

An Active Learning Approach to the Classification of Non-Sentential Utterances

Paolo Dragone¹, Pierre Lison²

¹Sapienza University of Rome

²University of Oslo

dragone.paolo@gmail.com, plison@ifi.uio.no

Abstract

English. This paper addresses the problem of classification of non-sentential utterances (NSUs). NSUs are utterances that do not have a complete sentential form but convey a full clausal meaning given the dialogue context. We extend the approach of Fernández et al. (2007), which provide a taxonomy of NSUs and a small annotated corpus extracted from dialogue transcripts. This paper demonstrates how the combination of new linguistic features and active learning techniques can mitigate the scarcity of labelled data. The results show a significant improvement in the classification accuracy over the state-of-the-art.

Italiano. *Questo articolo affronta il problema della classificazione delle non-sentential utterances (NSUs). Le NSUs sono espressioni che, pur avendo una forma incompleta, esprimono un significato completo dato il contesto del dialogo. Estendiamo l'approccio di Fernández et al. (2007), il quale fornisce una tassonomia per NSUs ed un piccolo corpus estratto da transcript di dialoghi. Questo articolo dimostra come, tramite l'utilizzo di nuove feature linguistiche in combinazione con tecniche di active learning, si riesce ad attenuare la scarsità di dati annotati. I risultati mostrano un miglioramento significativo dell'accuratezza rispetto allo stato dell'arte.*

1 Introduction

In dialogue, utterances do not always take the form of complete, well-formed sentences with a subject, a verb and complements. Many utterances – often called *non-sentential utterances*, or NSUs for

short – are fragmentary and lack an overt predicate. Consider the following examples from the British National Corpus:

A: How do you actually feel about that?

B: **Not too happy.** [BNC: JK8 168-169]

A: They wouldn't do it, no.

B: **Why?** [BNC: H5H 202-203]

A: [...] then across from there to there.

B: **From side to side.** [BNC: HDH 377-378]

Despite their ubiquity, the semantic content of NSUs is often difficult to extract automatically. Non-sentential utterances are indeed intrinsically dependent on the dialogue context for their interpretation – for instance, the meaning of "why" in the example above is impossible to decipher without knowing what precedes it.

This paper describes a new approach to the classification of NSUs. The approach builds upon the work of Fernández et al. (2007), which present a corpus of NSUs along with a taxonomy and a classifier based on simple features. In particular, we show that the inclusion of new linguistic features and the use of active learning provide a modest but significant improvement in the classification accuracy compared to their approach.

The next section presents the corpus used in this work and its associated taxonomy of NSUs. Section 3 describes our classification approach (extracted features and learning algorithm). Section 4 finally presents the empirical results and their comparison with the baseline.

2 Background

Fernández et al. (2007) provide a taxonomy of NSUs based on 15 classes, reflecting both the form and pragmatic function fulfilled by the utterance.

The aforementioned paper also presents a small corpus of annotated NSUs extracted from dialogue transcripts of the British National Corpus

NSU Class	Example		Frequency
Plain Ack. (Ack)	A: ...	B: <i>mmh</i>	599
Short Answer (ShortAns)	A: <i>Who left?</i>	B: <i>Bo</i>	188
Affirmative Answer (AffAns)	A: <i>Did Bo leave?</i>	B: <i>Yes</i>	105
Repeated Ack. (RepAck)	A: <i>Did Bo leave?</i>	B: <i>Bo, hmm.</i>	86
Clarification Ellipsis (CE)	A: <i>Did Bo leave?</i>	B: <i>Bo?</i>	82
Rejection (Reject)	A: <i>Did Bo leave?</i>	B: <i>No.</i>	49
Factual Modifier (FactMod)	A: <i>Bo left.</i>	B: <i>Great!</i>	27
Repeated Aff. Ans. (RepAffAns)	A: <i>Did Bo leave?</i>	B: <i>Bo, yes.</i>	26
Helpful Rejection (HelpReject)	A: <i>Did Bo leave?</i>	B: <i>No, Max.</i>	24
Check Question (CheckQu)	A: <i>Bo isn't here. okay?</i>		22
Sluice	A: <i>Someone left.</i>	B: <i>Who?</i>	21
Filler	A: <i>Did Bo ...</i>	B: <i>leave?</i>	18
Bare Modifier Phrase (BareModPh)	A: <i>Max left.</i>	B: <i>Yesterday.</i>	15
Propositional Modifier (PropMod)	A: <i>Did Bo leave?</i>	B: <i>Maybe.</i>	11
Conjunct (Conj)	A: <i>Bo left.</i>	B: <i>And Max.</i>	10
Total			1283

Table 1: Taxonomy of NSUs with examples and frequencies in the corpus of Fernández et al. (2007).

(Burnard, 2000). Each instance of NSU is annotated with its corresponding class and its antecedent (which is often but not always the preceding utterance). Table 1 provides an overview of the taxonomy, along the frequency of each class in the corpus and prototypical examples taken from Ginzburg (2012). See also e.g. Schlangen (2003) for related NSU taxonomies. Due to space constraints, we do not provide here an exhaustive description of each class, which can be found in (Fernández, 2006; Fernández et al., 2007).

3 Approach

In addition to their corpus and taxonomy of NSUs, Fernández et al. (2007) also described a simple machine learning approach to determine the NSU class from simple features. Their approach will constitute the baseline for our experiments. We then show how to extend their feature set and rely on active learning to improve the classification.

3.1 Baseline

The feature set of Fernández et al. (2007) is composed of 9 features. Four features capture some key syntactic and lexical properties of the NSU itself, such as the presence of yes/no words or wh-words in the NSU. In addition, three features are extracted from the antecedent utterance, capturing properties such as the mood or the presence of a marker indicating whether the utterance is complete. Finally, two features encode similarity mea-

asures between the NSU and its antecedent, such as the number of repeated words and POS tag sequences common to the NSU and its antecedent.

The classification performance of our replicated classifier (see Table 2) are in line with the results presented in Fernández et al. (2007) – with the exception of the accuracy scores, which were not provided in the original article.

3.2 Extending the feature set

In order to improve the classification accuracy, we extended the baseline features described above with a set of 23 additional features, summing up to a total of 32 features:

- **POS-level features:** 7 features capturing shallow syntactic properties of the NSUs, such as the initial POS tags and the presence of pauses and unclear fragments.
- **Phrase-level features:** 7 features indicating the presence of specific syntactic structures in the NSU and the antecedent, for instance the type of clause-level tags (eg. S, SQ, SBAR) in the antecedent or the initial phrase-level tag (eg. ADVP, FRAG, NP) in the NSU.
- **Dependency features:** 2 features signaling the presence of certain dependency patterns in the antecedent, for example the occurrence of a *neg* dependency in the antecedent.
- **Turn-taking features:** one feature indicating

whether the NSU and its antecedent are uttered by the same speaker.

- **Similarity features:** 6 features measuring the parallelism between the NSU and its antecedent, such as the local (character-level) alignment based on Smith and Waterman (1981) and the longest common subsequence at the word- and POS-levels, using Needleman and Wunsch (1970).

The phrase-level and dependency features were extracted with the PCFG and Dependency Parsers (Klein and Manning, 2003; Chen and Manning, 2014) from the Stanford CoreNLP API.

3.3 Active learning

The objective of active learning (AL) (Settles, 2010) is to interactively query the user to annotate new data by selecting the most informative instances (that is, the ones that are most difficult to classify). Active learning is typically employed to cope with the scarcity of labelled data. In our case, the lack of sufficient training data is especially problematic due to the strong class imbalance between the NSU classes (as exemplified in Table 1). Furthermore, the most infrequent classes are often the most difficult ones to discriminate. Fortunately, the dialogue transcripts from the BNC also contain a large amount of unlabelled NSUs that can be extracted from the raw transcripts using simple heuristics (syntactic patterns to select utterances that are most likely non-sentential).

The active learning algorithm we employed in this work is a pool-based method with uncertainty sampling (Lewis and Catlett, 1994). The sampling relies on *entropy* (Shannon, 1948) as measure of uncertainty. Given a particular (unlabelled) instance with a vector of feature values \mathbf{f} , we use the existing classifier to derive the probability distribution $P(C = c_i|\mathbf{f})$ for each possible output class c_i . We can then determine the corresponding entropy of the class C :

$$H(C) = - \sum_i P(C = c_i|\mathbf{f}) \log P(C = c_i|\mathbf{f})$$

A high entropy indicates the “unpredictability” of the instance. The most informative instances to label are therefore the ones with high entropy. As argued in Settles (2010), entropy sampling is especially useful when there are more than two classes, as in our setting. We applied the JCLAL active

NSU Class	Instances
Helpful Rejection	21
Repeated Acknowledgment	17
Clarification Ellipsis	17
Acknowledgment	11
Propositional Modifier	9
Filler	9
Sluice	3
Repeated Affirmative Answer	3
Factual Modifier	3
Conjunct Fragment	3
Short Answer	2
Check Question	2

Table 5: Class frequencies of the 100 additional NSUs extracted via active learning.

learning library¹ to extract and annotate 100 new instances of NSUs, which were then added to the training data. The distribution of NSU classes for these instances is shown in Table 5.

4 Evaluation

We compared the classification results between the baseline and the new approach which includes the extended feature set and the additional data extracted via active learning. All the experiments were conducted using the Weka package (Hall et al., 2009). Table 2 presents the results using the J48 classifier, an implementation of the C4.5 algorithm for decision trees (Quinlan, 1993), while Table 3 presents the results using Weka’s SMO classifier, a type of SVM trained using sequential minimal optimization (Platt, 1998). In all experiments, we follow Fernández et al. (2007) and remove from the classification task the NSUs whose antecedents are not the preceding utterance, thus leaving a total of 1123 utterances.

All empirical results were computed with 10-fold cross validation over the full dataset. The active learning (AL) results refer to the classifiers trained after the inclusion of the 100 additional instances. The results show a significant improvement of the classification performance between the baseline and the final approach using the SVM and the data extracted via active learning. Using a paired *t*-test with a 95% confidence interval between the baseline and the final results, the improvement in accuracy is statistically significant with a *p*-value of 6.9×10^{-3} . The SVM does

¹cf. <http://sourceforge.net/projects/jclal>.

Experimental setting	Accuracy	Precision	Recall	F_1 -Score
Train-set (baseline feature set)	0.885	0.888	0.885	0.879
Train-set (extended feature set)	0.889	0.904	0.889	0.889
Train-set + AL (baseline feature set)	0.890	0.896	0.890	0.885
Train-set + AL (extended feature set)	0.896	0.914	0.896	0.897

Table 2: Accuracy, precision, recall and F_1 scores for each experiment, based on the J48 classifier.

Experimental setting	Accuracy	Precision	Recall	F_1 -Score
Train-set (baseline feature set)	0.881	0.884	0.881	0.875
Train-set (extended feature set)	0.899	0.904	0.899	0.896
Train-set + AL (baseline feature set)	0.883	0.893	0.883	0.880
Train-set + AL (extended feature set)	0.907	0.913	0.907	0.905

Table 3: Accuracy, precision, recall and F_1 scores for each experiment, based on the SMO classifier.

NSU Class	Baseline			Final approach		
	Precision	Recall	F_1 -Score	Precision	Recall	F_1 -Score
Plain Acknowledgment	0.97	0.97	0.97	0.97	0.98	0.97
Affirmative Answer	0.89	0.84	0.86	0.81	0.90	0.85
Bare Modifier Phrase	0.63	0.65	0.62	0.77	0.75	0.75
Clarification Ellipsis	0.87	0.89	0.87	0.88	0.92	0.89
Check Question	0.85	0.90	0.87	1.00	1.00	1.00
Conjunct Fragment	0.80	0.80	0.80	1.00	1.00	1.00
Factual Modifier	1.00	1.00	1.00	1.00	1.00	1.00
Filler	0.77	0.70	0.71	0.82	0.83	0.78
Helpful Rejection	0.13	0.14	0.14	0.31	0.43	0.33
Propositional Modifier	0.92	0.97	0.93	0.92	1.00	0.95
Rejection	0.76	0.95	0.83	0.90	0.90	0.89
Repeated Ack.	0.74	0.75	0.70	0.77	0.77	0.77
Repeated Aff. Ans.	0.67	0.71	0.68	0.72	0.55	0.58
Short Answer	0.86	0.80	0.81	0.92	0.86	0.89
Sluice	0.67	0.77	0.71	0.80	0.84	0.81

Table 4: Precision, recall and F_1 score per class between the baseline (initial feature set and J48 classifier) and the final approach (extended feature set with active learning and SMO classifier).

not perform particularly well on the baseline features but scales better than the J48 classifier after the inclusion of the additional features. Overall, the results demonstrate that the classification can be improved using a modest amount of additional training data combined with an extended feature set. However, we can observe from Table 4 that some NSU classes remain difficult to classify. Distinguishing between e.g. *Helpful Rejections* and *Short Answers* indeed requires a deeper semantic analysis of the NSUs and their antecedents than cannot be captured by morpho-syntactic features alone. Designing appropriate semantic features for this classification task constitutes an interesting question for future work.

5 Conclusion

This paper presented the results of an experiment in the classification of non-sentential utterances,

extending the work of Fernández et al. (2007). The approach relied on an extended feature set and active learning techniques to address the scarcity of labelled data and the class imbalance. The evaluation results demonstrated a significant improvement in the classification accuracy.

The presented results also highlight the need for a larger annotated corpus of NSUs. In our view, the development of such a corpus, including new dialogue domains and a broader range of conversational phenomena, could contribute to a better understanding of NSUs and their interpretation.

Furthermore, the classification of NSUs according to their type only constitutes the first step in their semantic interpretation. Dragone and Lison (2015) focuses on integrating the NSU classification outputs for natural language understanding of conversational data, building upon Ginzburg (2012)’s formal theory of conversation.

References

- L. Burnard. 2000. Reference guide for the british national corpus (world edition).
- D. Chen and C. D. Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, volume 1, pages 740–750.
- P. Dragone and P. Lison. 2015. Non-sentential utterances in dialogue: Experiments in classification and interpretation. In *Proceedings of the 19th Workshop on the Semantics and Pragmatics of Dialogue*, page 170.
- R. Fernández, J. Ginzburg, and S. Lappin. 2007. Classifying non-sentential utterances in dialogue: A machine learning approach. *Computational Linguistics*, 33(3):397–427.
- R. Fernández. 2006. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. Ph.D. thesis, King’s College London.
- J. Ginzburg. 2012. *The Interactive Stance*. Oxford University Press.
- M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November.
- D. Klein and C. D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 423–430. Association for Computational Linguistics.
- D. D. Lewis and J. Catlett. 1994. Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the eleventh international conference on machine learning*, pages 148–156.
- S. B. Needleman and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453.
- J. Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14, Microsoft Research, April.
- R. J. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc.
- D. Schlangen. 2003. *A coherence-based approach to the interpretation of non-sentential utterances in dialogue*. Ph.D. thesis, University of Edinburgh. College of Science and Engineering. School of Informatics.
- B. Settles. 2010. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11.
- C. E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(3):379–423, July.
- T. F. Smith and M. S. Waterman. 1981. Identification of common molecular subsequences. *Journal of molecular biology*, 147(1):195–197.