 inter-annotator agreement, 0.66 weighted kappa)
- ▶ by using the selected linguistic features it is possible to automatically evaluate how useful a sentence is for an interpretation of a selected statutory term.
(0.696 agreement with gold s.)

... confirming H0 and H1.

Therefore, we have suggested a feasibility of the framework for the computational support for interpretation of statutory terms sketched in this talk.

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a common business purpose, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a **common business purpose**, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

29 U.S. Code 203 - Definitions

“Enterprise” means the related activities performed (either through unified operation or common control) by any person or persons for a **common business purpose**, and includes all such activities whether performed in one or more establishments or by one or more corporate or other organizational units including departments of an establishment operated through leasing arrangements, but shall not include the related activities performed for such enterprise by an independent contractor. [...]

List of Interpretive Sentences

The “**common business purpose**” requirement is not defined in the Act.

The utilization of a common service does not by itself establish a **common business purpose** shared by the owners of separate businesses.

Activities are performed for a **common business purpose** if they are “directed toward the same business objective or to similar objectives in which the group has an interest.”

In a situation such as this, in which the Court has concluded that there are no related activities, the fact of common ownership of the two businesses clearly is not sufficient to establish a **common business purpose**.

The Fifth Circuit has held that the profit motive is a **common business purpose** if shared.

Thank you!

Questions, comments and suggestions are welcome now
or any time at jas438@pitt.edu.