Few-shot text classification

With pre-trained word embeddings and a human in the loop

Problem

You have thousands of uncategorized pieces of content. You need categorized content in order to allow users to filter it.

Our Solution

Using our UI, manually label just a few pieces of content (even just one per category!) and get accurately predicted categories for the rest.

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

Few-Shot Learning: Learning from just a few labeled examples.

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We make the human do the “few shots”.

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

Few-Shot Learning: Learning from just a few labeled examples.

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We make the human do the “few shots”.

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

We don’t have big data

Few-Shot Learning: Learning from just a few labeled examples.

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We make the human do the “few shots”.

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

We don't have big data

Few-Shot Learning: Learning from just a few labeled examples.

We often don't have labeled data

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We make the human do the "few shots".

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

We don't have big data

Few-Shot Learning: Learning from just a few labeled examples.

We often don't have labeled data

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We're software engineers / UX designers

We make the human do the "few shots".

Transfer Learning: Taking the learnings gleaned from one task and applying them to another.

We don’t have big data

Few-Shot Learning: Learning from just a few labeled examples.

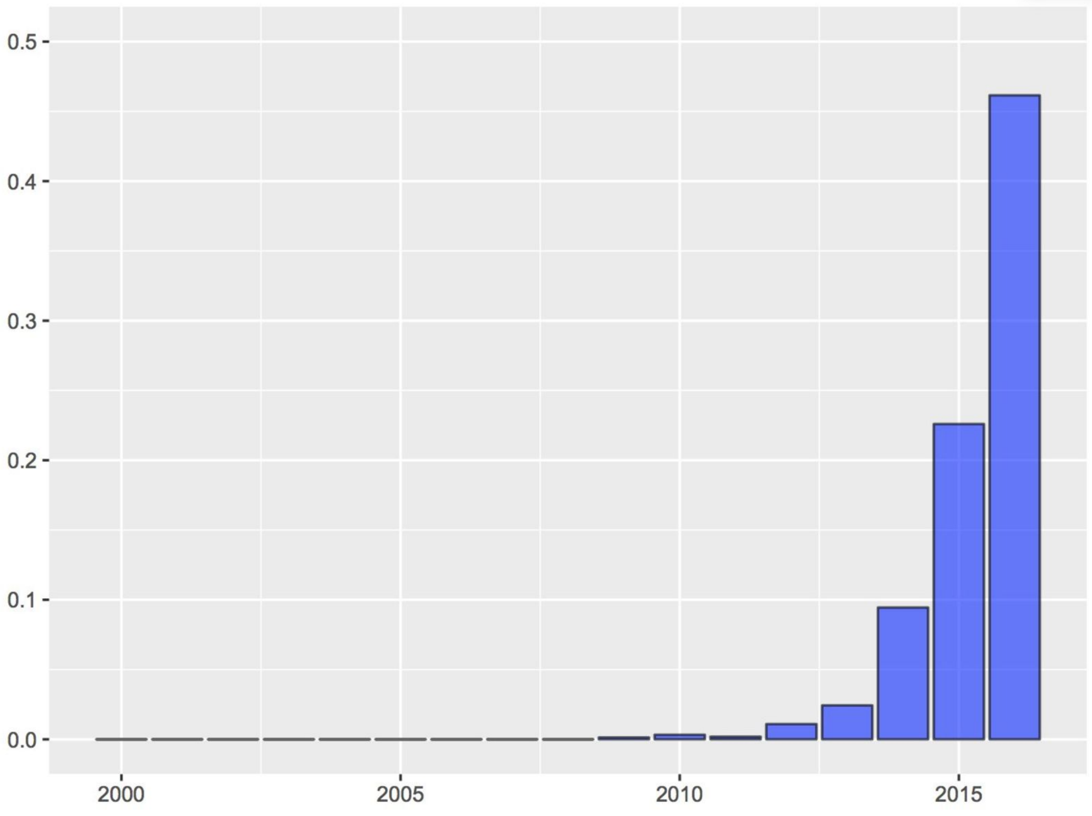
We often don’t have labeled data

Human-in-the-Loop Machine Learning: getting a human to help the machine learn.

We’re software engineers / UX designers

We make the human do the “few shots”.

OMG Word Embeddings!



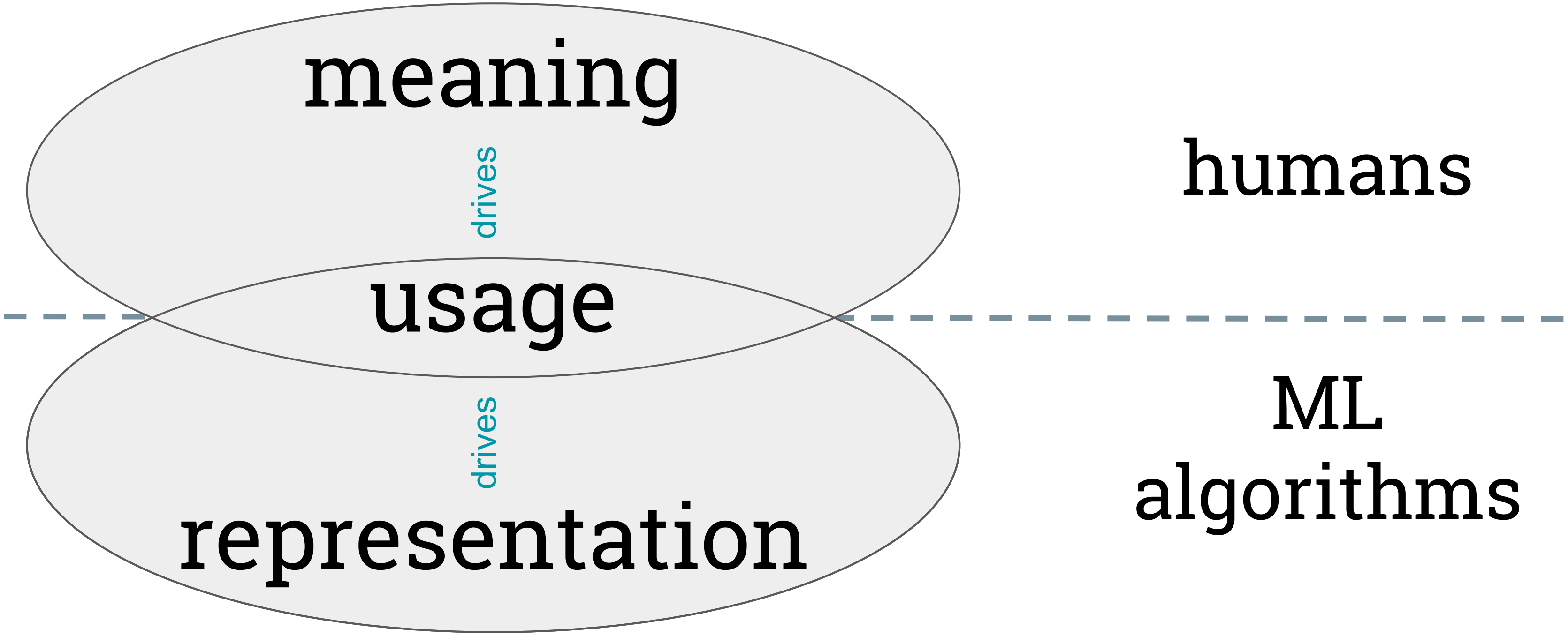
Fraction of papers in ACL, EMNLP, NAACL, EACL, CoNLL, TACL and CL mentioning "word embedding"

(data from ACL Anthology Network)

Source: David Bamman (@dbamman)

Word embeddings (aka word vectors) are learned numeric representations of words that capture the semantic relationships between them.

king - man + woman = queen


meaning
drives
usage
drives
representation
humans
ML
algorithms

From word vectors to document vectors...

Tried doc2vec, “Smooth Inverse Frequency” and other methods

Went with simple weighted average of the GloVe vectors for all words contained in the document (excluding stop words)

Document representation is completely independent of the batch it belongs to

The approach in a nutshell...

- Convert all documents to vectors
- Present the user with some documents in the UI and ask them to manually label at least one per category
- Use these as “category representatives”
- Compare all remaining docs to the representatives and for each one “predict” the category
- Score the “confidence” in each prediction and get the user to review the low confidence documents.

How accurate can this be?

Results on public labeled data sets

- Newsgroups: 20000 messages taken from 20 newsgroups on subjects like cars, religion, guns and baseball

- DBPedia: contains the first paragraph of the wikipedia page for ~0.5M entities in 15 categories, e.g. Animal, Film, Plant, Company, etc

- BBC: 2225 documents from the BBC news website corresponding to stories in five topical areas from 2004-2005 (business, entertainment, politics, sport, tech)

“Brute-force” one-shot accuracy results

Categories	No. of documents	Max accuracy
Animal,Film	8947	0.9984
Animal,Company	9220	0.9947
Animal,Film,Company	13867	0.9815
mideast,electronics	1063	0.9708
guns,hardware	1078	0.9703
autos,baseball	1046	0.9684
christian,guns	1099	0.9398
med,electronics	1112	0.9387
atheism,space	1004	0.9222
Animal,Plant	8730	0.9048
autos,baseball,space	1605	0.8939
baseball,hockey	1069	0.8772
politics,religion	774	0.8329

“Brute-force” one-shot accuracy results

Categories	No. of documents	Max accuracy
Animal,Film	8947	0.9984
Animal,Company	9220	0.9947
Animal,Film,Company	13867	0.9815
mideast,electronics	1063	0.9708
guns,hardware	1078	0.9703
autos,baseball	1046	0.9684
christian,guns	1099	0.9398
med,electronics	1112	0.9387
atheism,space	1004	0.9222
Animal,Plant	8730	0.9048
autos,baseball,space	1605	0.8939
baseball,hockey	1069	0.8772
politics,religion	774	0.8329

“Brute-force” one-shot accuracy results

Categories	No. of documents	Max accuracy
Animal,Film	8947	0.9984
Animal,Company	9220	0.9947
mideast,electronics	1063	0.9708
guns,hardware	1078	0.9703
christian,guns	1099	0.9398
med,electronics	1112	0.9387
atheism,space	1004	0.9222
[illegible]	8730	0.9048
autos,baseball,space	1605	0.8939

Films misclassified as animals:

“**Beasts of Prey** (육식동물 - Yukshik dongmul) also known as Carnivore and Carnivorous Animals is a 1985 South Korean film directed by Kim Ki-young.”

“**Brighty of the Grand Canyon** is a 1967 film based on a 1953 children's novel of the same name by Marguerite Henry a fictionalized account of a real-life burro named Brighty who lived in the Grand Canyon of the Colorado River from about 1892-1922.”

Animals misclassified as films:

“**California Memory** (Chinese: 加州萬里) is an American-bred Hong Kong based Thoroughbred racehorse. He was one of the nominees of 2010-2011 Hong Kong Horse of the Year.”

“**Military Move** (Chinese:軍事行動) is a New Zealand thoroughbred racehorse. On March 6 2010 he won the 135th running of the New Zealand Derby. The gelding is trained by Shaune Ritchie who was the strapper for the hugely successful and popular 1985 Derby winner Bonecrusher. Bonecrusher was trained by Ritchie's father Frank.”

“Brute-force” one-shot accuracy results

Categories	No. of documents	Max accuracy
Animal,Film	8947	0.9984
Animal,Company	9220	0.9947
Animal,Film,Company	13867	0.9815
mideast,electronics	1063	0.9708
guns,hardware	1078	0.9703
autos,baseball	1046	0.9684
christian,guns	1099	0.9398
med,electronics	1112	0.9387
atheism,space	1004	0.9222
Animal,Plant	8730	0.9048
autos,baseball,space	1605	0.8939
baseball,hockey	1069	0.8772
politics,religion	774	0.8329

How do we surface the likely representatives?

- Use **Latent Dirichlet Allocation (LDA)** to find the most "topic-y" docs!

- Present documents to the user ordered according to the LDA results

Results with LDA

LDA N is the maximum accuracy obtained trying all combinations of the top N documents surfaced by LDA.

Categories	No. of categories	No. of documents	Max accuracy	LDA 12	LDA 24
Animal,Film	2	8947	0.9984	0.9875908328675238	0.9928451648965902
Animal,Company	2	9220	0.9947	0.9847038403124322	0.9909958776307225
Animal,Film,Company	3	13867	0.9815	0.9557847663012118	0.9609780727062897
mideast,electronics	2	1063	0.9708	0.9641847313854854	0.9641847313854854
guns,hardware	2	1078	0.9703	0.9591078066914498	0.9591078066914498
autos,baseball	2	1046	0.9684	0.904214559387	0.944444444444
christian,guns	2	1099	0.9398	0.9161349134001823	0.9161349134001823
med,electronics	2	1112	0.9387	0.9027027027027027	0.9117117117117117
atheism,space	2	1004	0.9222	0.844311377245509	0.8702594810379242
Animal,Plant	2	8730	0.9048	--	0.7102428964252979
autos,baseball,space	3	1605	0.8939	0.7715355805243446	0.8508114856429463
baseball,hockey	2	1069	0.8772	0.7722586691658857	0.7722586691658857
politics,religion	2	774	0.8329	0.7927461139896373	0.7927461139896373

Results with LDA

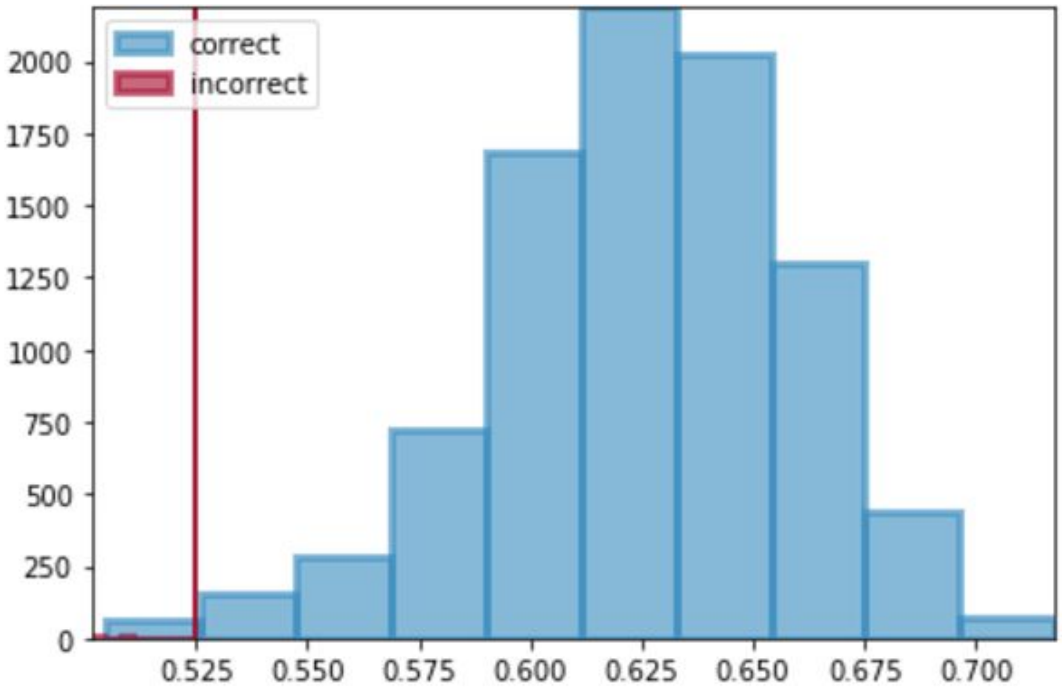
LDA N is the maximum accuracy obtained trying all combinations of the top N documents surfaced by LDA.

Categories	No. of categories	No. of documents	Max accuracy	LDA 12	LDA 24
Animal,Film	2	8947	0.9984	0.9875908328675238	0.9928451648965902
Animal,Company	2	9220	0.9947	0.9847038403124322	0.9909958776307225
Animal,Film,Company	3	13867	0.9815	0.9557847663012118	0.9609780727062897
mideast,electronics	2	1063	0.9708	0.9641847313854854	0.9641847313854854
guns,hardware	2	1078	0.9703	0.9591078066914498	0.9591078066914498
autos,baseball	2	1046	0.9684	0.904214559387	0.944444444444
christian,guns	2	1099	0.9398	0.9161349134001823	0.9161349134001823
med,electronics	2	1112	0.9387	0.9027027027027027	0.9117117117117117
atheism,space	2	1004	0.9222	0.844311377245509	0.8702594810379242
Animal,Plant	2	8730	0.9048	--	0.7102428964252979
autos,baseball,space	3	1605	0.8939	0.7715355805243446	0.8508114856429463
baseball,hockey	2	1069	0.8772	0.7722586691658857	0.7722586691658857
politics,religion	2	774	0.8329	0.7927461139896373	0.7927461139896373

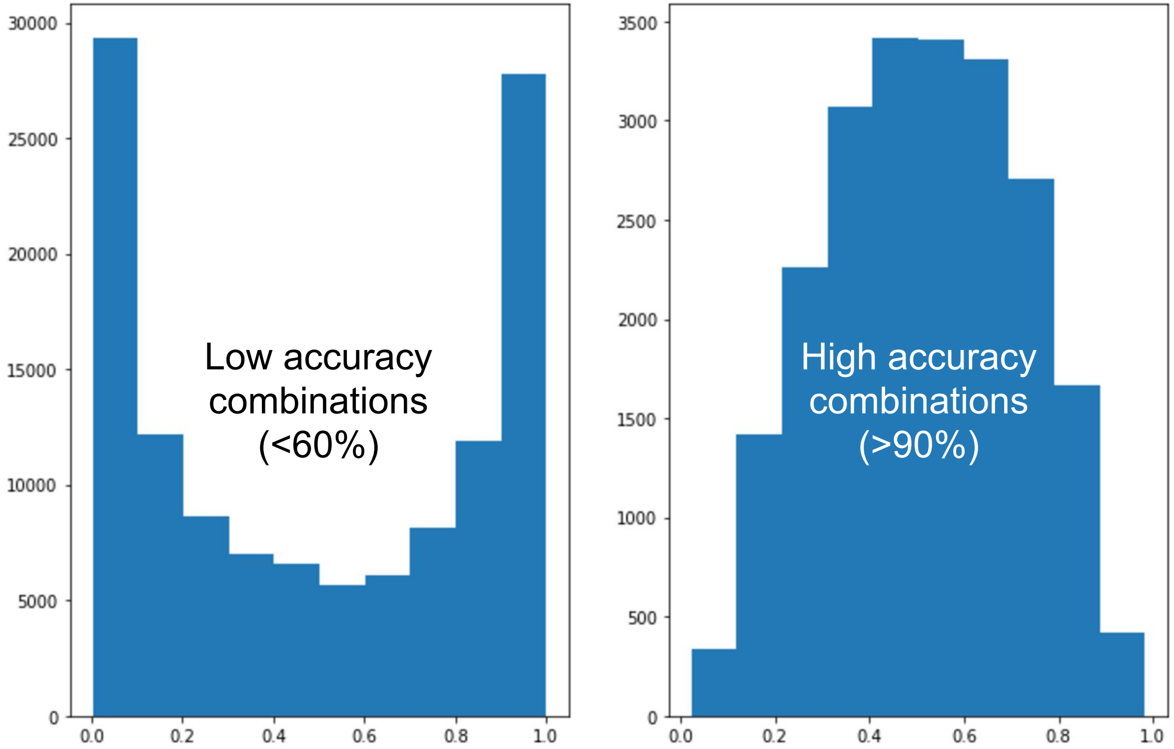
Categories	No. of categories	No. of documents	LDA 12	LDA 24
Village,Film	2	9307	0.9980655561526062	0.9984954325631381
MeanOfTransportation,NaturalPlace	2	8619	0.9886271324126726	0.990019728443774
Village,Company	2	9580	0.9886197536020046	0.9886197536020046
Building,NaturalPlace	2	8595	0.9757942511346445	0.9771907366461073
Album,Film	2	9388	0.9711272107393991	0.9745365437886213
Film,Company	2	9567	0.9643491897543126	0.9730266596968112
autos,hockey	2	1075	0.9412861136999068	0.9524697110904008
Artist,Athlete	2	9657	0.9481097876747799	0.9481097876747799
Film,WrittenWork	2	9491	0.9338181051744124	0.9454104752871746
Building,EducationalInstitution	2	9864	0.9337862502534983	0.9337862502534983
med,religion	2	904	0.8403547671840355	0.8802660753880266
space,med	2	1124	0.7923351158645277	0.8787878787878788
autos,guns	2	1044	0.8253358925143954	0.8656429942418427
business,entertainment,politics,sport,tech	5	2225	--	0.8639639639639639
autos,electronics	2	1073	0.811391223155929	0.8608776844070962
Animal,Film,Company,Village	4	18527	--	0.855800896183
religion,mideast	2	855	0.7620164126611958	0.7936694021101993
christian,atheism	2	1026	0.634765625	0.6630859375
atheism,religion	2	784	0.6202046035805626	0.6202046035805626
Village,Animal	2	8960	--	--

Categories	No. of categories	No. of documents	LDA 12	LDA 24
Village,Film	2	9307	0.9980655561526062	0.9984954325631381
MeanOfTransportation,NaturalPlace	2	8619	0.9886271324126726	0.990019728443774
Village,Company	2	9580	0.9886197536020046	0.9886197536020046
Building,NaturalPlace	2	8595	0.9757942511346445	0.9771907366461073
Album,Film	2	9388	0.9711272107393991	0.9745365437886213
Film,Company	2	9567	0.9643491897543126	0.9730266596968112
autos,hockey	2	1075	0.9412861136999068	0.9524697110904008
Artist,Athlete	2	9657	0.9481097876747799	0.9481097876747799
Film,WrittenWork	2	9491	0.9338181051744124	0.9454104752871746
Building,EducationalInstitution	2	9864	0.9337862502534983	0.9337862502534983
med,religion	2	904	0.8403547671840355	0.8802660753880266
space,med	2	1124	0.7923351158645277	0.8787878787878788
autos,guns	2	1044	0.8253358925143954	0.8656429942418427
business,entertainment,politics,sport,tech	5	2225	--	0.8639639639639639
autos,electronics	2	1073	0.811391223155929	0.8608776844070962
Animal,Film,Company,Village	4	18527	--	0.855800896183
religion,mideast	2	855	0.7620164126611958	0.7936694021101993
christian,atheism	2	1026	0.634765625	0.6630859375
atheism,religion	2	784	0.6202046035805626	0.6202046035805626
Village,Animal	2	8960	--	--

Confidence in predictions



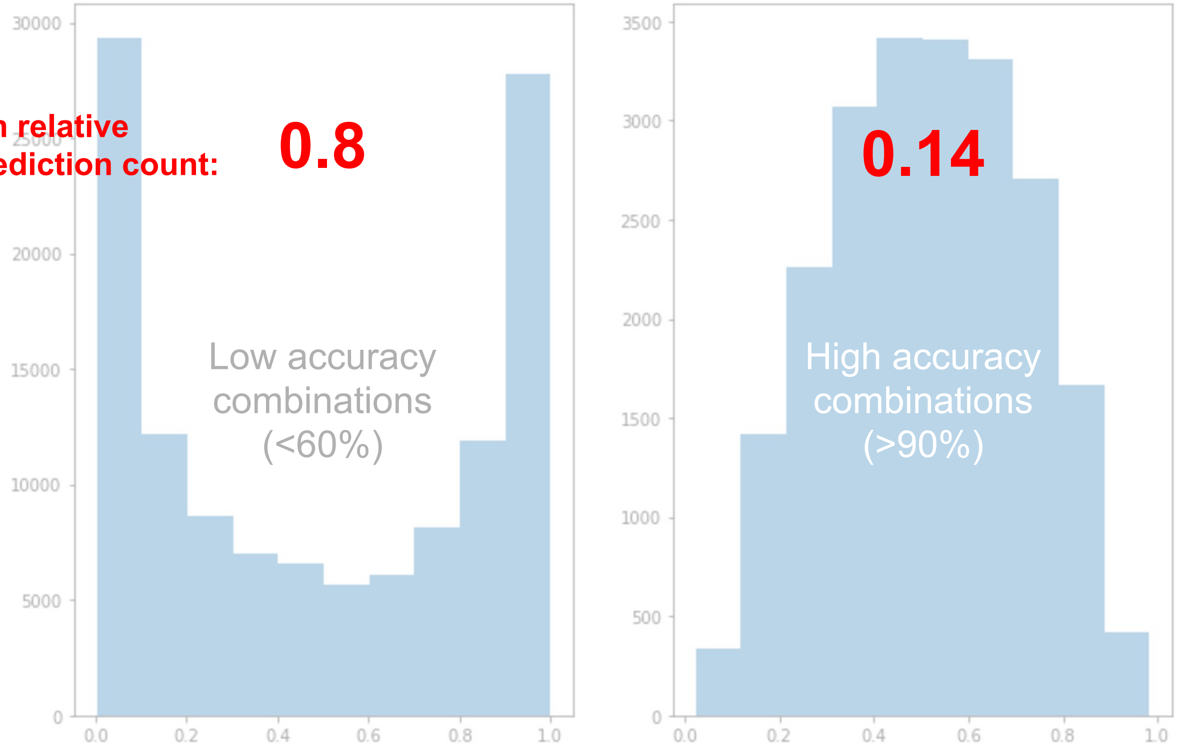
The importance of similar lengths in category reps



Relative length of the category A representative, in a 2-category classifier, so .5 means equal lengths

The importance of similar lengths in category reps

Correlation between relative length & relative prediction count:



Relative length of the category A representative, in a 2-category classifier, so .5 means equal lengths

3-pronged approach to improving the system

1. Better document vectors

2. Better ways of identifying potential category representatives.

3. Better guidelines for the user / overall better UX.

3-pronged approach to improving the system

1. Better document vectors

2. Better ways of identifying potential category representatives.

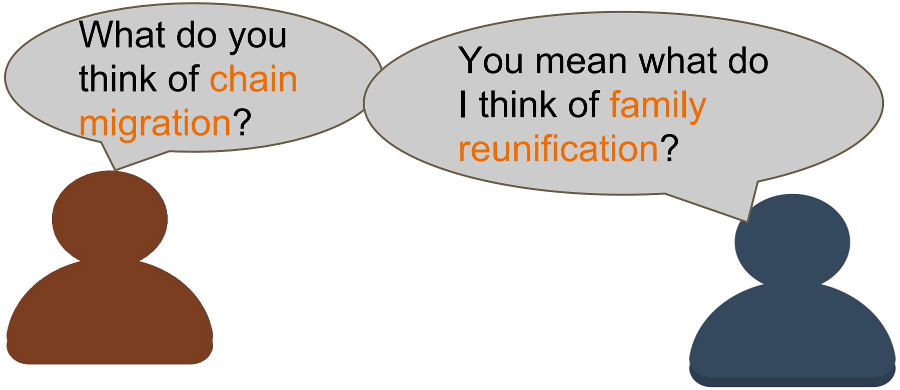
3. Better guidelines for the user / overall better UX.

Next steps

- Multi-label classification
- Working with partially labeled datasets
- Other languages? Use FastText pre-trained embeddings?
- Topic stance detection...

Next steps

- Multi-label classification
- Working with partially labeled datasets
- Other languages? Use FastText pre-trained embeddings?
- Topic stance detection...



Thanks :)

@katherinebailey