Near-Zero-Shot-Classification with a Little Help from WordNet

Anton Alekseev[1], Elena Tutubalina[2,4], Sejeong Kwon[3], Sergey Nikolenko[1]

[1]Steklov Mathematical Institute at St. Petersburg, Russia
[2]Kazan Federal University, Russia
[3]Samsung Research, Korea
[4]Sber AI, Russia

16 December 2021

Overview

- **Goal**: explore the constructive side of online reviews:
  - advice
  - tips
  - requests
  - suggestions
- **Data**: reviews about software and hotels
- **Methods**: entailment-based zero-shot approaches in a label-fully-unseen fashion

Data: SemEval 2019 Task 9

Two domains:

- Software forums (UserVoice.com)
  - many suggestions in the form of requests
  - highly technical vocabulary

- Hotel reviews (TripAdvisor)

Task/Domain		Suggestions		Non-Suggestions	
A	Software development forums (Uservoice)	Training	2085	Training	6415
		Validation	296	Validation	296
		Testing	87	Testing	746
B	Hotel reviews (TripAdvisor)	Training	2085	Training	6415
		Validation	404	Validation	404
		Testing	348	Testing	476

Table 1. Dataset statistics. The Train/validation/test split is shown for suggestions and non-suggestions for the two subtasks of [24].

Data

	Suggestion	Non-suggestion
SDE forums	The proposal is to add something like: // Something happened update your UI or run your business logic	I write a lot support ticket on this, but no one really cares on this issue.
Hotel reviews	For a lovely breakfast, turn left out of the front entrance - on the next corner is a cafe with fresh baked breads and cooked meals.	A great choice!!

Table 2. Samples from the SemEval2019-Task9 suggestion mining dataset: software development (SDE) forums and hotel reviews.

Subtask A: train/dev/test – software forums; **single-domain**

Subtask B: development and test sets – hotel reviews; **cross-domain**

Original Work: MNLI Models for Text Classification

MNLI model (e.g. mBART)
https://huggingface.co/facebook/bart-large-mnli

premise	hypothesis
text	**'This text is about [LABEL]'**

0.2	entailment
0.1	contradiction
0.7	neutral

Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach

Wenpeng Yin, Jamaal Hay, Dan Roth
Cognitive Computation Group
Department of Computer and Information Science, University of Pennsylvania
{wenpeng,jamaalh,danroth}@seas.upenn.edu

Abstract

Zero-shot text classification (0SHOT-TC) is a challenging NLU problem to which little at-

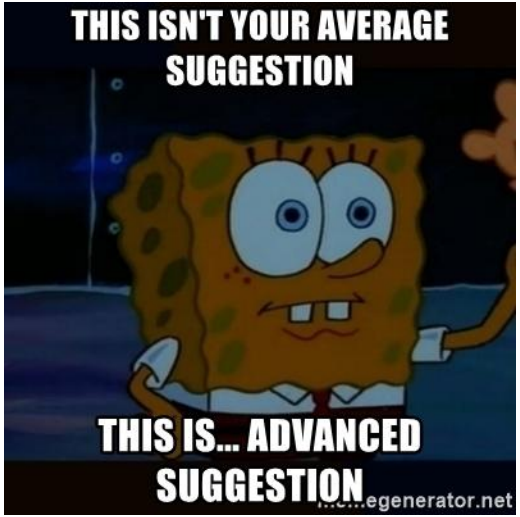
Approach 1: “This is a suggestion.”

- We use the following statements as hypotheses:
  - “This text is a suggestion.”
  - “This text is not a suggestion.”
  - “This text suggests/is suggesting.”

- Results were arguably unsatisfactory…
  - Suggestion can be a plea, a question, a request etc.
  - Non-suggestions are even more diverse: questions, comments, jokes, complaints and so on.

Approach 2: WordNet definitions

- We use definitions of “suggestion” from WordNet:
  - *suggestion.n.01* (“This text is an idea that is suggested”)
  - *suggestion.n.02* (“This text is a proposal offered for acceptance or rejection”)
  - *suggestion.n.04* (“This text is persuasion formulated as a suggestion”).
- Some definitions were discarded as irrelevant
  - *suggestion.n.05*: the sequential mental process in which one thought leads to another by association
- Results were arguably unsatisfactory again…



Approach 3: Near-Zero-Shot Learning

- Zero-shot learning: no target task data used, no gold labels seen
- Few-shot learning: using a small number of annotated target task data
- ≠ ‘Near-Zero-Shot Learning’: tuning the zero-shot predictions selection mechanism on development set

Approach 3: Near-Zero-Shot Learning

- There are a wide variety of possible “message types” about suggestions
- We have formulated the labels as “This text is a [LEMMA]” -- now we need to select a set of labels
- There is the WordNet synset *message.n.02*: “what a communication that is about something is about”
- We, explore the hyperonyms of *message.n.02* in WordNet:

✓	*offer.n.02*	offer	something offered (as a proposal or bid)
	opinion.n.02	opinion	a message expressing a belief about something...
✓	*promotion.n.01*	promotion	a message issued in behalf of some product or...
✓	*proposal.n.01*	proposal	something proposed (such as a plan or...
	refusal.n.02	refusal	a message refusing to accept something that...
✓	*reminder.n.01*	reminder	a message that helps you remember something
✓	*request.n.01*	request	a formal message requesting something that is...
	respects.n.01	respects	(often used with ‘pay’) a formal expression of...

Results

Premise	Task	Dev. set		Test set	
		F1	Acc.	F1	Acc.
"This text is [not] a suggestion." (A1)	A	0.6727	0.5152	0.1961	0.1536
	B	0.6616	0.5	0.5806	0.4163
"This text is [not] suggesting."(A1)	A	0.6712	0.5118	0.1898	0.1393
	B	0.6617	0.4988	0.5840	0.4175
3 definitions VS "This text is **not** a suggestion." (A2)	A	0.6689	0.5051	0.1925	0.1237
	B	0.6656	0.5025	0.5876	0.4175
"The best subset" (A3)	A	0.7517	0.7568	0.4479	0.8283
	B	0.4635	0.6361	0.4841	0.6699

Table 4. Results of Approaches 1 & 2 and the best label subset from Approach 3 (Subtask A).

Big difference with the results in the supervised setting:

A: OleNet (Baidu; ensemble): 0.7812 (F1) [test set]

B: OleNet (Baidu; ensemble): 0.8579 (F1) [test set]

[Potamias et al.] (learned rules): 0.858

Random uniform sampling:

A: 0.1734 (F1)

B: 0.4566 (F1)

Results

Size	Labels subset	F1	Accuracy
4	offer,proposal,reminder,request	0.5963	0.6845
	guidance,offer,proposal,reminder	0.5833	0.6723
	offer,promotion,proposal,reminder	0.5804	0.6772
5	guidance,offer,proposal,reminder,request	0.6030	0.6820
	offer,promotion,proposal,reminder,request	0.6006	0.6869
	offer,proposal,reminder,request,submission	0.5997	0.6857
6	guidance,offer,promotion,proposal,reminder,request	0.6073	0.6845
	guidance,offer,proposal,reminder,request,submission	0.6063	0.6833
	direction,guidance,offer,proposal,reminder,request	0.6054	0.6820
7	guidance,offer,promotion,proposal,reminder,request,submission	0.6105	0.6857
	direction,guidance,offer,promotion,proposal,reminder,request	0.6096	0.6845
	direction,guidance,offer,proposal,reminder,request,submission	0.6087	0.6833
8	(all 8 labels)	**0.6129**	0.6857

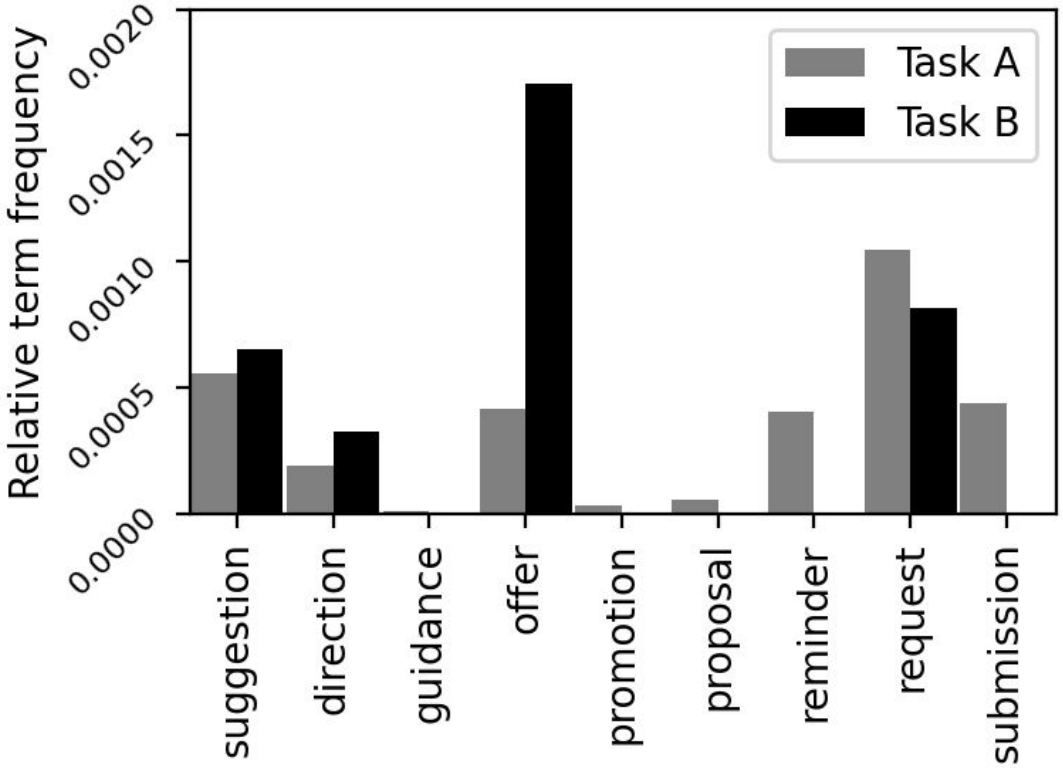
Table 7. Test set, Subtask B. Top-3 label subsets results for each subset size from 4 to 8 selected by F1 measure.

Best subset on the dev set

Random uniform sampling: 0.4566 (F1)

“The best subset” (A3, dev set from Task A): 0.4841 (F1)

Results


Relative term frequency
0.0000
0.0005
0.0010
0.0015
0.0020
Task A
Task B
suggestion
direction
guidance
offer
promotion
proposal
reminder
request
submission

Summary

- Prediction quality is far from that in the supervised setting
- Several approaches to *label-fully-unseen zero-shot* suggestion mining
- Word ‘message’ hyponyms-based method outperforms direct labeling

  “This text is [not] a suggestion”
- Best-performing subset of the hyponyms is domain-dependent
- Finding the best subset of hyponyms is an interesting direction for further work

https://arxiv.org/pdf/2111.12956.pdf