

A Corpus Study of Levin's Verb Classification

Jianguo Li, Kirk Baker, Chris Brew

Department of Linguistics
The Ohio State University

ACL
March 15th, 2008

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Verb Classification

Lexical semantic classes of verbs play an important role in structuring complex predicate information in a lexicon, thereby avoiding redundancy in verb descriptions and enabling generalizations across semantically similar verbs with respect to their usage.

- ▶ Many scholars hypothesize that the behavior of a verb, particularly with respect to the expression of arguments and the assignment of semantic roles is to a large extent driven by deep semantic regularities (Dowty, 1991; Green, 1974; Goldberg, 1995; Levin 1993).
- ▶ There exist several manually-created verb classifications: Levin, VerbNet, and FrameNet, which have been widely used in various natural language processing tasks.

Introduction

Verb Classification

Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data

Machine Learning Method

Results

Discussion

Conclusion

Analysis

Acknowledgments

Levin's Verb Classification

- ▶ Levin's verb classification rests on the assumption that verbs that exhibit the same diathesis alternation share certain semantic components and form a semantically coherent class.
 - ▶ Dative alternation:
 - ▶ *Mary sent John a book.* NP-V-NP-NP
 - ▶ *Mary sent a book to John.* NP-V-NP-PP(to)
 - ▶ Benefactive alternation:
 - ▶ *Mary bought John a book.* NP-V-NP-NP
 - ▶ *Mary bought a book for John.* NP-V-NP-PP(for)
- ▶ Levin classifies 3,014 English verbs into 197 verb classes.

Introduction

Verb Classification

Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data

Machine Learning Method

Results

Discussion

Conclusion

Analysis

Acknowledgments

Goals

What is the relation between Levin's approach, her theoretical motivations, and the corpus data?

- ▶ Levin classifies verbs based on their syntactic behavior. *What are the most informative features in automatic verb classification?*
- ▶ Levin pre-selects syntactic frames that display alternations. *Are these frames important for classification?*
- ▶ Levin classifies verbs based on whether a given verb participates in a particular alternation, without regard to frequency. *Does the frequency information, derived from corpus data, benefit or hurt automatic verb classification?*

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces

Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Subcategorization Frames

- ▶ Subcategorization frames are obviously relevant to alternation behaviors, and have been adopted in earlier work on automatic verb classification (Schulte im Walde, 2000; Brew and Schulte im Walde, 2002; Korhonen et al., 2003).
- ▶ Subcategorization frames generally only contain syntactic information, lexical information is largely disregarded.
- ▶ *Mary sent a book to John.* {NP-V-PPto}

Lexical Co-occurrences

- ▶ Levin does not address whether verbs in the same class tend to share their neighboring words.
- ▶ Unlike subcategorization frames, co-occurrences mostly convey lexical information only, and are generally considered not particularly sensitive to argument structures.
- ▶ Conventional co-occurrence features also adopt a stop list to filter out function words (e.g. prepositions, determiners).
- ▶ *Mary sent a book to John.* {*Mary, book, John*}

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Subcategorization Frames + Co-occurrences

- ▶ We mix syntactic and lexical information together by combining subcategorization frames and co-occurrence features in hope that they are complementary to each other, and therefore yield better performance for automatic verb classification.
- ▶ *Mary sent a book to John.* {NP-V-PPto, *Mary, sent, book, John*}

Dependency Relations

- ▶ A dependency relation is represented as a triple $\langle V, D, R \rangle$ where V represents the anchoring verb, D its dependent, and R the syntactic relations between V and D .
- ▶ Dependency relations encode both syntactic and lexical information, and have been shown to be informative and reliable in a wide range of lexical acquisition tasks, such as detection of polysemy, and word sense disambiguation.
- ▶ *Mary sent a book to John.* $\{ \langle \text{send}, \text{Mary}, \text{subj} \rangle, \langle \text{send}, \text{book}, \text{obj} \rangle, \langle \text{send}, \text{John}, \text{PPto} \rangle \}$

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Adapted Co-occurrences

- ▶ Conventional co-occurrence features generally ignore function words. However, some of the function words, prepositions in particular, are known to carry great amount of syntactic information that is related to the lexical meanings of verbs.
- ▶ We adapt the conventional co-occurrence features by keeping all prepositions.
- ▶ Adapted co-occurrence features contain at least some degree of syntactic information.
- ▶ *Mary sent a book to John.* {*Mary, book, to, John*}

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces

Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Corpus

- ▶ To collect each type of feature, we use the English Gigaword Corpus, which consists of samples of recent newswire text data collected from four distinct international sources of English newswire.

Introduction

Verb Classification
GoalsAutomatic Verb
Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Feature Extraction: Subcategorization Frames and Dependency Relations

These more linguistically informed features are constructed based on the grammatical relations generated by the CCG parser (Clark and Curran, 2007).

<i>Mary sent a book to John</i>
(det book_3 a_2)
(doj sent_1 book_3)
(doj to_4 John_5)
(ioj sent_1 to_4)
(ncsubj sent_1 Mary_0 _)

- ▶ Subcategorization: {NP-V-NP-PP_{to}}
- ▶ Dependency Relations: {⟨send, Mary, subj⟩, ⟨send, book, obj⟩, ⟨send, John, PP_{to}⟩}

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Feature Extraction: Co-occurrences and Adapted Co-occurrences

- ▶ Co-occurrence features are collected using a flat 4-word lemmatized window, meaning that 4 words to the left/right of each target verb are considered potential features. However, we eliminate all co-occurrence features that are in a stopwords list, which consists of about 200 closed class words including mainly prepositions, determiners, complementizers and punctuation (e.g. Rohde-et-al, 2004).
- ▶ We adapt conventional co-occurrence features by
 - ▶ keeping all prepositions.
 - ▶ replacing all the verbs in the neighboring contexts of the target verb with their POS tags (to reduce data sparsity).

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces

Experimental Setup

Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Verb Classes

Early work has focused only on a small set of verbs or a small number of verb classes (30 classes and 153 verbs in Schulte im Walde (2000), 3 classes and 60 verbs in Merlo and Stevenson (2001), and 15 classes and 835 verbs in Joanis-et-al. (2007)).

- ▶ We test the applicability of different feature sets to distinctions among 48 classes involving 1,300 verbs. To our knowledge, this is, by far, the largest investigation of English verb classification.
- ▶ Levin48 is obtained by selecting the monosemous members of the Levin verb classes that have at least 10 monosemous members with a frequency of at least 100 in the Gigaword Corpus.

Building the Frequency Matrix

We represent the semantic space for verbs as a matrix of counts, where each row corresponds to a target Levin verb v , and each column represents a given feature f . Each cell represents the frequency $w_{v,f}$ of a verb v occurring with a feature f .

	NP-V	NP-V-NP	NP-V-NP-PP _{to}	NP-V-NP-NP
add	10	100	15	2
give	5	1	80	150
throw	2	50	130	30
write	5	30	50	100

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup

Preprocessing Corpus Data

Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Introduction

Verb Classification
GoalsAutomatic Verb
ClassificationFeature Spaces
Experimental Setup

Preprocessing Corpus Data

Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Normalization Methods

To reduce undue influence of outlier features, we employ a normalization technique - *correlation* (Rohde-et-al, 2004), which is supposed to help reduce the range of extreme values while having little effect on others. Each raw frequency $w_{v,f}$ will be replaced with its normalized value $w'_{v,f}$.

$$\text{correlation} : \frac{T w_{v,f} - \sum_j w_{v,j} \sum_i w_{i,f}}{(\sum_j w_{v,j} (T - \sum_j w_{v,j}) \sum_i w_{i,f} (T - \sum_i w_{i,f}))^{1/2}}$$
$$T = \sum_i \sum_j w_{i,j} \quad (1)$$

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data

Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Classifier

Bayesian Multinomial Logistic Regression (BMR)

- ▶ BMR performs the so-called 1-of-k classification (Madigan-et-al, 2005).
- ▶ It has been shown to be very efficient with handling large number of features and extremely sparsely populated matrices, which characterize the data we have for automatic verb classification.
- ▶ We also tried Support Vector Machines (SVMs). However, BMR generally outperforms SVMs for our task.

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Classification Tasks

If we could exhaust all the possible n -way ($2 \leq n \leq 48$) classification tasks with the 48 Levin classes that we will investigate, it would allow us to draw a firmer conclusion about the general applicability and scalability of a particular feature set. However, the number of classification tasks grows really huge when n takes on certain value (e.g. $n = 20$).

- ▶ When $n = 2$, we perform all the possible 1,028 classification tasks.
- ▶ When $n = 48$, we perform the only one possible classification task.
- ▶ When $n = 5, 10, 20, 30, 40$, we randomly select 100 tasks for each n .

Classification Tasks

- ▶ For each n -way task, we randomly select 9 verbs (the smallest classes have 10 verbs) from each class as training data, and repeat this process 10 times.
- ▶ The accuracy for each n -way task is computed by averaging the results from these 10 trials.
- ▶ The accuracy reported for the overall n -way classification for each selected n is obtained by averaging the results from each individual n -way task for that particular n .
- ▶ For each trial, each feature set is trained and tested on the same training/test split.

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Classification Results

SCF: Subcategorizations

CO: Co-occurrences

DR: Dependency Relations

ACO: Adapted Co-occurrences

n-way	No of tasks	Baseline	SCF	CO	SCF+CO	DR	ACO
2-way	1,028	50	84.0	77.8	82.9	83.1	80.9
5-way	100	20	71.9	69.8	71.1	73.1	70.1
10-way	100	10	65.8	70.4	77.3	76.4	73.0
20-way	100	5	51.4	58.8	65.8	62.1	60.1
30-way	100	3.3	46.7	48.6	57.8	55.7	51.8
40-way	100	2.5	43.6	47.3	55.1	53.9	49.9
48-way	1	2.2	39.1	42.4	52.8	50.6	46.8

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Subcategorization - Most Informative?

Both syntactic and lexical information are useful in deriving Levin-style verb classification.

- ▶ Subcategorization frames are not the most effective features in deriving Levin-style verb classification. Although they achieve the highest accuracy on the 2-way classification, their accuracy drops more drastically as n gets bigger.
- ▶ Co-occurrence features outperform subcategorization frames on every n -way classification when $n \geq 10$.
- ▶ The feature sets that include both syntactic and lexical information generally work better than either of them used alone.

Introduction

Verb Classification
GoalsAutomatic Verb
ClassificationFeature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Levin-selected Frames - More Useful?

One explanation for the poor performance achieved by using subcategorization frames could be that we use all the frame types generated by the CCG parser in our experiment (**CCG-SCF**). A potentially better way of doing this would be to use Levin-selected frame set.

- ▶ **Unfiltered-Levin-SCF**: We match each subcategorization frame generated by the CCG parser to one of the 78 Levin-defined frames.
- ▶ **Filtered-Levin-SCF**: We employ a statistical test (Binomial Hypothesis Test) to filter out noisy subcategorization frames.

Task	CCG-SCF	Unfiltered-Levin-SCF	Filtered-Levin-SCF
48-way	39.1	39.7	40.8

Introduction

Verb Classification
GoalsAutomatic Verb
ClassificationFeature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method

Results

Discussion

Conclusion
Analysis

Acknowledgments

Frequency Information - Any Benefits?

Levin classifies verbs based on whether a given verb participates in a particular alternation, without regard to frequency. So far, we have used frequency information in our experiments. What if we ignore the frequency information and use binary features for the classification tasks?

Task	CCG-SCF	Unfiltered-Levin-SCF	Filtered-Levin-SCF
48-way (frequency)	39.1	39.7	40.8
48-way (binary)	33.9	33.6	34.2

Conclusion

Our experiment on automatic verb classification, guided by Levin's lexical semantic theory, failed to yield what we would have predicted by extrapolation of Levin's theory.

- ▶ Subcategorization frames are not the most informative features in deriving Levin-style verb classification. Lexical information is useful too.
- ▶ Levin-selected frames do not work much better than parser-defined frames.
- ▶ Frequency information proves to be a benefit even though it is not used in Levin's classification.

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Introduction

Verb Classification
Goals

Automatic Verb
Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion

Analysis

Acknowledgments

Analysis

- ▶ Is this due to the deficiencies of the techniques, or the theories?
- ▶ What insights can our experiments provide into linguistic theories?

Techniques

- ▶ The CCG parser we used makes errors in extracting subcategorization frames, even though it is one of the state-of-the-art parsers. The F-measure value for extracting subcategorization frames from the Wall Street Journal is about 80%. In contrast, errors arising from generating co-occurrence features are almost negligible.
- ▶ The noise introduced by the parser is likely to have a negative impact on the performance of the classifier when subcategorization frames are chosen as features.

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion

Analysis

Acknowledgments

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion

Analysis

Acknowledgments

Theories

- ▶ Levin's classification is based on the examples constructed by linguists.
- ▶ Information mined from corpus data can be used to test linguistic theories by taking into account realistic examples that do not constitute idealizations of linguistic phenomena.

Dative Alternation

- ▶ Dative Alternation: Lexical semantic theories (Dowty 1991, Levin 1993) have predicted that the class of PREVENTION-OF-POSSESSION (e.g. *cost*, *deny*) are not allowed to occur in the dative frame (NP-V-NP-PP_{to}).
 - ▶ *The car costs Beth \$5000. *The car costs \$5000 to Beth.*
 - ▶ *Ann denies Beth the ice cream. *Ann denies the ice cream to Beth.*
- ▶ Corpus data have shown that these verbs occur in both dative and double-object frame (NP-V-NP-NP). In fact, the frequency ratio between this class of verbs occurring in dative and double-object frame is about 1:9.
 - ▶ *It will cost jobs to the Californians in the televised profession.*
 - ▶ *The employer has denied a leave to the teacher.*

Introduction

Verb Classification
Goals

Automatic Verb Classification

Feature Spaces
Experimental Setup
Preprocessing Corpus Data
Machine Learning Method
Results

Discussion

Conclusion
Analysis

Acknowledgments

Acknowledgments

- ▶ NSF grant 0347799
- ▶ OSU Clippers (computational linguistic discussion group)