The Impact of Selectional Preference Agreement on Semantic Relational Similarity

Bryan Rink   Sanda Harabagiu
University of Texas at Dalals
{bryan,sanda}@hlt.utdallas.edu

Abstract

Relational similarity is essential to analogical reasoning. Automatically determining the degree to which a pair of words belongs to a semantic relation (relational similarity) is greatly improved by considering the selectional preferences of the relation. To determine selectional preferences, we induced semantic classes through a Latent Dirichlet Allocation (LDA) method that operates on dependency parse contexts of single words. When assigning relational similarities to pairs of words, if the agreement of selectional preferences is considered alone, a correlation of 0.334 is obtained against the manual ranking outperforming the previously best reported score of 0.229.

1 Introduction

In natural language, words participate often in a variety of semantic relations. Both linguists and psychology researchers have been interested in categorizing semantic relations and to understand their usage in language and cognition. One particular interesting usage of semantic relations is provided by analogical reasoning. As reported by Gentner (1983) and Holyoak and Thagard (1996), whenever a new situation arises, humans tend to search for an analogous situation from their past experience. Analogical reasoning relies on relational similarity, as reported by Turney (2006) and Turney (2008). In analogical reasoning, the degree of relational similarity is an estimation of the likelihood of applicability of the knowledge transfer (from past to present). Thus, as postulated in the recent SemEval 2012 Task 2 (Jurgens et al., 2012), the automatic analysis of relational similarity may have practical benefits of indicating the appropriateness of an analogy.

Relational similarity, as reported in Turney (2006), is one of the forms of similarity, the other one being provided by attributional similarity. Relational similarity evaluates the correspondence between relations (Medin et al., 1990), while attributional similarity evaluates the correspondence between attributes. As stated by Turney: "*When two words have a high degree of attributional similarity, we call them* synonyms. *When two word* pairs *have a high degree of relational similarity, we say they are* analogous."

We claim that there is a special property that arguments of relations need to share. The arguments of relations are words which are predications of binary facts, properties, actions, etc. As such, we are aware from the work of Resnik (1996) that words which appear as arguments of a predicate define the selectional preferences of the predicate. Moreover, Pantel et al. (2007) have extended the notion of predicate selectional preferences to "relational selectional preferences" of binary relations. For a binary relation $r(x, y)$, the semantic classes $C(x)$ which can be instantiated for the argument x as well as $C(y)$, the semantic classes which can be instantiated for the argument y constitute the relational selectional preferences of the binary relation. Thus we believe and show in this paper that semantic relations have selectional preferences and that word pairs x:y are more similar to a relation when those words are more admissible under the relational selectional preferences.

Consider the semantic relation REFERENCE-*Expression*, with prototypical word pairs *smile*:*friendliness*, *lamentation*:*grief*, and *hug*:*affection*. In these pairs, the first word can be seen as a physical expression of the emotional state represented by the second word. Word pairs which are prototypical of the relation

Category	Example word pairs	Relations
CLASS-INCLUSION	flower:tulip, weapon:knife, clothing:shirt, queen:Elizabeth	5
PART-WHOLE	car:engine, fleet:ship, mile:yard, kickoff:football	10
SIMILAR	car:auto, stream:river, eating:gluttony, colt:horse	8
CONTRAST	alive:dead, old:young, east:west, happy:morbid	8
ATTRIBUTE	beggar:poor, malleable:molded, soldier:fight, exercise:vigorous	8
NON-ATTRIBUTE	sound:inaudible, exemplary:criticized, war:tranquility, dull:cunning	8
CASE RELATIONS	tailor:suit, farmer:tractor, teach:student, king:crown	8
CAUSE-PURPOSE	joke:laughter, fatigue:sleep, gasoline:car, assassin:death	8
SPACE-TIME	bookshelf:books, coast:ocean, infancy:cradle, rivet:girder	9
REFERENCE	smile:friendliness, person:portrait, recipe:cake, astronomy:stars	6

Table 1: The ten categories of semantic relations. Each word pair has been taken from a different subcategory of each major category.

should be assigned a high degree of membership for the REFERENCE-*Expression* relation, while word pairs such as *discourse*:*relationship* and *anger*:*slap* should not, either because the word pair expresses a different relation, or because the pair is in the wrong order (slap is an expression of anger, not the other way around). Table 1 shows the ten top-level categories of relations we consider, which is further divided into 79 relations covering multiple parts of speech (adjective, noun, adverb, and verb).

We show that a model which independently considers the semantic classes of each word in a word pair is effective at assigning degrees of membership (relational similarity). For instance, knowing that the relation REFERENCE-*Expression* selects for emotional states in the first argument (e.g., *grief*, *friendliness*, *affection*) and expressions of emotion in the second argument (e.g., *smile*, *hug*, *lamentation*) helps in determining word pair candidates which don't adhere to those classes. Clearly word pairs whose arguments do not fit these preferences should be given a lower degree of relatedness to the relation. We describe a method for inducing semantic classes for use as selectional preferences and a method for determining the distributions over argument classes for a relation. While selectional preferences are not the only phenomena responsible for assigning degrees of membership for word pairs to semantic relations, we choose to model it alone in this paper to examine its importance. We show that modeling selectional preference alone produces results which are better than the previously reported results for measuring relational similarity.

The rest of this paper is structured as follows: Section 2 gives some perspective on previous work, Section 3 describes how we used an LDA model to induce semantic classes. Section 4 describes the dataset we use for measuring relational similarity. Section 5 describes how the induced semantic classes are used to model the selectional preferences of semantic relations. Section 5 describes how we determine the extent to which a word pair matches a relation's selectional preferences. Section 6 gives our experimental setup and the results of our evaluation. Section 7 analyzes the types of semantic classes that were automatically induced and Section 8 concludes the paper.

2 Previous work

Prior work on relational similarity (Jurgens et al., 2012; Rink and Harabagiu, 2012; Turney, 2005, 2006) has understandably focused the actual relation between a pair of words under consideration. These approaches have all considered how the two words co-occur in a large corpus and what contexts can be found near the words when they co-occur. Contextual information is useful for determining the relationship between two words. Therefore we believe the selectional preference agreement method can complement these approaches. The best-performing relational similarity approach at the SemEval 2012 Task 2 utilized a graphical model to determine patterns likely to be found between the two words of a word pair within a large corpus (Rink and Harabagiu, 2012). Word pairs were then ranked by their likelihood of occurring with those patterns. Constraints on the arguments were not directly addressed. One of the limitations of the approach is that word pairs which never occurred near each other in the corpus could not be ranked, which occurred regularly for some relations. The approach we present does not have this sparsity issue because we treat the relation arguments independently.

The literature on selectional preferences has focused largely on well-known relations such as syntactic relations (Mechura, 2008; Resnik, 1996; Ó Séaghdha, 2010), considering typical subjects and direct objects of verbs, or typical nouns modified by specific adjectives. These approaches usually focus on semantic classes of nouns at the exclusion of other parts of speech. One recent example relevant to our work is a set of LDA-inspired models proposed by Ó Séaghdha (2010). His models directly induce semantic classes for each predicate (verb or adjective). One consequence of such approaches is that the semantic classes differ based on the type of relationship being modeled: verb-object, noun-noun, or adjective-noun. The set of classes derived for nouns which are objects of verbs will be different than the classes derived for nouns which are modified by adjectives for instance. In our approach we induce semantic classes independently of the relations whose selectional preferences we are modeling. We take this approach because our relational data consists only of word pairs with no context. Further, some of the word pairs may never occur in the same sentence even in a large corpus (e.g., *signature:acknowledgment*) yet we can still check the admissibility of the words as arguments to the desired relation (e.g, X represents Y).

An extension to Latent Dirichlet Allocation model has been used before by Ritter and Etzioni (2010) to model semantic relations and their selectional preferences. There are two distinct reasons their approach is not well-suited to the relational similarity task. First, they were additionally inducing the set of relations present in their data, while in the relational similarity task we aim to determine membership to an existing set of relations. The second difference in their approach is the large size of their dataset. While we were able to train our models using on average around 40 word pairs per relation, their data contained all tuples matching a relation over a large corpus.

There has been much previous research effort on inducing semantic classes as well. Most approaches use some form of context around words to induce the classes. Older approaches simply used a bag of words context (Roark and Charniak, 1998), but this leads to induced classes containing more paradigmatically similar words rather than syntagmatically similar words (Widdows and Dorow, 2002). More recent approaches have utilized a subset of semantically-rich syntactic relations such as verb-object, noun modifier, coordination, and preposition (Baroni and Lenci, 2010; Widdows and Dorow, 2002). Lin and Pantel (2001) induce semantic classes using dependency parse contexts. Their approach is based on a vector space rather than the probabilistic setting of an LDA. Rahman and Ng (2010) use a factor graph with various semantic, morphological, and grammatical features to induce a set of semantic classes with the goal of performing better named entity recognition. Pantel (2003) uses short contextual patterns to inform a clustering approach to category induction.

3 Inducing semantic word classes

We consider a *semantic class* to be a set of words which share a semantic property. For example, the semantic property "male" forms a semantic class which includes the words "man, bull, boy, boyfriend, groom". Under this definition, words can belong to many semantic classes. For example "man" could belong to semantic classes for "man", "adult", and "human". We adopt the "distributional hypothesis" that the meaning of words can be inferred from their context. We follow existing approaches which use syntactic dependency context (Lin and Pantel, 2001) for inducing semantic classes. n The basis of our model for selectional preference agreement uses a set of semantic word classes induced using a Latent Dirichlet Allocation model (Blei et al., 2003). The data for this model is structured differently than a standard LDA, so that rather than inducing topic distributions for documents, we induce semantic class distributions for words. We begin with a large corpus of documents and dependency parses (De Marneffe and Manning, 2008) for all the documents. Every time a word occurs in the corpus we collect all of the dependency edges which include the word. We then concatenate the label on the dependency edge and the other word to form what we call a *dependency context*. For instance, the syntactic dependency *sadness* $\xleftarrow{dobj}$ *expressed* would generate one dependency context for *sadness*: "*←dobj expressed*" and one dependency context for *expressed*: "→*dobj sadness*". Figure 1 shows the most frequent dependency contexts for the word *sadness*.

We train our LDA model by forming a pseudo-document for each unique word in the corpus consisting of all of the dependency contexts for that word, with repetitions. Figure 1 shows a small part of the pseudo-document formed for the word *sadness*. After forming such pseudo-documents, the LDA can be trained in the usual way to infer the parameters of the model.

More formally, the generative story for this LDA can be written as:

1. For each semantic class k, draw a distribution over dependency contexts $\phi_k \sim Dirichlet(\beta)$
2. For each unique word in the corpus w, draw a distribution over semantic classes $\theta_w \sim Dirichlet(\alpha)$
3. For each dependency context k of word w in the corpus, draw a semantic class $z_{w,k} \sim Multinomial(\theta_w)$
4. Draw a dependency context $d_{w,k} \sim Multinomial(z_{w,k})$

The LDA model trained on the pseudo-documents formed from dependency contexts will form two clusterings, a clustering of dependency contexts and a clustering of words. We argue that the clustering of words represent semantic classes. We evaluate this claim in Section 7. As an example of dependency context clustering, a *person* semantic class could be induced which would often be assigned to dependency contexts such as "←nsubj said", "→amod young", "→amod famous", or "→amod teenage".

The trained LDA model also assigns each pseudo-document a distribution over semantic classes (θ_w). Because the pseudo-documents correspond to unique words from the corpus, we can assess the affinity of each word to each semantic class and, in turn, compare two words to each other. For example, the most frequent dependency contexts for *sadness* include ←DOBJ *expressed*, →AMOD *great*, and →AMOD *deep*. Some other words sharing these contexts include *hope*, *sorrow*, *regret*, and *satisfaction*. We would therefore expect them to be clustered together by the LDA model. This allows us to compare two words based on the similarity of their semantic class distributions (θ_w).

Freq.	Context
1485	→det the
1233	←dobj expressed
978	→amod great
857	→det a
757	→punct "
601	→amod deep
532	→poss his
388	←prep_of sense
386	←prep_with is
342	→amod profound
318	←conj_and shock
300	←dobj express
297	←nsubj is
279	→poss their
259	→det the
248	←dobj feel
212	←dobj expressing
193	←dobj felt
189	←prep_with tinged
184	→conj_and anger
184	←conj_and anger
180	→det The
176	→det some
170	←nsubj 's
163	←prep_of lot

Figure 1: A compact representation of the pseudo-document associated with the word *sadness*. The most frequent contexts are shown.

4 SemEval-2012 Relational Similarity Task

The dataset we use for evaluating the degrees of relational similarity was developed as part of SemEval 2012 Task 2 - Measuring Degrees of Relational Similarity (Jurgens et al., 2012). In the task, organizers focused on 79 categories of relations taken from Bejar et al. (1991), which can be partitioned into the ten broader categories listed in Table 1. The task of obtaining word pairs that match closely with each type of relation was crowd-sourced to Amazon Mechanical Turk in two phases. In the first phase, participants were shown a description of the relation along with several prototypical word pairs. Then, they were asked to provide additional word pairs belonging to the same relation. The second phase focused on determining the similarity of each word pair to the relation. Participants were shown a description of the relation, several prototypical word pairs, and a set of four word pairs collected in Phase 1. They were then asked to choose both the word pair among those four which best represented the relation, and the word pair which least represented the relation. Each word pair appeared in multiple

Word	Semantic class distribution									
	44	17	13	47	24	32	36	41	3	45
sadness	.71	.07	.04	.04	.02	0	0	0	0	0
happiness	.73	.02	.01	0	0	0	0	0	0	0
sorrow	.75	.02	.01	.06	0	0	0	0	0	0
terror	.06	0	0	.26	0	0	0	0	0	.47
amusement	.19	0	0	0	0	.54	0	0	0	0
agreement	.01	.08	0	0	0	0	.03	0	.82	0
smile	.04	.40	.09	0	.33	0	0	0	0	0
nod	0	.42	.03	0	.02	0	.11	0	0	0
laugh	.01	.23	.31	0	.15	0	0	0	0	0
kiss	.04	.22	.19	0	.09	0	.03	.12	0	0
intoxicate	0	0	.2	0	0	0	0	.55	0	0

Table 2: A portion of the semantic class distribution vectors for several words participating in word pairs belonging to the REFERENCE:*Expression* relation.

Phase 2 questions, sometimes being chosen as the most representative, and other times being chosen as the least representative. This setup is known as a MaxDiff (Louviere and Woodworth, 1991) problem and is effective at deciding an absolute ranking among items without requiring participants to order all items.

Using the data collected from Amazon Mechanical Turk, the organizers were able to create a ranked list of word pairs for each relation in the following manner. Each word pair was assigned a score equal to the percentage of times it was chosen as the most representative minus the percentage of times it was chosen as the least representative. The word pairs were then ranked based on this score. An example ranking is shown in Figure 2. The goal of the SemEval task was to most accurately reproduce this ranking using automatic methods.

Word Pair	Similarity
laugh:happiness	50
nod:agreement	46
laugh:amusement	44
tears:sadness	44
crying:sadness	40
tears:sorrow	36
laughter:amusement	34
scream:terror	26
lie:dishonesty	16
laugh:hilarity	14
yawn:boredom	8
frown:discontent	6
frown:sadness	-2
sigh:exhaustion	-8
frown:anger	-28
wink:friendliness	-48
exhaustion:sigh	-50
anger:slap	-56
hilarity:laugh	-58
discourse:relationship	-60
friendliness:wink	-68

Figure 2: Ranking of a subset of the word pairs for the relation REFERENCE:*Expression* chosen by participants

5 Measuring selectional preference agreement

In order to measure how well a word pair matches the selectional preferences of a relation we must first model the selectional preferences for each argument of each relation. This is done using the induced semantic word classes described earlier.

We model the selectional preferences for an argument position of a relation using a distribution over semantic classes. These distributions are determined by first gathering all of the word pairs belonging to a relation (as collected in Phase 1). For each word pair $w1$:$w2$, we retrieve the semantic class distribution associated with each word (θ_{w1} and θ_{w2}). The distributions for all of the words appearing as a first argument are then averaged to obtain a class distribution for the first argument, which we call σ_1. This is repeated to obtain a distribution for the second argument as well to obtain σ_2. We then repeat this procedure for all relations to obtain selectional preferences for them. The assumption we make about our dataset is that the average word pair which needs to be ranked is representative of the arguments for that relation, or at least, that the contributions of non-representative word pairs will not overwhelm the contributions of those which are representative.

Measuring the agreement of a single word to the selectional preferences of a relation is then done by comparing the semantic class distribution associated with the word (θ_w) to the average distribution computed for that argument position (σ_1 or σ_2) of the relation. We consider several possible vector similarity metrics such as cosine similarity. Table 2 shows the most significant elements of the semantic class distributions (θ_w) for several words participating in the REFERENCE:*Expression* relation. The top half of Table 2 shows distributions for words participating in the second argument of a word pair, and the bottom half shows words participating in the first argument of a word pair (except *intoxicate*). Table 2 illustrates that similar words have similar vectors in the induced semantic class space. The words *sadness*, *happiness*, and *sorrow* are semantically similar. We have also included an outlier word *intoxicate* to show that the words in the bottom of the figure were not similar simply because they were all verbs. The distribution for *intoxicate* is zero for many of the classes that are significant for the other words confirming that we are capturing semantics beyond just part of speech.

We model relational similarity (how closely a word pair belongs to a relation) using only the selectional preferences of the relation. For a word pair w_1:w_2 we measure its relational similarity to relation r as:

$$sim(r, w_1 : w_2) = \frac{s(\theta_{w_1}, \sigma_{r,1}) + s(\theta_{w_2}, \sigma_{r,2})}{2} \quad (1)$$

where θ_w is the LDA-induced semantic class distribution for word w, $\sigma_{r,n}$ is the selectional preference distribution for the n^{th} argument of relation r, and s is a similarity measure between vectors.

The measure in (1) compares each word pair's semantic class distributions against the average for all word pairs assigned to a relation. The similarities for all word pairs belonging to a relation are computed and the pairs are then ranked. Next, this ranking is compared against the ranking produced by annotators such as the ranking in Figure 2. Note that only the order in the ranking is considered, the particular similarity values are not. We chose to average over all class distributions for an argument position to capture a soft membership of each class to the selectional preferences. We evaluate several vector similarity measures in the next section.

6 Evaluation of the relational similarity method

For training our LDA model we used a corpus consisting of the 8 million documents from English Gigaword (LDC2009T13) (Parker and Consortium, 2009) and the 4 million documents from the 2011-12-01 dump of Wikipedia[1]. The dependency parses were obtained by using the Stanford dependency parser[2] (De Marneffe and Manning, 2008). The textual content from the Wikipedia XML files was extracted using WP2TXT (http://wp2txt.rubyforge.org/). Due to the large size of this corpus we used a parallel implementation of LDA known as PLDA (Liu et al., 2011) across eight quad-core machines. The parameters for the LDA were the suggested defaults of $\alpha = 0.1$ and $\beta = 0.01$. We arbitrarily chose 50 topics, but this is clearly a parameter that requires further investigation. Additionally, our input to the LDA only consisted of 3,357 pseudo-documents, corresponding to all of the unique words in all of the word pairs that we were interested in ranking. While this contains many commonly used words in English, many other words are not covered and the data would have to be expanded for use in other tasks.

We used the official testing set from the SemEval 2012 Task 2 (Jurgens et al., 2012), which consisted of 69 relations (another ten were released for training but we do not make use of them). The relations had an average of 40 word pairs, ranging from 25 to 45. We evaluate the performance of the relational similarity model using a Spearman correlation score between the model's word pair ranking and the ranking produced by the annotation effort. This is the same evaluation metric used during the official SemEval 2012 Task 2 (Jurgens et al., 2012). Table 3 shows the results of our approach under several common similarity measures. We expected the measures designed for probability distributions (Jensen Shannon/Hellinger) to perform best, however our evaluation showed that vector space metrics (cosine/Tanimoto) performed slightly better. During the official evaluation, the best performing system achieved a correlation of 0.229. The model presented in this paper achieved a significantly higher correlation of 0.334 using the Tanimoto

[1]http://dumps.wikimedia.org/
[2]http://nlp.stanford.edu/software/lex-parser.shtml

Model	Correlation
Best SemEval 2012 system	0.229
Jensen Shannon divergence	0.324
Hellinger distance	0.326
Cosine similarity	0.332
Tanimoto coefficient	0.334
Generalized Dice coefficient	0.307

Table 3: Spearman's correlation scores between rankings produced by our approach over different similarity metrics and the gold rankings made available for SemEval 2012 Task 2

metric, which is similar to cosine similarity defined as:

$$Tanimoto(a, b) = \frac{a \cdot b}{||a||^2 + ||b||^2 - a \cdot b} \quad (2)$$

The effectiveness of the simple model presented in this paper shows two things: (1) an LDA model can be used effectively to induce semantic classes from English text using dependency parse contexts, and (2) that those semantic classes can be used to model selectional preferences in semantic relations. These results also show the high importance of selectional preference agreement when measuring the degree to which a pair of words belongs to a semantic relation. This model outperforms reported results, without taking into consideration the actual relation between the two arguments of a word pair. Future work will involve combining the selectional preferences approach with a approach that also models the dependence between the two arguments.

7 Analysis of the induced semantic classes

We first present a manual inspection of the semantic class space that was induced by the LDA, followed by a more analytical evaluation. Table 4 illustrates the top dependency contexts associated with four semantic classes that were prominent for relation REFERENCE:*Expression* in Table 2. Table 5 shows the top words associated with the same four semantic classes. All of the top 16 words for class 44 are categorized as abstract entities in WordNet. Many of them can be further categorized as states (*independence*, *love*, *freedom*, *confidence*, *security*). We can see from the top dependency contexts of class 44 listed in Table 4 the types of contexts which indicate a state: ←prep_of lack, ←prep_of level, ←prep_of sense. From Table 2 we can see that class 44 is the predominant class for several emotional states participating in the first argument of a REFERENCE:*Expression* relation, so it is reassuring to see that this class consists of states.

Class 44	Class 17	Class 13	Class 24
access	day	take	white
progress	time	come	red
confidence	man	mean	black
independence	game	done	light
ability	victory	look	blue
freedom	question	understand	green
relationship	number	love	hair
responsibility	deal	call	suit
experience	member	give	rain
growth	team	ask	color
future	case	live	yellow
strength	state	agree	breeze
authority	sign	concerned	dress
love	person	remember	flag
security	record	read	shirt
life	attack	hear	smoke

Table 5: The top words (descending) occurring with with semantic classes 44, 17, 13, and 24.

The words in class 17 seem less related, but have some broad similarities. For instance, they appear to be countable nouns expressed in the singular form. When we examine the dependency contexts for class 17 we can understand why this is. The contexts include →det another, →amod first, →det every, →amod only, etc. These determiners and adjectives cannot modify mass nouns and the set of top words for the class do appear to fall in the category of countable nouns.

Class 44	Class 17	Class 13	Class 24
←prep_of lack	→det another	→nsubj I	→amod white
←nn process	→amod first	←ccomp said	→dobj wearing
→amod economic	→amod big	→neg n't	→amod black
←nn talks	←prep_of kind	→punct ”	←prep_of pair
→amod political	←prep_of part	→neg not	→amod red
←prep_of kind	←dobj made	→nsubj they	→conj_and white
←prep_of level	→det every	→nsubj we	→amod green
→amod national	→amod only	→nsubj you	→amod blue
→amod great	→det any	→nsubj We	←conj_and red
←dobj expressed	→amod single	→aux do	→amod calm
←prep_of sense	→amod second	→nsubj he	←amod light
→amod public	→amod major	→aux to	←dobj wear
←dobj claimed	→det each	→complm that	←conj_and black
←nn plan	←nsubj came	→aux did	→amod dark
←nn agreement	→amod biggest	→aux does	→amod heavy
←dobj made	→amod great	→aux would	←dobj wore
←prep_of loss	→predet such	→nsubj They	←appos C.
→amod social	←dobj make	→nsubj who	←amod chips
←dobj give	←nsubj 's	→nsubj people	←prep_in dressed
→amod full	→advmod just	→nsubj You	→punct
←prep_of moment	←dobj has	→dobj it	←amod card

Table 4: The top dependency contexts for semantic classes 44, 17, 13, and 24. Some contexts which are common across many semantic classes were omitted.

Semantic class 13 consists largely of actions taken by humans. The dependency contexts reveal how this cluster came about: →nsubj I, →neg n't, →nsubj they, →aux would, etc. These dependencies apply to verbs, and many of them specifically contain pronouns (you, I) reserved primarily for humans. From Table 2 class 13 was largest for the "expression" words *smile*, *nod*, *laugh*, *kiss* which obviously are actions usually preformed by humans.

Semantic class 24 appears to contain words which are often described using colors or shades (e.g., dark, light). Examples for colors would include white flag, white suit, black smoke, while examples for shades would include dark hair and dark shirt, but also colors themselves as in dark green and light blue.

Overall, it appears that using the LDA model on dependency contexts performed well at clustering words into semantic classes, picking up on common-place but subtle linguistic phenomena such as countable nouns, and whether a verb tends to have a person as a subject.

We now present a more quantitative assessment of the induced semantic class space. We follow the evaluation proposed by Widdows and Dorow (2002). They selected the ten categories of objects shown in the first column of Table 6, along with a prototypical member word for each category. Using the prototype word as a seed, its twenty nearest neighbors are determined. The most appropriate distance metric for our approach is to use the Tanimoto coefficient between the semantic classes distributions of two words. The lists of nearest neighbors produced using our induced class distributions are illustrated in Table 6. Neighbors which are not subsumed by the WordNet synset represented in the first column have been italicized. Our method achieves a precision of only 59.4% on this evaluation. The results are considerably below previous approaches which have achieved 82% (Widdows and Dorow, 2002) and 90.5% (Davidov and Rappoport, 2006), however our method has several disadvantages in this comparison. Firstly, we have only generated semantic class vectors for the 3,357 words which occurred in the word pairs in the relation dataset which limits our recall. This particularly affects the retrieval of "easy" but rare neighbors of a word such as *fortepiano* from the seed *piano*. This also caused us to choose different seed words for the categories crimes, body parts, and academic subjects because the seeds used in prior literature did not

Class	Seed Word	Neighbors
crimes	theft	*abuse destruction* rape infringement *crimes violence* crime *explosion famine eruption scandal discrimination accident* assault *crash damage punishment controversy snowstorms slavery*
places	park	mall zoo hall marina stadium castle mill airport aquarium cafeteria hotel factory warehouse gym firehouse restroom shrine house casino garage
tools	screwdriver	knife trowel *mattress* spatula broom *stool* scalpel *flashlight stethoscope pillow microphone leash pouch beaker lid faucet pane fingertip glove scepter*
vehicle conveyance	train	ship *link* craft bus truck boat van airplane *route highway* wagon *mountain* vessel vehicle car *engine* kayak sedan rocket
musical instruments	piano	violin clarinet cello guitar flute *rock* bass *fairy jazz blues television art computer music* keyboard *dance soap opera cinema* Throughout
clothes	shirt	hat sweater frock blouse *earring* wig yarmulke *tiara* coat scarf *necklace bracelet* skirt *breeze eyeshadow* burka pants sandal ballpoint
body parts	neck	wrist ear finger nose waist mouth spine toe *glove coffin* foot eye *couch* hands fingers *door* During penis legs *lawn*
academic subjects	philosophy	geography logic chemistry religion composition psychology anatomy algebra architecture *voice vision* geometry genealogy *image* discourse art *memory signature* history *conception*
foodstuffs	cake	egg salad apple *cane* pie soup *blender* carrot *leaf* omelette *cigarette* pizza *pot polymer* dish beer *oven glass* dessert

Table 6: The nearest neighbors for nine seed words. Italics mark words which do not match the class of the seed word.

appear in the word pair corpus. Secondly, the previous approaches utilizing this evaluation metric have limited their class induction space to only nouns. Therefore, the candidate neighbors under the previous approaches are restricted to nouns, whereas our approach conflated words with the same surface form, but different parts of speech. The effects of this are quite clear for the tools category. Certain tool words which are also used as verbs are absent from our top neighbors such as rake, plow, and shovel, however they are top neighbors of each other. Both of these limitations can be alleviated, but are not addressed in this paper. We believe the results from Table 6 show that our semantic space based on an LDA model and Tanimoto coefficient do correspond to a semantic class space. While alternative semantic class induction techniques may improve our relational similarity results, this approach does show the merit in modeling the relational selectional preferences by semantic class membership of the relation arguments.

8 Conclusion

We showed that a simple model based on LDA using dependency parse contexts can be used effectively to model selectional preferences of semantic relations. Further, we can achieve state of the art results for measuring relational similarity by using only the agreement between a word pair and the expected semantic classes for the relation's arguments. While there remains more work to be done towards incorporating additional types of information beyond just argument semantic classes, our current results are promising. Future improvements to the method would include the use of word senses (or simply part of speech) information to form more semantically coherent classes, and incorporating information about relations into the semantic class induction process.

References

Baroni, M. and Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Comput. Linguist.*, 36(4):673–721.

Bejar, I. I., Chaffin, R., and Embretson, S. E. (1991). *Cognitive and psychometric analysis of analogical problem solving. Recent research in psychology.* Springer-Verlag Publishing.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.

Davidov, D. and Rappoport, A. (2006). Efficient unsupervised discovery of word categories using symmetric patterns and high frequency words. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 297–304. Association for Computational Linguistics.

De Marneffe, M. C. and Manning, C. D. (2008). The Stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8. Association for Computational Linguistics.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2):155 – 170.

Holyoak, K. and Thagard, P. (1996). *Mental leaps: Analogy in creative thought.* MIT press.

Jurgens, D. A., Mohammad, S. M., Turney, P. D., Holyoak, and J., K. (2012). SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012).*

Lin, D. and Pantel, P. (2001). Induction of semantic classes from natural language text. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, pages 317–322, New York, NY, USA. ACM.

Liu, Z., Zhang, Y., Chang, E. Y., and Sun, M. (2011). PLDA+: Parallel Latent Dirichlet Allocation with Data Placement and Pipeline Processing. *ACM Transactions on Intelligent Systems and Technology, special issue on Large Scale Machine Learning*.

Louviere, J. J. and Woodworth, G. G. (1991). Best-worst scaling: A model for the largest difference judgments. Technical report, Working paper. University of Alberta.

Mechura, M. (2008). *Selectional Preferences, Corpora and Ontologies*. PhD thesis, University of Dublin.

Medin, D., Goldstone, R., and Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, 1(1):64–69.

Ó Séaghdha, D. (2010). Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL-10)*, Uppsala, Sweden.

Pantel, P. (2003). *Clustering by committee*. PhD thesis, Citeseer.

Pantel, P., Bhagat, R., Coppola, B., Chklovski, T., and Hovy, E. H. (2007). ISP: Learning Inferential Selectional Preferences. In *North American Chapter of the Association for Computational Linguistics*, pages 564–571.

Parker, R. and Consortium, L. D. (2009). *English gigaword fourth edition*. Linguistic Data Consortium.

Rahman, A. and Ng, V. (2010). Inducing fine-grained semantic classes via hierarchical and collective classification. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, pages 931–939, Stroudsburg, PA, USA. Association for Computational Linguistics.

Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61(1-2):127 – 159. ¡ce:title¿Compositional Language Acquisition¡/ce:title¿.

Rink, B. and Harabagiu, S. (2012). UTD: Determining Relational Similarity Using Lexical Patterns. In *Proceedings of the 6th International Workshop on Semantic Evaluation (SemEval 2012), in conjunction with the First Joint Conference on Lexical and Computational Semantics (* SEM 2012)*.

Ritter, A. and Etzioni, O. (2010). A latent dirichlet allocation method for selectional preferences. In *In Proceedings of the Association for Computational Linguistics ACL2010*.

Roark, B. and Charniak, E. (1998). Noun-phrase co-occurrence statistics for semiautomatic semantic lexicon construction. In *Proceedings of the 17th international conference on Computational linguistics - Volume 2*, COLING '98, pages 1110–1116, Stroudsburg, PA, USA. Association for Computational Linguistics.

Turney, P. D. (2005). Measuring Semantic Similarity by Latent Relational Analysis. In *Proceedings of the 19th international joint conference on Artificial intelligence*, pages 1136–1141.

Turney, P. D. (2006). Similarity of semantic relations. *Comput. Linguist.*, 32(3):379–416.

Turney, P. D. (2008). A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*, COLING '08, pages 905–912, Stroudsburg, PA, USA. Association for Computational Linguistics.

Widdows, D. and Dorow, B. (2002). A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1*, pages 1–7. Association for Computational Linguistics.