Slang Generation as Categorization

Zhewei Sun (zheweisun@cs.toronto,edu)
Department of Computer Science
University of Toronto

Richard Zemel (zemel@cs.toronto.edu)
Department of Computer Science
University of Toronto, Vector Institute

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science, Cognitive Science
University of Toronto

Acknowledgements
Thanks to Suzanne Stevenson, Barend Beekhuizen, and Renato Ferreira Pinto Junior for constructive comments

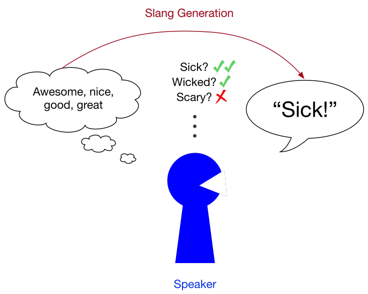
Introduction

- Slang is ubiquitous in human language yet its cognitive bases are not well understood.
- Previous research has characterized slang as a social phenomenon (Labov, 1972, 2006), and in the context of cognitive science, focused on comprehension of metaphors (Kao, Bergen, & Goodman, 2014).
- We examine whether it is feasible to forecast lexical choices for novel usages of slang with linguistic resources alone, using categorization models inspired by two key ideas from historical word sense extension: semantic chaining (Ramiro et al., 2018) and parallel semantic change (Lehrer, 1985, Xu & Kemp, 2015).

- **Task**: Given a slang sense such as "awesome, nice" as illustrated in Figure 1, we wish to predict the word choice made by the speaker among possible alternative words within our vocabulary. In the illustrated case, the target word *sick* might be chosen if its existing senses relate to the novel slang sense, and words similar to the target word *sick* such as *wicked* might also have a good chance of being chosen.

- We found that categorization models capture substantial predictability in the emergence of novel slang usages. Adding collaborative filtering further enhances both the accuracy and generalizability of the models.

Figure 1: Illustration of the slang generation problem



Computational Formulation

- Given a slang sense S, find a word w in our lexicon that best captures its meaning by estimating the distribution:

$$P(w|S) \propto P(S|w)P(w)$$

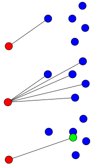
- **Categorization** approach: Compute the likelihood of S given existing definitions $E_w = \{W_1, W_2, \cdots .W_n\}$ of word w:

$$P(w|S) \propto P(S|w) = P(S|E_w) = P(S|W_1, W_2, \cdots, W_n)$$

- **Collaborative Filtering**: On top of estimating the maximum likelihood, we also integrate over a set of neighboring candidate words:

$$P(w|S) \propto \sum_{w' \in \mathcal{L}(w)} P(w, w'|S) = \sum_{w' \in \mathcal{L}(w)} P(w|w')\underbrace{P(w'|S)}_{\text{same as above}}$$

- Similarity between the slang and existing definitions are computed using one of the categorization models:
  - 1NN (Ramiro et al., 2018):
    $Sim(S, E_w) = \max_{W_i \in E_w} Sim(S, W_i)$
  - Exemplar (Nosofsky, 1986):
    $Sim(S, E_w) = \sum_{W_i \in E_w} Sim(S, W_i)$
  - Prototype (Rosch, 1975):
    $Sim(S, E_w) = Sim(S, E_w^{prototype})$
    $E_w^{prototype} = \mathbb{E}_{W_i \in E_w} W_i$



- Similarity between pairs of definitions and words are computed as follows with two free h parameters:

$$Sim(S, W_i) = \exp\left(-\frac{d(S, W_i)^2}{h_s}\right)$$

$$P(w|w') \propto sim(w, w') = \exp\left(-\frac{d(w, w')^2}{h_w}\right)$$

Materials and Methods

- We leverage the following resources in our experiments:
  - **Online Slang Dictionary** (OSD; *onlineslangdictionary.com*) for slang lexical entries and their definitions.
  - **WordNet** for existing definitions of words.
  - **fastText** for pretrained word embeddings.

- We exclude acronyms and remove possibly conventionalized slang definitions that have significant word overlap in the set of content words.
- Our resulting dataset contains 4,256 slang definitions from 2,128 distinct words. These slang definitions are split into a 90% training set for parameter learning and a 10% test set for model evaluation.
- Definition sentences are represented by summing fastText embeddings of the corresponding set of contents words.
- The free kernel width parameters are learned by optimizing the negative log-likelihood of the posterior on the training definitions using quasi-newton methods.

Results

Figure 2: Summary of results in terms of both AUC and Expect Rank measures.

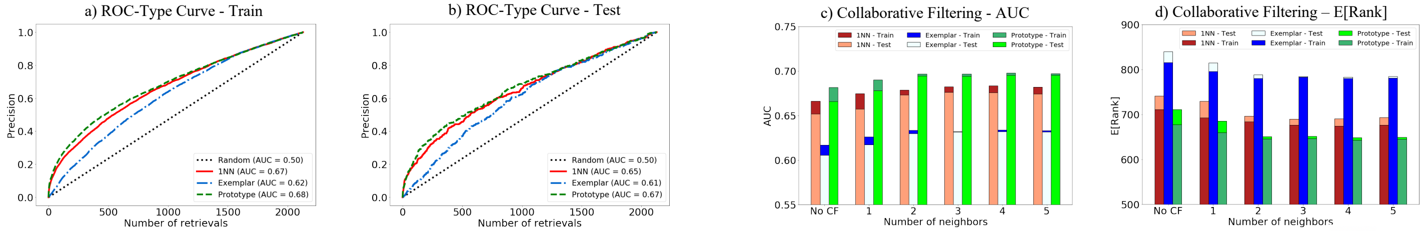


Table 1: Example of model success.

Ground truth target word [w]:	*wicked*
Slang sense in OSD [s]:	impressive.
Corresponding WordNet senses [E]:	(1) morally bad in principle or practice; (2) having committed unrighteous acts; (3) intensely or extremely bad or unpleasant in degree or quality; (4) naughtily or annoyingly playful; (5) highly offensive; arousing aversion or disgust.
Model expected rankings [E(Rank)]:	(1NN): 93/2128; (Exemplar): 369/2128; (Prototype): 33/2128
Top ranked words:	(1NN): *bonzer, spot, point, tall, grand*; (Exemplar): *broken, play, cut, point, heavy*; (Prototype): *bonzer, good, tall, grand, hot*

Table 2: Example that illustrate how collaborative filtering helps predicting slang word choice.

Ground truth target word [w]:	*scary*
Slang sense in OSD [s]:	ugly, weird.
Corresponding WordNet senses [E]:	provoking fear terror.
5 neighboring words used in collaborative filtering [$\mathcal{L}(w)$]:	*freaky, crazy, nightmare, awesome, stupid*

- We assess our models by ranking all candidate words according to the posterior distribution $P(w|S)$ from the categorizations models. We then compute the Area-Under-Curve (AUC) statistics for the Receiver-Operator Type (ROC) curve (Figure 2a, 2b) and the expected rank of the corresponding ground-truth words.
- All three types of categorization models perform better than chance, while 1NN and Prototype both perform better than the Exemplar model.
- Collaboratively filtered models achieve better AUC and expected rank on both the training set and testing set compared to their respective basic categorization models (Figure 2c, 2d).

Conclusion

Our categorization framework was able to capture substantial predictability in slang generation without explicitly modeling external social variables, where predictability can be further enhanced by collaboratively filtering on semantically similar words.

References

Kao, J. T., Bergen, L., & Goodman, N. D. (2014). Formalizing the pragmatics of metaphor understanding. In Proceedings of the 36th annual conference of the cognitive science society (pp. 719–724). Cognitive Science Society; Labov, W. (1972). Language in the inner city: Studies in the black english vernacular. University of Pennsylvania Press; Labov, W. (2006). The social stratification of english in new york city. Cambridge University Press; Lehrer, A. (1985). The influence of semantic fields on semantic change. Historical Semantics: Historical Word Formation, 29, 283–296; Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General, 115, 39–57; Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. Proceedings of the National Academy of Sciences, 115, 2323–2328; Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192–233; Xu, Y., & Kemp, C. (2015). A computational evaluation of two laws of semantic change. In Proceedings of the 37th annual conference of the cognitive science society (pp. 2703–2708). Cognitive Science Society.