A Supervised Approach for Enriching the Relational Structure of Frame Semantics in FrameNet

Shafqat Mumtaz Virk and **Philippe Muller** and **Juliette Conrath**
IRIT, Toulouse University
118 Route de Narbonne
31062 Toulouse, France

Abstract

Frame semantics is a theory of linguistic meanings, and is considered to be a useful framework for shallow semantic analysis of natural language. FrameNet, which is based on frame semantics, is a popular lexical semantic resource. In addition to providing a set of core semantic frames and their frame elements, FrameNet also provides relations between those frames (hence providing a network of frames i.e. FrameNet). We address here the limited coverage of the network of conceptual relations between frames in FrameNet, which has previously been pointed out by others. We present a supervised model using rich features from three different sources: structural features from the existing FrameNet network, information from the WordNet relations between synsets projected into semantic frames, and corpus-collected lexical associations. We show large improvements over baselines consisting of each of the three groups of features in isolation. We then use this model to select frame pairs as candidate relations, and perform evaluation on a sample with good precision.

1 Introduction

In the area of formal linguistics, frame semantics is a theory of meanings, which was introduced by Charles J. Fillmore and his colleagues back in the early 1980's. Frame semantic analysis (Das et al., 2014), which is based on frame semantics, itself is a type of shallow semantic analysis in which the focus is on predicates and their arguments. Frame semantic analysis is abstracting away from single verbal predicates, used in other semantic analysis approaches such as Propbank-based work (Kingsbury and Palmer, 2012), to "semantic frames" (Fillmore, 1982). The idea in frame semantics is to group predicates referring to similar events, processes, and/or object types under the umbrella term 'a semantic frame', which can be expressed with different parts of speech and with arguments that can have various syntactic realizations. For instance, the sentence in example (1-a) could be seen as introducing a commerce frame, indicated by the lexical unit *buy*. This frame has a set of necessary arguments, a *buyer*, some *goods*, and a *seller*, which are realized by role fillers *Lester*, *a car* and *Jimmy* in the example. Frames can be realized with different lexical units and syntactic constructions, so that sentences (1-b)-(1-c) would have a similar meaning with respect to the frame in question.

(1)   a.   [Lester]$_{\text{buyer}}$ [bought]$_{\text{commerce}}$ [a car]$_{\text{goods}}$ [from Jimmy]$_{\text{seller}}$.
           b.   [Jimmy]$_{\text{seller}}$ [sold]$_{\text{commerce}}$ [a car]$_{\text{goods}}$ [to Lester]$_{\text{buyer}}$.
           c.   [Lester]$_{\text{buyer}}$'s [purchase]$_{\text{commerce}}$ of [Jimmy]$_{\text{seller}}$'s [car]$_{\text{goods}}$ (...)

Most recent NLP work on such semantic frames are based on FrameNet (henceforth FN) (Baker et al., 1998), a resource listing frames, how they can be evoked, and their argument types. FrameNet also lists how the frames are organized with respect to each other, with a set of frame-to-frame relations. If FrameNet based analysis has proven useful for certain tasks such as information extraction (Surdeanu et al., 2003), question-answering (Shen and Lapata, 2007), or coreference resolution (Ponzetto and Strube, 2006), its frame structure is assumed to be useful on its own for semantic analysis (Burchardt et al.,

2005), or paraphrase extraction (Hasegawa et al., 2011). It also provides additional information for semantic analysis in contextual role linking, as demonstrated in (Li et al., 2015), or to improve prediction of roles, as in the work of Kshirsagar et al. (2015). The relations are also useful to build thesauri and to retrieve semantically related words (Ruppenhofer et al., 2006). However, one big issue with FrameNet is its partial coverage of the lexicon and the intended set of frames, which translates also into a partial coverage of frame relations. Numerous studies address FrameNet's lack of lexical coverage (Pennacchiotti et al., 2008; Das and Smith, 2012; Pavlick et al., 2015). In contrast, little work has been done on extending frame relations except by Ovchinnikova et al. (2010), in which cluster of frames are proposed based on collocation information, along with principles to be respected when adding new relations, or Pennacchiotti and Wirth (2009), which defines a generic notion of "relatedness" between frames, with no semantics to compare this to FrameNet relations. Both approaches are unsupervised, and provide little evaluation of the relevance to the intended FN structure.

We present a supervised approach to enrich FrameNet's relational structure by training a model on the existing set of frame relations and a rich set of features combining linguistic and structural information. Further, we also leverage external resources such as WordNet. The resulting model is used to predict new frame-to-frame relations whose validity is then manually evaluated. The rest of the paper is organized as follows: Section 2 presents in more detail FN frame relations, Section 3 presents our methodology and the three group of features we use for training a supervised model. Section 4 presents our experimental results based (1) on a separate test set taken from already existing FrameNet relations, and (2) a human evaluation of newly proposed relations.

2 FrameNet Relations

FrameNet is a lexical semantic resource which is based on the theory of frame semantics (Fillmore, 1982). A set of script-like descriptions, known as semantic frames, of real world situations is collected and maintained along with the participants of those situations. Each of the semantic frames has a set of associated words (known as frame evoking elements, FEE) which can evoke a particular predicate. The participants of a situation (called frame elements, FE instead of the more classical term 'semantic role') are also identified for each frame. In addition, each semantic frame is coupled with example sentences taken from naturally occurring natural language text. FrameNet-1.5 has 1230 semantic frames, 11829 lexical units, and 173018 example sentences.

As mentioned previously, FrameNet has defined a set of frame-to-frame relations, and has proposed connections of certain frames to certain other frames, thus providing a network. The backbone of this network is built on a hierarchy of predicate types, and typical sub-events of specific situations. Event participants, also called frame elements, of the connected frames may also have one-to-one connections. Some of these frames are introduced purely for the coherence of the structure and may not have associated lexical items ("unlexicalized frames"). The following is a list of relations defined between two frames X and Y:

- Subframe(X,Y): a complex scenario X, e.g. CRIMINAL_PROCESS, is composed of typical sub-events (e.g. Y can be TRIAL, SENTENCING...).

- Precedes(X,Y): X is before Y in a typical scenario, e.g. TRIAL precedes SENTENCING.

- Inheritance(X,Y): A relation between frames where one frame is a more specific version of another frame e.g. ANIMALS inherits from BIOLOGICAL_ENTITY.

- Causative_of(X,Y): X is a potential cause of Y, e.g AIMING is causative_of HIT_OR_MISS.

- Inchoative_of(X,Y): X is an inchoative of Y, e.g. ROTTING is inchoative of BEING_ROTTEN.

- Perspectivized_in(X,Y): X represents a specific perspective of Y, e.g. COMMERCE_BUY and COMMERCE_SELL are two different perspectives of COMMERCE_TRANSFER_GOODS.

- Using(X,Y): X somehow involves Y, e.g. PEOPLE_BY_AGE uses the AGE frame.

- See_also(X,Y): Frames that have some similarities, but need to be differentiated carefully (e.g. MEASURABLE_ATTRIBUTE and DIMENSION).

Figure (1) gives an example of a subset of relations around the CRIMINAL PROCESS Frame, that will be used for illustration in the rest of the paper.
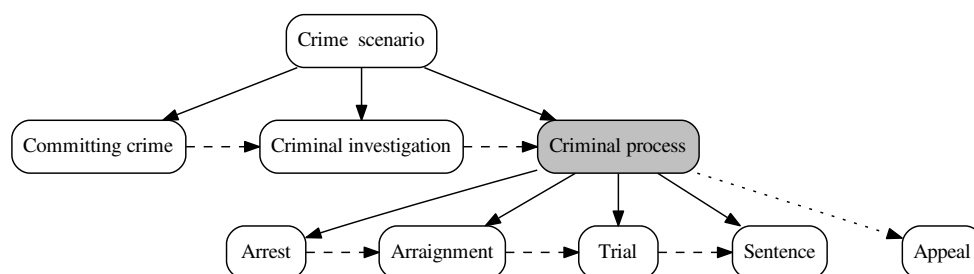

Figure 1: Example sub-graph around the CRIMINAL PROCESS frame; Solid, dashed and dotted lines respectively indicate the "SubFrame", "Precede" and "Using" relations. Only subframes of CRIMINAL PROCESS are shown.

3   Experimental Design and Features

Our goal is to provide a method to enrich FrameNet structures with new relations where supervision is given by the existing frame relations. Our experimental methodology is thus two-fold: 1) we trained a discriminative model to predict frame-to-frame relations in a supervised setting using the existing relations listed in the original FrameNet and selected counter-examples; and 2) we applied this model to predict likely frame-to-frame relations that are not already listed. We used rich features from three different sources, taking inspiration from the unsupervised relatedness measures of (Pennacchiotti and Wirth, 2009), which are based on lexical collocations, but adding more features: structural features from the existing FrameNet network, information from WordNet relations between synsets projected into semantic frames, and different additional corpus-collected lexical associations. Since we want to focus on event-denoting predicates, which have arguably richer and potentially more interesting argument structures, and are more likely to be related in common scenarios, we restricted ourselves to frames with at least one verb trigger. This restricts the existing relations to 824 frame pairs. Next we describe the different set of features used in our experiments.

3.1   WordNet Based Features

WordNet provides useful lexical and semantic relations between synonym sets. The most important among those relations are hypernymy/hyponymy and meronymy/honolymy. Hypernym and hyponym (and troponymy for verbs) are super-subordinate relations and link more general synsets to specific ones and could be relevant for Frame inheritance, and thus make the first source of information for predicting frame relations. As we are interested in relationships between pairs of frames, instead of pair of synsets, we need to transfer knowledge about relations between synsets to useful features between frames whose verbs appear in WordNet synsets. There have been attempts at matching WordNet sense inventory to FrameNet frames (Shi and Mihalcea, 2005; Tonelli and Pighin, 2009) but since both resources are incomplete, we decided to compute frame relatedness using an existing WordNet relation between synsets which include verbs from the given frames, irrespective of their senses. This certainly introduces some noise, that we hope to control with redundancies in frame-to-frame associations. For our purposes, we have divided WN-based features into two groups: (1) occurrence-based features (2) similarity-based features. Occurrence-based features are simple counts of existence of a particular relation between a pair

from all possible pairs of all senses of all lexical units of the two candidate frames. By taking the existence of a particular relation between senses and summing them up, we are projecting the knowledge about WordNet's synset relations to FrameNet's frames.

For a given pair of frames (F_i, F_j), and a given WordNet relation R (i.e. hypernym, hyponym, all meronym relations, all holonyms relations, attributes, entailments, causes, also_sees) the number of occurrence-based feature values were computed using the following formula:

$$number\_rel(F_i, F_j, R) = \sum_{(lu_i, lu_j) \in LU_{F_i} \times LU_{F_j}} \left[\sum_{(s_{lu_i}, s_{lu_j}) \in S_{lu_i} \times S_{lu_j}} R(s_{lu_i}, s_{lu_j}) = True \right]$$

LU_{F_i} and LU_{F_j} are sets of lexical units of frames F_i and F_j respectively, while S_{lu_i} and S_{lu_j} are sets of various senses of the corresponding lexical units in WordNet. The values of the semantic similarity based features were computed using a similar formula (but averaged) for a given similarity function f and two frames F_1 and F_2:

$$similarity(F_i, F_j, f) = \frac{1}{|LU_{F_i} \times LU_{F_j}|} \times \sum_{(lu_i, lu_j) \in LU_{F_i} \times LU_{F_j}} \left[\frac{\sum_{(s_{lu_i}, s_{lu_j}) \in S_{lu_i} \times S_{lu_j}} f(s_{lu_i}, s_{lu_j})}{|S_{lu_i} \times S_{lu_j}|} \right]$$

The semantic similarity measures are classical wordnet similarity measures: Path, Wu-Palmer (Wu and Palmer, 1994) and Leacock-Chodorow (Leacock and Chodorow, 1998), computed using NLTK's (Bird, 2006) Python interface. They all take into consideration the path between synsets in WordNet's hierarchy with different normalization factors. For instance Wu-Palmer is based on the depths of two synsets in the WordNet hierarchy along with the depth of the least common subsumer. The path similarity measure is a semantic association measure which is based on the shortest path that connects two synsets in the taxonomy in which the senses occurred. LCH similarity takes into consideration the depth of the taxonomy in addition to the shortest path, and is computed as $-log(p/2d)$, where p denotes the shortest path and d is the depth. For the purpose of projecting WordNet's synset knowledge to FrameNet's frames, we are summing up and averaging the corresponding semantic similarity values of all possible combinations of various senses of all lexical units of the two candidate frames.

3.2 FrameNet Based Features

FrameNet network structure can also be a useful resource to compute frame-relatedness between non explicitly related frames. (Pennacchiotti and Wirth, 2009) proposed to apply equivalents of Wu-Palmer similarity and Hirst-St.Onge's (Hirst and St-Onge, 1998) measure to FrameNet's hierarchy. They also suggested to take into account frame element overlap (the proportion of predicate roles having the same name). Here, the frame element overlap similarly is based on the number of frame-elements (role names) shared by the two candidate frames. Hirst-St.Onge measure is a semantic similarity measure which is based on the path connecting two WordNet synsets and how often one needs to change direction to reach one from the other in the network. A "change of direction" would be for instance a path along an hypernym relation then an hyponym relation (which yields a likely co-hyponym). The intuition is that two synsets are semantically closer if they are connected through a "not too long path which does not change direction too often". This can be applied similarly over FrameNet's relations, even though some of these are vaguer than WordNet relations. An example path in figure 1 would be the path (SubFrame^{-1}+Using) between SENTENCE and APPEAL.

In addition to the hierarchy-based features, frame verbal definitions can potentially provide useful information for frame relatedness, and we decided to also use a definition overlap as a feature, with a Jaccard between the sets of open-class words of each definition (open(F) being the set of open-class words in F definition):

$$def\_overlap(F_1, F_2) = \frac{\|open(F_1)\| \cup \|open(F_2)\|}{\|open(F_1)\| \cap \|open(F_2)\|}$$

3.3 Corpus-Based Features

In addition to the previous sets of features, which make use of existing resources or of the existing structure we want to expand, we derived clues from large corpora where frame lexical units can be observed in the same contexts, indicating a potentially regular relation between their corresponding frames. We collected both general ococcurences, in the spirit of (Pennacchiotti and Wirth, 2009), and targetted associations of verbs in explicit discursive contexts, inspired by (Conrath et al., 2014).

Frame Coocurrences We used statistics of frame co-occurrences from a large-scale corpus to predict frame relatedness, relying on the intuition that related frames tend to co-occur more often in the same context within a given corpus. The context could be either a document, a sentence, or a specific number of sentences. We computed the co-occurrence of frames as point-wise mutual information (PMI) within the GigaWord corpus (Graff et al., 2003). Similar information has already been used in (Pennacchiotti and Wirth, 2009), but we deal with the frame ambiguity problem differently.

Ideally, one would like to compute frame co-occurrences statistics from a frame annotated corpus which is big enough for machine learning tasks. FrameNet does provide a frame annotated corpus, but it is too small to be truly useful for making generalizations. A workaround is to use an unannotated corpus together with FN's frame-evoking lexicon to decide which frame is being triggered by a particular lexical unit. The problem is now to resolve the ambiguity in cases where a lexical unit can potentially trigger more than one frame. (Pennacchiotti and Wirth, 2009) suggested a weighted co-occurrence measure, which gave lower weights to the co-occurrence of ambiguous words. The probabilities of sense occurrences of lexical units were learned from the WordNet sense tagged corpus SemCor. Our approach is slightly different in the sense that instead of learning word-sense probabilities first and then mapping it to frames, we directly learn the probabilities of lexical units triggering particular frames from FrameNet's annotated corpus using the ratio of the number of frames triggered by a lexical unit lu in the FrameNet corpus and of the total number of occurrences of lu in the corpus. This is arguably more direct and only use FrameNet information, although both approaches need annotated data. The probabilities are then used to compute a weighted PMI between frames (the weighting function simply sums up the probabilities of a lexical unit triggering a particular frame F over the entire GigaWord).

Lexical Coocurrences We also used as features measures of semantic similarity derived from corpus-based specific associations between lexical items. If we can find valid lexical associations to match the targeted frame relations, we can obtain relevant association measures. We adapted the method of (Conrath et al., 2014), in which semantic associations between verbs are derived from cooccurrences between two verbs and certain classes of discourse markers, using mutual information measures. This shallow discourse analysis can be seen as extraction of typical semantic relations between verbs. Using the Penn Discourse Treebank (Prasad et al., 2008) list of markers, grouped into semantic classes, we computed six types of associations between verbs: (1) causal, (2) temporal, (3) continuation, (4) contrast, (5) disjunction, (6) elaboration. Each corresponds to a coherent set of markers and can provide clues to corresponding FrameNet relations: causative-of for (1), precede for (2) and (3), subframe for (6), and (4) and (5) for some form of looser relatedness that is sometimes encoded as "see-also" or "using" in FrameNet. For each class we implemented the three best measures according to (Conrath et al., 2014), which correspond to different normalizations or combinations of the verbs and discourse relation cooccurrences : normalized PMI (Evert, 2005), a specificity measure taken from (Mirroshandel et al., 2013), and a combined association measure they call w_combined. They are supposed to capture different aspects of the targeted lexical associations. Let $P(V_1, V_2, R)$ be the probability of the association of the two verbs V_1, V_2 with the given semantic relation R:

$$\mathrm{NPMI}(V_1, V_2, R) = \mathrm{PMI}(V_1, V_2, R)/(-2\log(P(V_1, V_2, R)))$$

$$specificity(V_1, V_2, R) = \frac{1}{3} \times \left(\frac{P(V_1, V_2, R)}{\sum_i P(V_1, V_i, R)} + \frac{P(V_1, V_2, R)}{\sum_i P(V_i, V_2, R)} + \frac{P(V_1, V_2, R)}{\sum_i P(V_1, V_2, R_i)} \right)$$

$$W_{combined}(V_1, V_2, R) = \frac{1}{3}(w_{V_1} + w_{V_2} + w_R)$$

with: $w_{V_1} = \frac{P(V_1, V_2, R)}{\max\limits_{i}(P(V_i, V_2, R))}$, $w_{V_2} = \frac{P(V_1, V_2, R)}{\max\limits_{i}(P(V_1, V_i, R))}$, and $w_R = \frac{P(V_1, V_2, R)}{\max\limits_{i}(P(V_1, V_2, R_i))}$.

We computed the corresponding association values for frame-pairs using the same averaging scheme as for the previous features over groups of lexical units, see 3.1

4 Experiments and Results

The previous features were computed for two sets of frame-pairs: (1) All those frames-pairs which have any of the frame-relations mentioned in section 2. As mentioned previously, we restricted ourselves to frames denoting events, i.e. which have at least one verb trigger. There are 824 such instances. (2) The set of all unrelated frame-pairs. From these two sets, a balanced set of 1648 frame-pairs (824 positive, and 824 negative) was collected, 10% of which was set aside for testing the chosen model. The remaining 90% of the balanced set was used to train and evaluate different models with different feature combinations. The best combination, evaluated by cross-validation, was evaluated on the test set, and then used to train a new model on the full balanced set, which we applied to all "unrelated" frame pairs (i.e. to predict likely frame relations between frame pairs that are not already listed in the FrameNet), with a fixed *a priori* confidence threshold in order to focus first on precision of the candidate pairs[1].

4.1 Training a Supervised Model

For the first part of our method, we trained a binary classifier to decide whether a given pair of frames is related or not. To provide negative examples, we randomly sampled the set of unrelated frame pairs. Of course we don't know the real proportion of frames that should be (ideally) related, but we suspect that such a relational structure is likely to be sparse, and thus taking random pairs should yield mostly truly unrelated pairs.

Assuming the actual relational structure is sparse, we probably face an imbalanced classification problem, in which we want to identify primarily the minority class. That is why we chose to balance the number of positive and negative instances during training, since it means we are doing majority class undersampling, a classic simple way of addressing class imbalance, and it is also a simple way to evaluate the relevance of our features. This is likely to generate too many candidate pairs on the test, degrading precision while helping recall of the positive class, something we can control *a posteriori* on new instances by imposing a confidence threshold on the classifier output (see below).

We used a Random Forest classifier with 1000 estimators and a minimum of 10 instances for splitting, and the implementation provided in the scikit-learn package (Pedregosa et al., 2011). Random Forest is a robust classifier used in a variety of tasks and perform best among the classifiers evaluated on cross-validation on the training set. As baselines, we tested each separate group of features: WordNet transfer, corpus cooccurences, and FrameNet features. Note that a majority/random baseline would give a score of 50% with the balanced setting we chose. We observed that the combination of features proves very useful, as cross-validation accuracies for the three groups are much lower than the full model. Interestingly, all groups of features seem necessary, as even removing the group which performs worse by itself lowers the training accuracy by 5% (ablations of each group of features were tried by cross-validation on the train only). Final results are presented in Table 1. Not surprisingly, the most informative features come from FrameNet itself, since positively related frames are more likely to be in "denser" parts of the network, at least those parts that have been the focus of the lexicographic work, and this can be seen as a bias of the model. Note however that (1) these denser areas of the network might still be incomplete and (2) the other sources of information definitely contribute to the overall result, adding 10 points over

---

[1]An archive can be found at `http://www.irit.fr/~Philippe.Muller/fn_structure.zip` which contains instances and their extracted features, the manually annotated sample of frame relation instances, and various scripts to reproduce the experimental results.

FrameNet based features. We provide confidence intervals on the scores (which are proportion estimations), and we can see the different models' intervals don't overlap much, even on the small test sample.

Model	Cross-Validation / training	Test	Confidence Interval
WordNet-based	60.3	63.4	2.5/7.4
Frame information	78.1	79.3	2.1/6.2
Corpus-based	64.8	60.4	2.4/7.5
Full Model	87.3	88.4	1.7/4.9

Table 1: Accuracy of predictions with balanced dataset (in %); confidence intervals $\pm$ are given at 95% for the train/test accuracy.

4.2 Finding New Relations

We used the previous supervised model on the set of all possible pairs of frames to predict potential new relations. Since the model has been trained on a balanced set, we probably over-generate positive labels, so as a first basic step to improve the precision of the extraction, we set an arbitrary cutoff for the confidence level according to the model, at the value 0.8. This yielded over 1500 pairs, out of which we randomly selected 100 pairs for validation. We mixed the pairs with an equal number of 100 "distractors" (random pairs absent from FrameNet), to prevent annotation bias. Note that, again, we have no *a priori* way of ensuring these distractors are not actual unknown relations, but we expect true relations to be rare in this set. Two of the authors then judged if one of the FN relations could hold, based on the summary of the relations from the FrameNet book, and a description of the frames as defined in FN, i.e. a short definition, sometimes a few example sentences, a list of frame element names (arguments of the frame), and a list of lexical units that can evoke the frame. They had to label each pair as Yes/No, and were asked to indicate a tentative label from the set of FrameNet relations (there could be 0, 1, or 2 relations according to their level of confidence in their decision). We did the same procedure on the top 100 frame pairs according to the classifier, also mixed with 100 random pairs. We computed inter-annotator agreement with Cohen's Kappa which was 0.65, generally considered as an acceptable agreement above 0.6, at least for semantic tasks. Raw agreement was 0.83. To estimate agreement on the labels, we consider judgments for common Yes decisions, and we considered the intersection of proposed labels (if two labels were proposed and one was common we counted that as a positive). Keeping only pairs for which each annotator gave an answer (only half of the pairs), the Kappa was at 0.5, mainly because of agreement on inheritance pairs. Given that a Kappa of 0.5 indicates moderate agreement, even with the lenient setting we provided, the relation types need some clarification. We can still tentatively conclude that the inheritance relation can be annotated, but in the remaining, we only discuss relatedness.

Frame 1	Frame 2
Building	Manufacturing
Evidence	Reasoning
Motion_directional	Path_shape
Cause_motion	Motion
Cause_impact	Cause_motion
Intentional_traversing	Path_shape
Discussion	Statement
Make_cognitive_connection	Relating_concepts
Destroying	Killing
Cause_harm	Cause_to_fragment

Table 2: Top ten true new relations predicted, according to the probability given by the model.

Frame 1	Frame 2
Filling	Removing
Change_position_on_a_scale	Motion
Bringing	Ride_vehicle
Communication	Evidence
Change_position_on_a_scale	Path_shape
Getting_vehicle_underway	Removing
Change_direction	Path_shape
Bringing	Change_direction
Change_direction	Removing
Change_direction	Departing

Table 3: Top ten false positive new relations predicted, according to the probability given by the model.

Both authors adjudicated the differences in the relatedness question, and only pairs with both positive judgments were finally considered positive. Out of the top 100 positive candidates, 63 were labeled as positive by human annotators giving a 63% precision, and 52 out of the 100 selected above the 0.8 threshold, with respectively ± 9.3 and $\pm 9.8\%$ confidence interval. This means a large proportion of proposed frame pairs could be considered for addition, and that proportion does not decrease too quickly when the confidence of the classifier decreases. We cannot estimate recall, as we don't know how many relations there should be overall, but we observed that only 4 pairs out of the distractors were judged positive, which seems to indicate relations are indeed rare with respect to all possible pairs. Note that estimating recall among a subsample of n frames would require a number of judgments equal to $n^2 \times$ the number of relation types. Table 2 and 3 shows a list of top 10 true and false positive relations predicted by our model.

We did an error analysis on the 37 false positive in the top 100 test set, and came to the conclusion that most frame pairs were picked up as related because they involve related lexical units (and often with different senses), rather obviously, and that they involve either (1) co-hyponyms at different levels of granularity; or (2) frame describing opposite events, such as Removing and Filling, which do not correspond to any FrameNet relation. Case (2) is arguably a FrameNet problem, since a lot of frames involve antonyms, and could indicate here missing frames at a higher level up the hierarchy of inheritance, an inconsistency noted in previous work (Hasegawa et al., 2011). As an example the two adjectives *easy* and *difficult* are part of the same Frame DIFFICULTY, while the verbs *empty* and *fill* are in separate frames EMPTYING and FILLING. Case (1) is an interesting perspective for improvement: relations should not be considered between frame pairs in isolation, but with respect to all existing or predicted relations, in a global way. If a frame F_1 is already related to a frame F_2, there should not be a relation between F_1 and, e.g., subframes of F_2. Note also that most Frames that are siblings in a subframe relation also have explicit relations, usually a "Precede" relation (see figure 1), but also vaguer links such as "Using".

5   Conclusion and Future Directions

Our main contribution is to present the first supervised method to provide candidate relations between FrameNet frames, using a rich set of structural and linguistic features inspired by previous unsupervised work which did not provide evaluation with respect to FrameNet's set of relations. A secondary contribution is in showing that Frame relation annotation is possible with good agreement, based only on Frame descriptions, although labeling proves to be more difficult. This also raised the question of the completeness of relation types in FrameNet, or at least the consistency of certain decisions that are made in grouping lexical predicates. As a perspective, we plan to evaluate our model predictions at various levels of confidence, to determine the best compromise between precision and recall when providing relation instance candidates for human validation. It would also be interesting to combine this approach with work on lexical expansion of FN, such as (Pavlick et al., 2015), or use an unsupervised method to

pre-filter candidate pairs, using looser lexical association resources, such as the Moby thesaurus (Ward, 1996).

An interesting perspective given the conclusions of the error analysis would be to try to predict new relations in a global way, using principles on the well-formedness of such ontological/lexical relationships, following the ontological principles given in (Ovchinnikova et al., 2010), or work on lexical taxonomy induction, see e.g. (Navigli et al., 2011).

As mentioned previously, FrameNet does provide a list of FE-to-FE relations for each of the connected frame pairs. In this paper, we have proposed ways to predict new frame-to-frame relations, and we plan to explore the possibilities of automatically linking frame-elements of the newly proposed frame-to-frame relations.

Acknowledgments

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada, August. Association for Computational Linguistics.

Steven Bird. 2006. Nltk: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics.

Aljoscha Burchardt, Anette Frank, and Manfred Pinkal. 2005. Building text meaning representations from contextually related frames - a case study. In *Proceedings of the Sixth International Workshop on Computational Semantics*, pages 12–. to appear.

Juliette Conrath, Stergos Afantenos, Nicholas Asher, and Philippe Muller. 2014. Unsupervised extraction of semantic relations using discourse cues. In *International Conference on Computational Linguistics - COLING 2014*, pages pp. 2184–2194. Association for Computational Linguistics.

Dipanjan Das and Noah A. Smith. 2012. Graph-based lexicon expansion with sparsity-inducing penalties. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 677–687, Montréal, Canada, June. Association for Computational Linguistics.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. *Comput. Linguist.*, 40(1):9–56, March.

Stefan Evert. 2005. *The statistics of word cooccurrences*. Ph.D. thesis, Stuttgart University.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Charles J. Fillmore, 1982. *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Graff, David, and Christopher Cieri. 2003. English gigaword ldc2003t05. Web Download. Philadelphia: Linguistic Data Consortium.

Yoko Hasegawa, Russell Lee-Goldman, Albert Kong, and Kimi Akita. 2011. Framenet as a resource for paraphrase research. *Constructions and Frames*, 3(1):104–127.

Graeme Hirst and David St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum (Fellbaum, 1998).

Paul Kingsbury and Martha Palmer. 2012. From treebank to propbank. In *Proceedings of LREC 2012*.

Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith, and Chris Dyer. 2015. Frame-semantic role labeling with heterogeneous annotations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 218–224, Beijing, China, July. Association for Computational Linguistics.

Claudia Leacock and Martin Chodorow. 1998. Combining local context and wordnet similarity for word sense identification. In Fellbaum (Fellbaum, 1998).

Ru Li, Juan Wu, Zhiqiang Wang, and Qinghua Chai. 2015. Implicit role linking on chinese discourse: Exploiting explicit roles and frame-to-frame relations. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1263–1271, Beijing, China, July. Association for Computational Linguistics.

Seyed Abolghasem Mirroshandel, Alexis Nasr, and Benoît Sagot. 2013. Enforcing subcategorization constraints in a parser using sub-parses recombining. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 239–247, Atlanta, Georgia, June. Association for Computational Linguistics.

Roberto Navigli, Paola Velardi, and Stefano Faralli. 2011. A graph-based algorithm for inducing lexical taxonomies from scratch. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pages 1872–1877, Barcelona, Spain.

Ekaterina Ovchinnikova, Laure Vieu, Ro Oltramari, Stefano Borgo, and Theodore Alex. 2010. Data-driven and ontological analysis of framenet for natural language reasoning. In *In Proc. of LREC10*.

Ellie Pavlick, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze, and Benjamin Van Durme. 2015. Framenet+: Fast paraphrastic tripling of framenet. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 408–413, Beijing, China, July. Association for Computational Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Marco Pennacchiotti and Michael Wirth. 2009. Measuring frame relatedness. In Alex Lascarides, Claire Gardent, and Joakim Nivre, editors, *EACL*, pages 657–665. The Association for Computer Linguistics.

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 457–465, Honolulu, Hawaii, October. Association for Computational Linguistics.

Simone Paolo Ponzetto and Michael Strube. 2006. Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 192–199, New York City, USA, June. Association for Computational Linguistics.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *In Proceedings of LREC*.

Josef Ruppenhofer, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2006. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 12–21, Prague, Czech Republic, June. Association for Computational Linguistics.

Lei Shi and Rada Mihalcea. 2005. Putting the pieces together: Combining FrameNet, VerbNet, and WordNet for robust semantic parsing. In *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico.

Mihai Surdeanu, Sanda Harabagiu, John Williams, and Paul Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 8–15, Sapporo, Japan, July. Association for Computational Linguistics.

Sara Tonelli and Daniele Pighin. 2009. New features for framenet - wordnet mapping. In *CoNLL'09: Thirteenth Conference on Computational Natural Language Learning*, Boulder, CO, USA, 06/2009. ACL, ACL.

G. Ward. 1996. Moby thesaurus. Moby project.

Zhibiao Wu and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138, Las Cruces, New Mexico, USA, June. Association for Computational Linguistics.