Project related paper

Character-level Convolutional Networks for Text Classification

Yuhui Lin

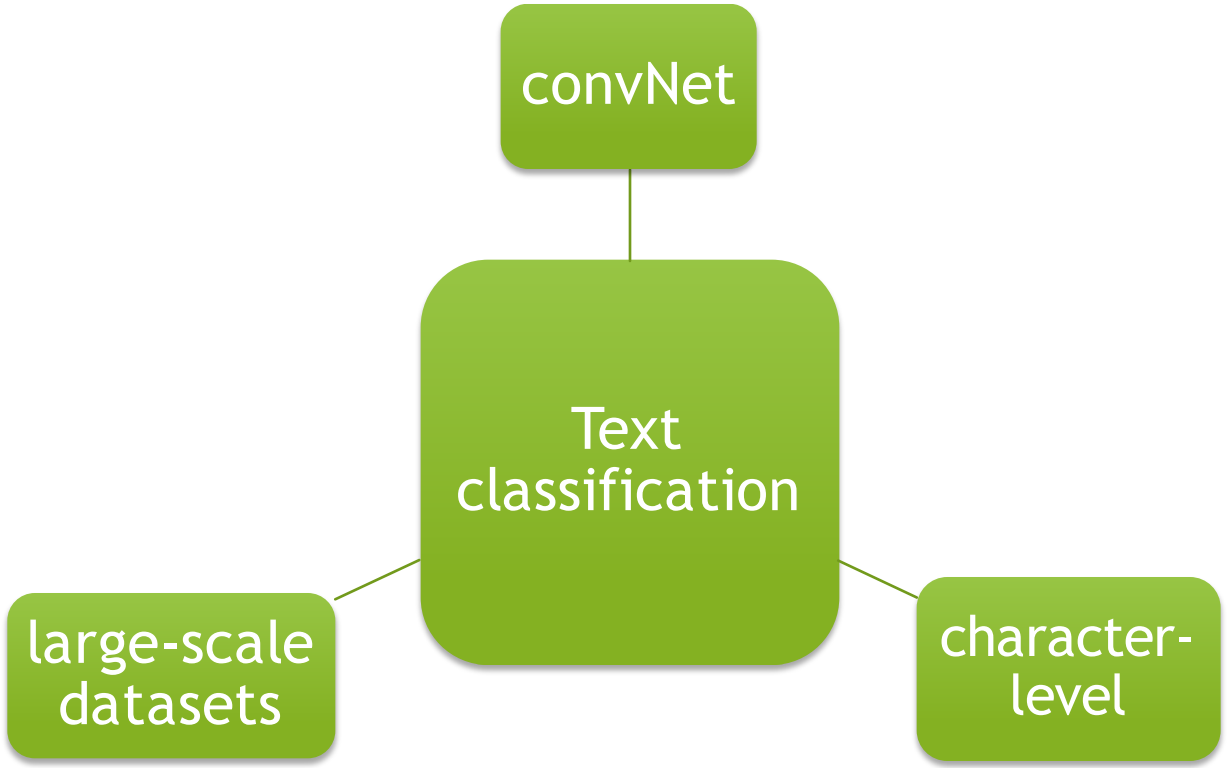
Introduction

- Introduce text classification
- The task is to assign a document to one or more predefined classes or categories.
- Topic classification:
  - Arts, business, sports
- Sentiment classification:
  - Negative, neutral, positive
- Functional classification:
  - Thethis, essay, announcement

Background

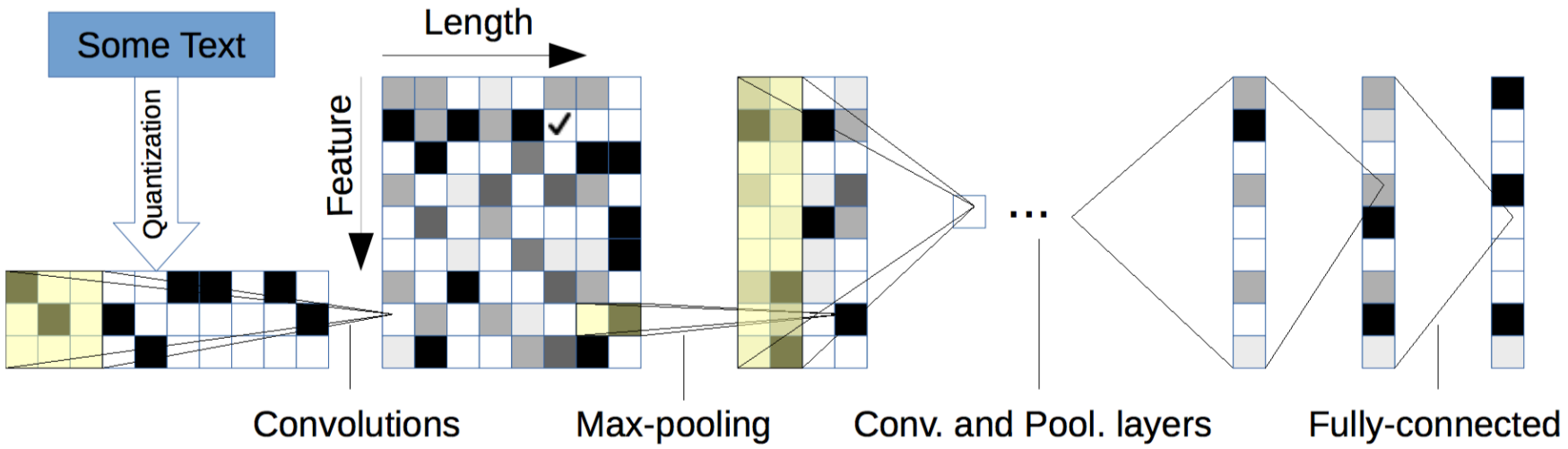
- The success of convNet in following field:
  - Computer vision
  - Speech recognition
- convNet with raw signal
  - Image pixels,
  - voice,
  - text character(English letter, punctuation, space)
- convNet with other NLP problems
  - Part-of-speech tagging
  - Information retrieval

Innovation


convNet
Text classification
large-scale datasets
character-level

Model design

The model has 9 layers deep with 6 convolutional layers and 3 fully-connected layers.

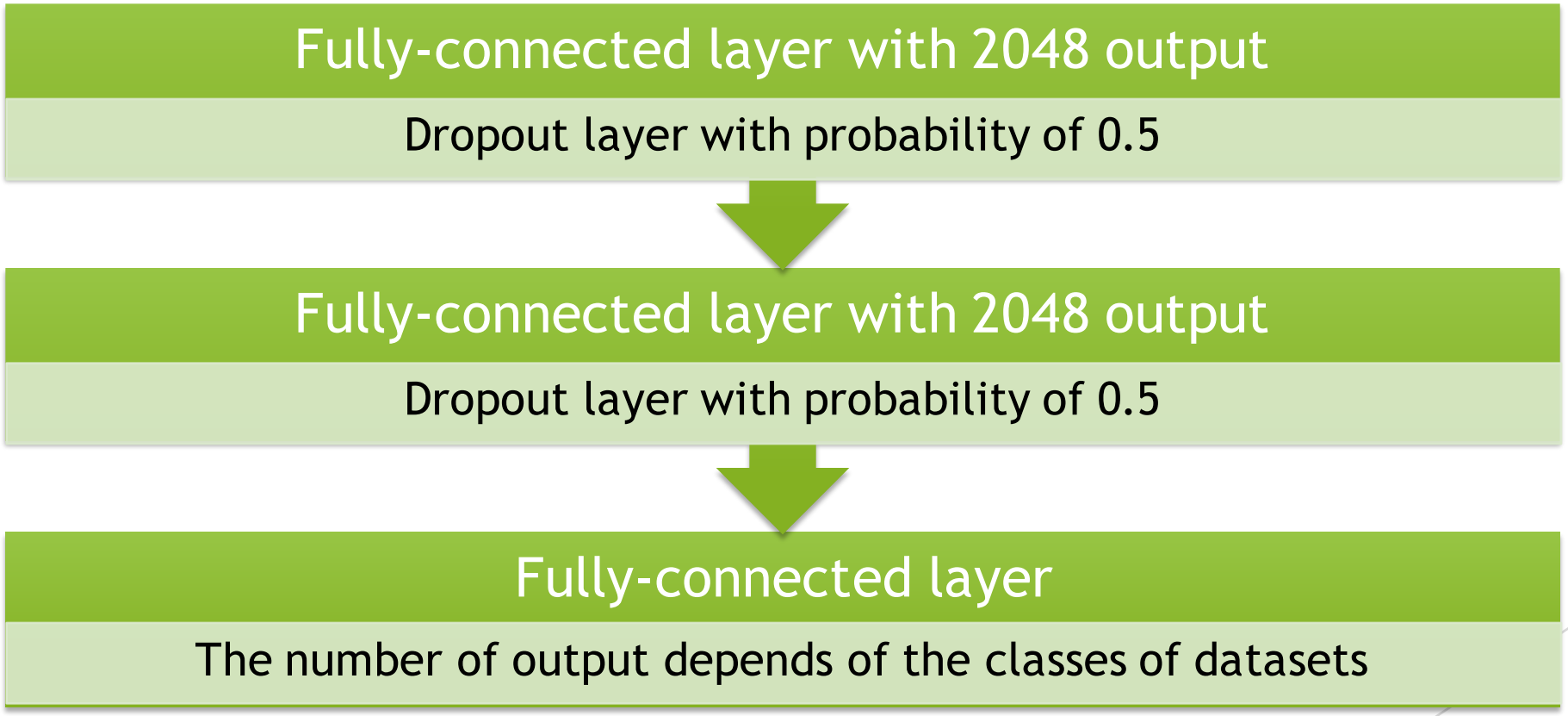


Convolutional Layers

layer	#features	# kernel	Max Pooling
1	1024	7	3
2	1024	7	3
3	1024	3	N/A
4	1024	3	N/A
5	1024	3	N/A
6	1024	3	3

The convolutional layers have stride 1 and pooling layers are all non-overlapping ones.

Fully-connected Layers

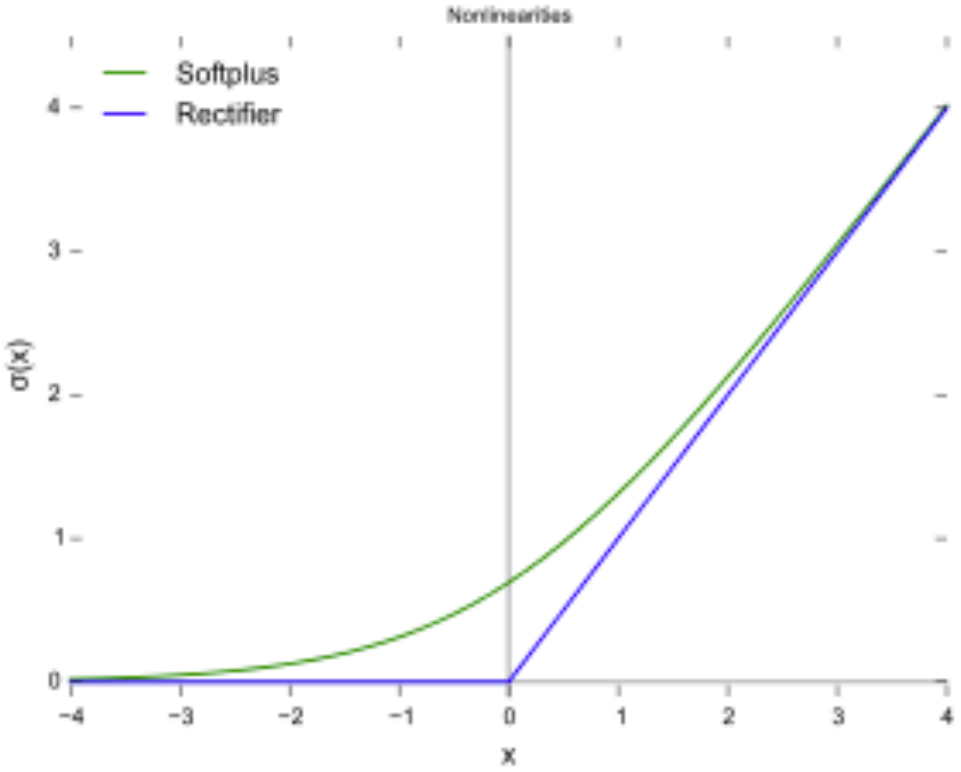


Gradient descent

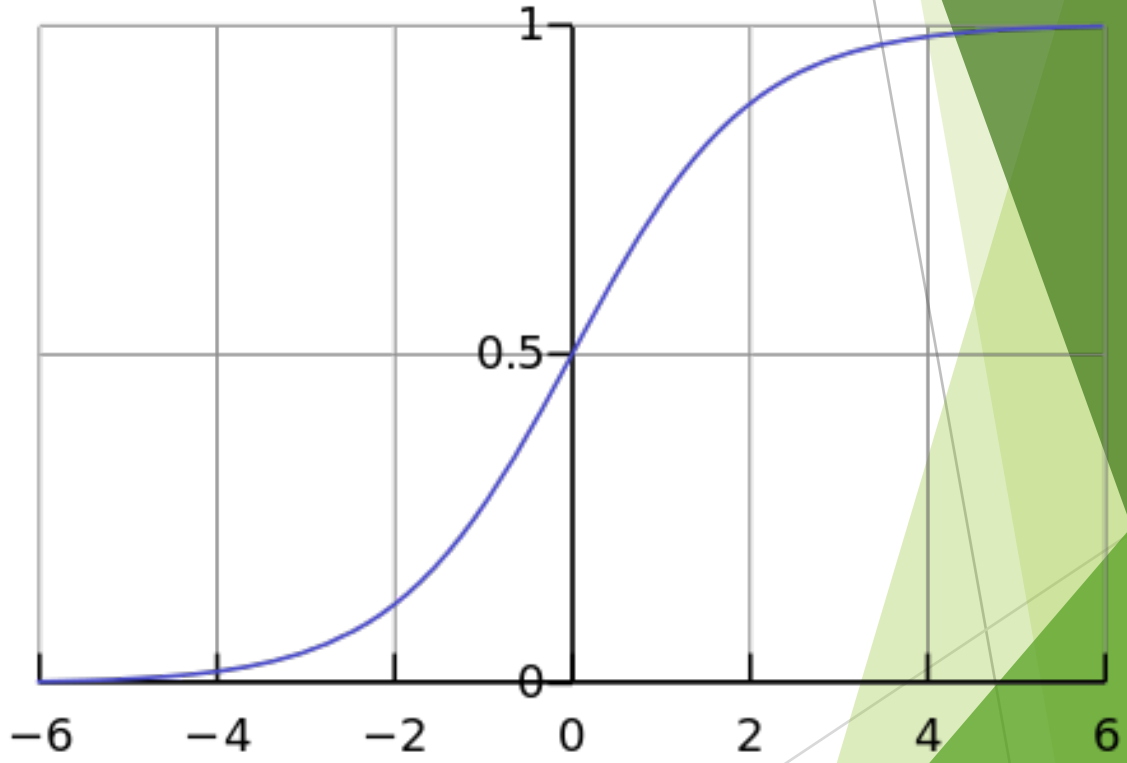
- Mini batch gradient descent with momentum:

```
loop maxEpochs times
  loop until all data items used
    for-each batch of items
      compute a gradient for each weight and bias
      accumulate gradient
    end-batch
    use accumulated gradients to update each weight and bias
  end-loop all item
end-loop
```

Non-linearity activation function ReLUs



ReLu $f(x) = \max(0, x)$



Sigmoid function

Data Augmentation

- data augmentation techniques are useful for reducing generalization error
- traditional techniques don't apply for text
- replace words by using an English thesaurus

Comparison Model

- Bag-of-words with softmax
- N-grams with softmax
- K-means on word2vec with softmax
- Word-based ConvNets
- Word-based LSTM

Large-scale Datasets

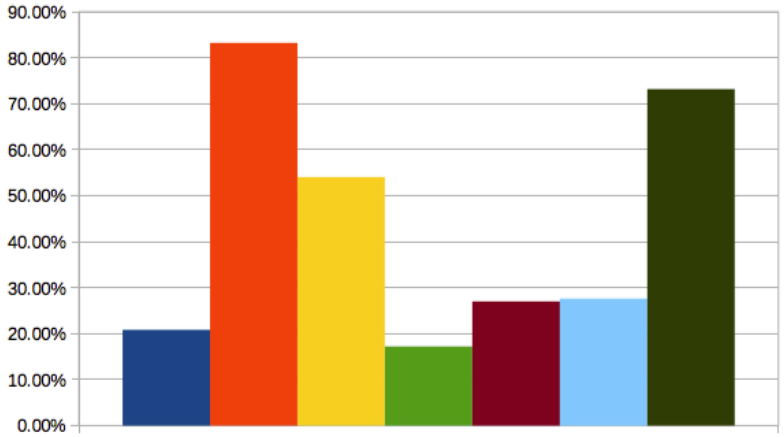
8 large-scale datasets:

Dataset	Classes	Train Samples	Test Samples	Epoch Size
AG's News	4	120,000	7,600	5,000
Sogou News	5	450,000	60,000	5,000
DBPedia	14	560,000	70,000	5,000
Yelp Review Polarity	2	560,000	38,000	5,000
Yelp Review Full	5	650,000	50,000	5,000
Yahoo! Answers	10	1,400,000	60,000	10,000
Amazon Review Full	5	3,000,000	650,000	30,000
Amazon Review Polarity	2	3,600,000	400,000	30,000

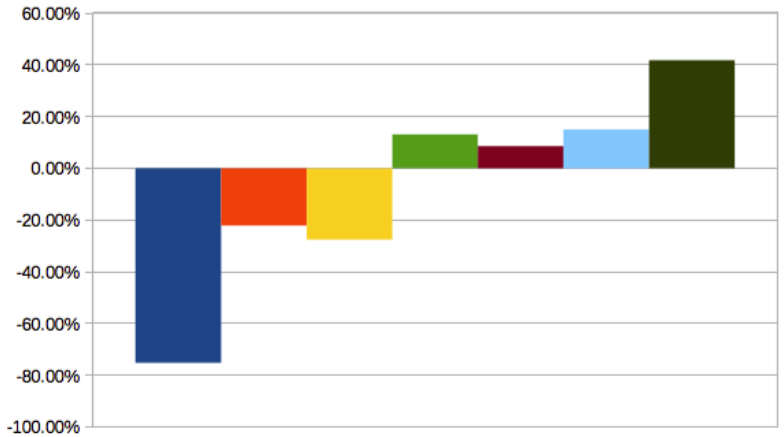
Result: error rate

Model	AG	Sogou	DBP.	Yelp P.	Yelp F.	Yah. A.	Amz. F.	Amz. P.
BoW	11.19	7.15	3.39	7.76	42.01	31.11	45.36	9.60
BoW TFIDF	10.36	6.55	2.63	6.34	40.14	28.96	44.74	9.00
ngrams	7.96	2.92	1.37	**4.36**	43.74	31.53	45.73	7.98
ngrams TFIDF	**7.64**	**2.81**	**1.31**	4.56	45.20	31.49	47.56	8.46
Bag-of-means	**16.91**	**10.79**	**9.55**	**12.67**	**47.46**	**39.45**	**55.87**	**18.39**
LSTM	13.94	4.82	1.45	5.26	41.83	29.16	40.57	6.10
Lg. w2v Conv.	9.92	4.39	1.42	4.60	40.16	31.97	44.40	5.88
Sm. w2v Conv.	11.35	4.54	1.71	5.56	42.13	31.50	42.59	6.00
Lg. w2v Conv. Th.	9.91	-	1.37	4.63	39.58	31.23	43.75	5.80
Sm. w2v Conv. Th.	10.88	-	1.53	5.36	41.09	29.86	42.50	5.63
Lg. Lk. Conv.	8.55	4.95	1.72	4.89	40.52	29.06	45.95	5.84
Sm. Lk. Conv.	10.87	4.93	1.85	5.54	41.41	30.02	43.66	5.85
Lg. Lk. Conv. Th.	8.93	-	1.58	5.03	40.52	28.84	42.39	5.52
Sm. Lk. Conv. Th.	9.12	-	1.77	5.37	41.17	28.92	43.19	5.51
Lg. Full Conv.	9.85	8.80	1.66	5.25	38.40	29.90	40.89	5.78
Sm. Full Conv.	11.59	8.95	1.89	5.67	38.82	30.01	40.88	5.78
Lg. Full Conv. Th.	9.51	-	1.55	4.88	38.04	29.58	40.54	5.51
Sm. Full Conv. Th.	10.89	-	1.69	5.42	**37.95**	29.90	40.53	5.66
Lg. Conv.	12.82	4.88	1.73	5.89	39.62	29.55	41.31	5.51
Sm. Conv.	15.65	8.65	1.98	6.53	40.84	29.84	40.53	5.50
Lg. Conv. Th.	13.39	-	1.60	5.82	39.30	**28.80**	40.45	**4.93**
Sm. Conv. Th.	14.80	-	1.85	6.49	40.16	29.84	**40.43**	5.67

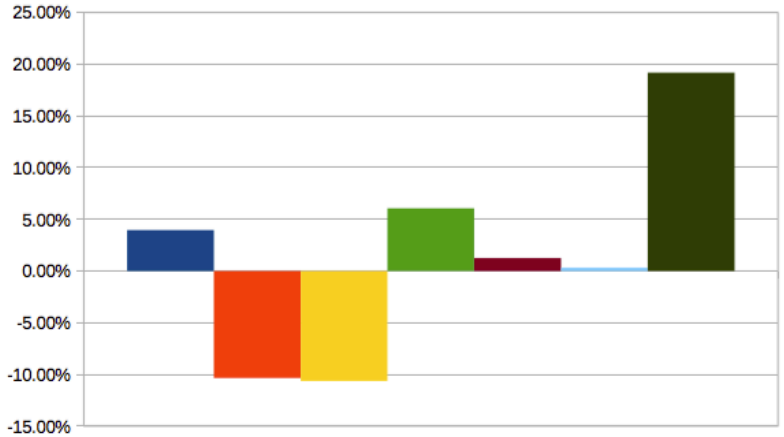
Character-level ConvNet is an effective



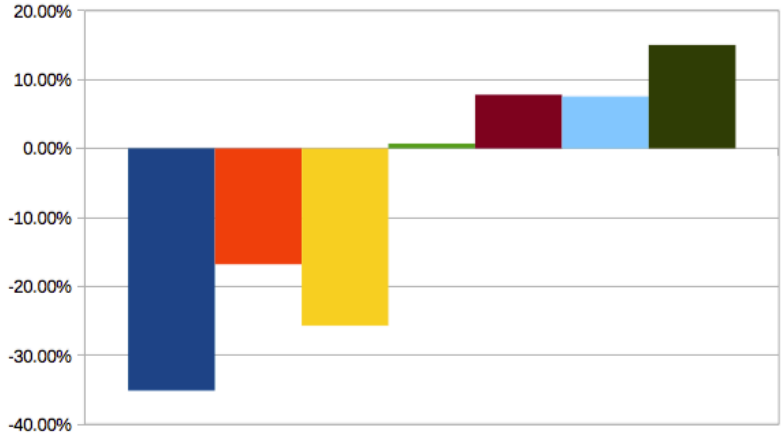
(a) Bag-of-means



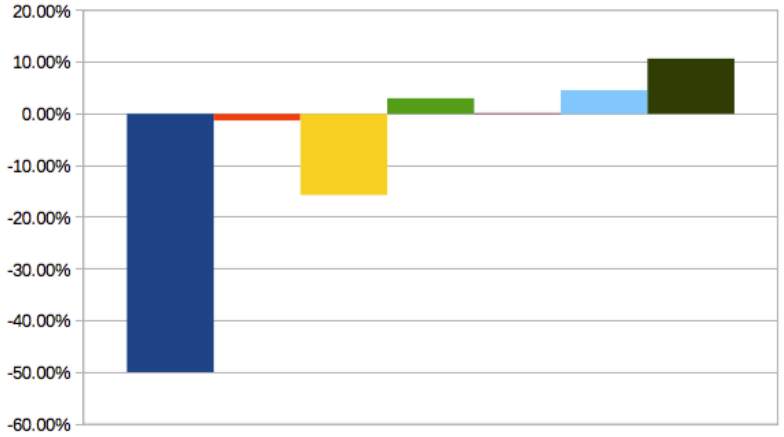
(b) n-grams TFIDF



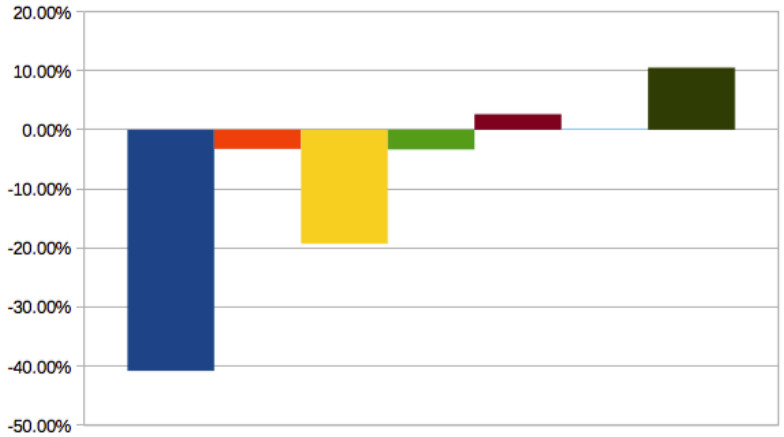
(c) LSTM



(d) word2vec ConvNet



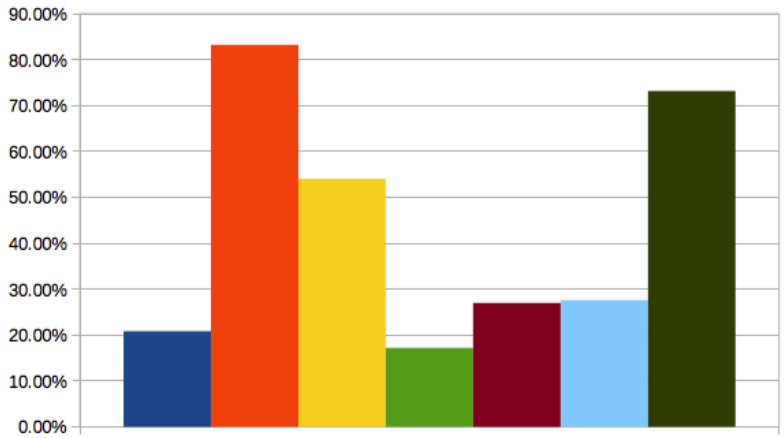
(e) Lookup table ConvNet



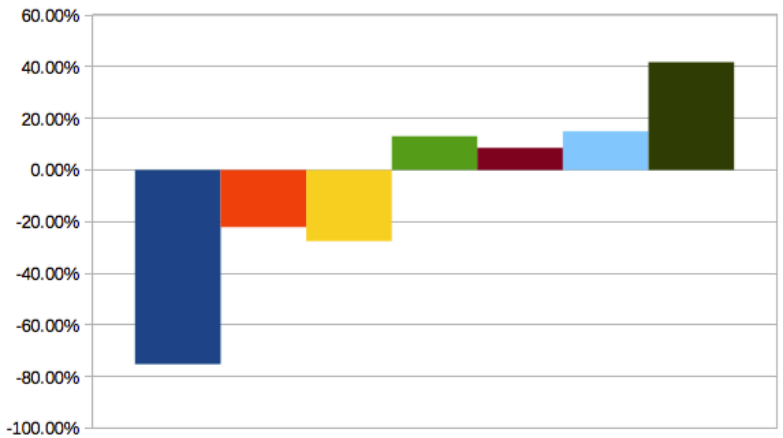
(f) Full alphabet ConvNet

AG News DBPedia Yelp P. Yelp F. Yahoo A. Amazon F. Amazon P.

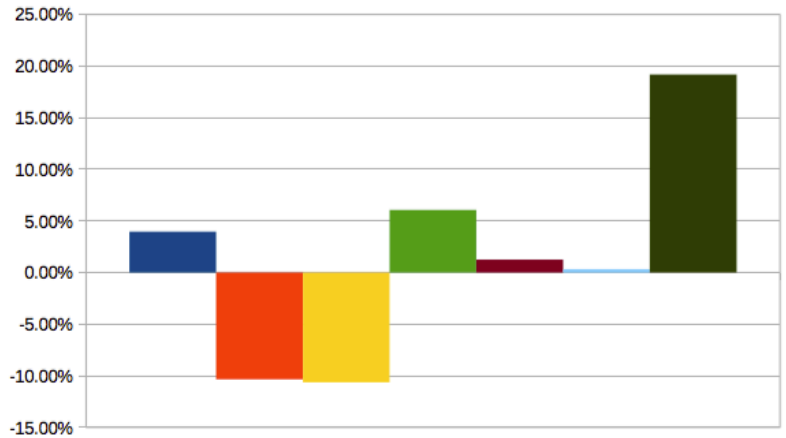
Dataset size forms a dichotomy between traditional and ConvNets models.



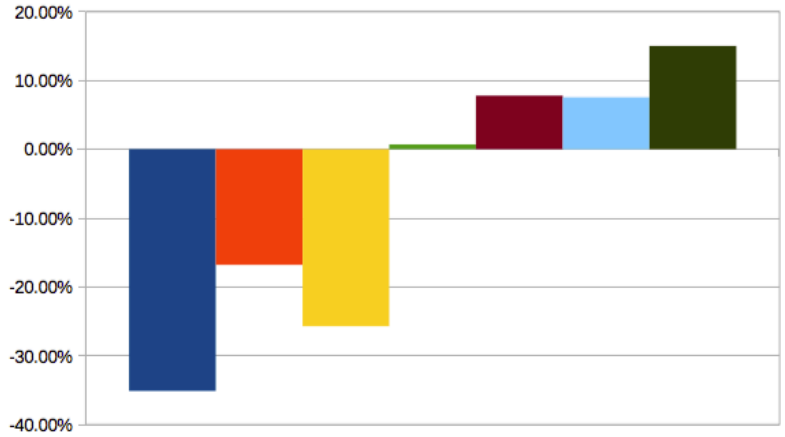
(a) Bag-of-means



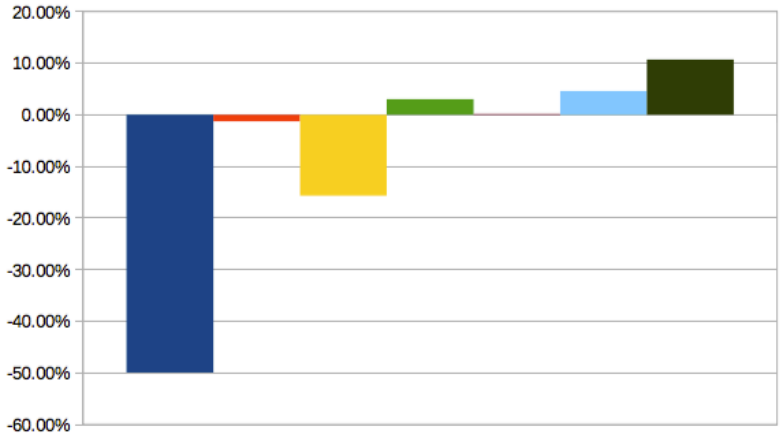
(b) n-grams TFIDF



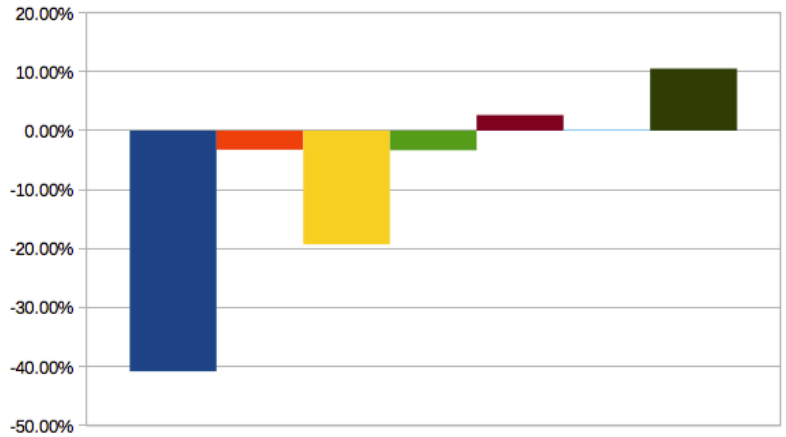
(c) LSTM



(d) word2vec ConvNet



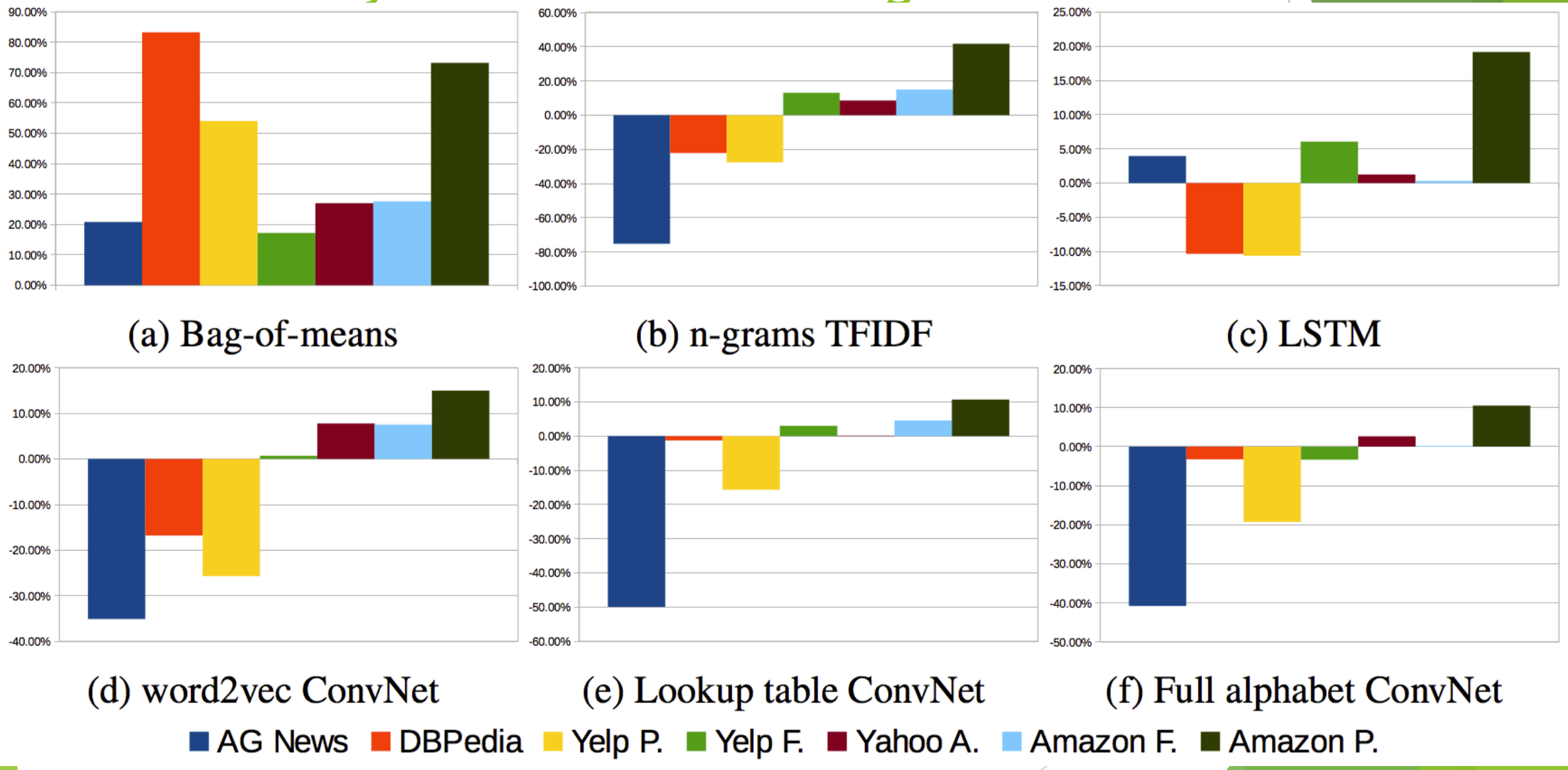
(e) Lookup table ConvNet



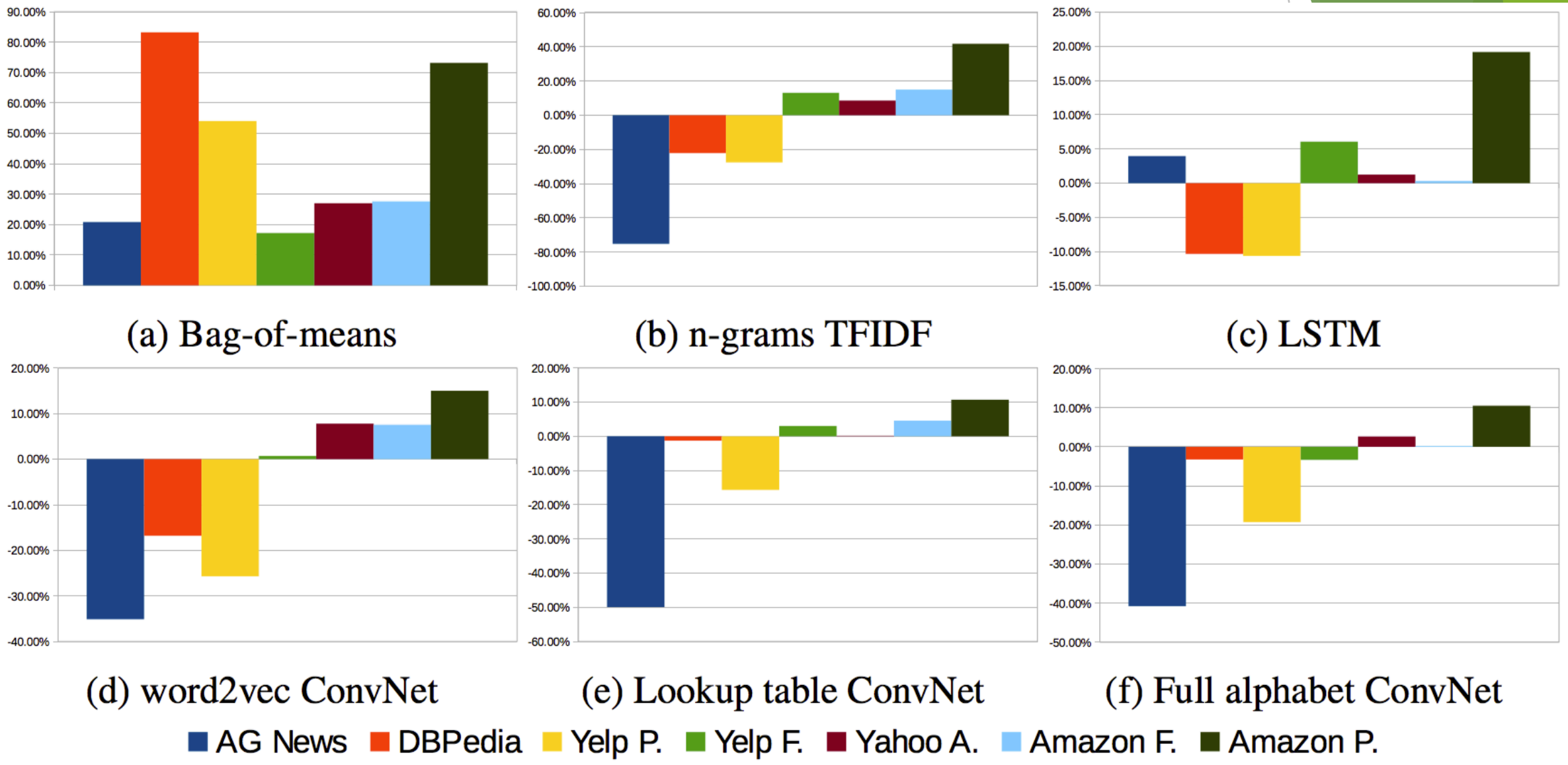
(f) Full alphabet ConvNet

AG News DBPedia Yelp P. Yelp F. Yahoo A. Amazon F. Amazon P.

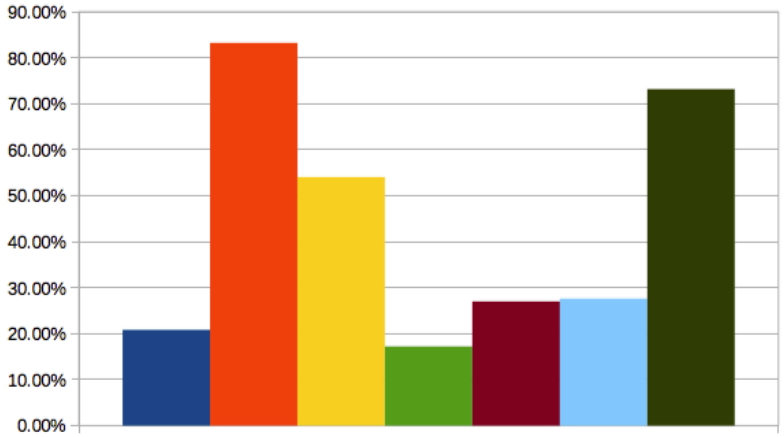
ConvNets may work well for user-generated data.



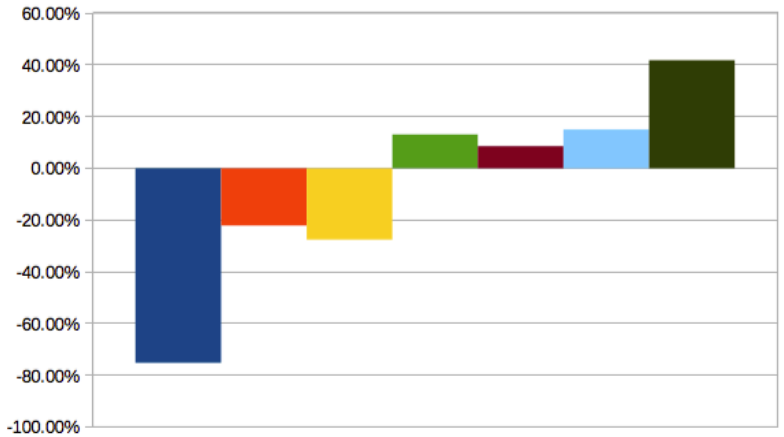
uppercase and lowercase letters could make a difference.



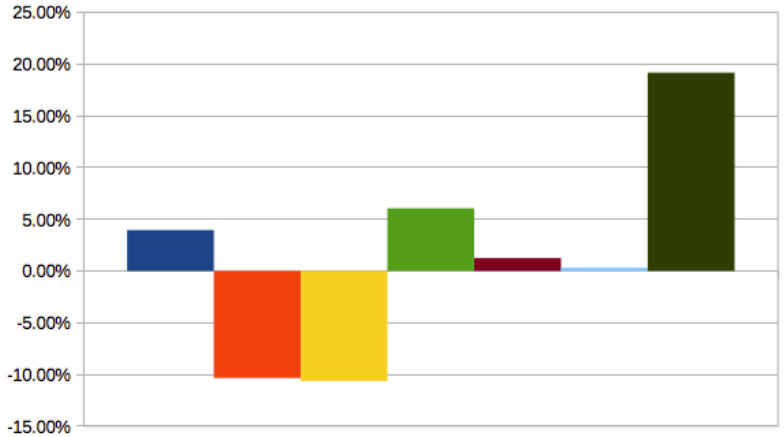
There is no free lunch.



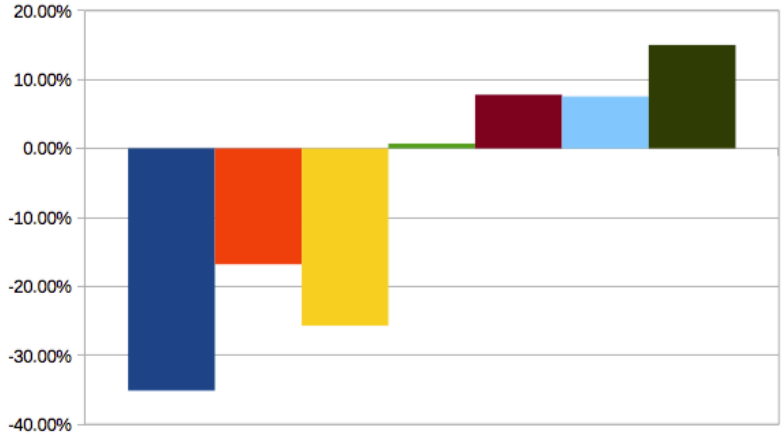
(a) Bag-of-means



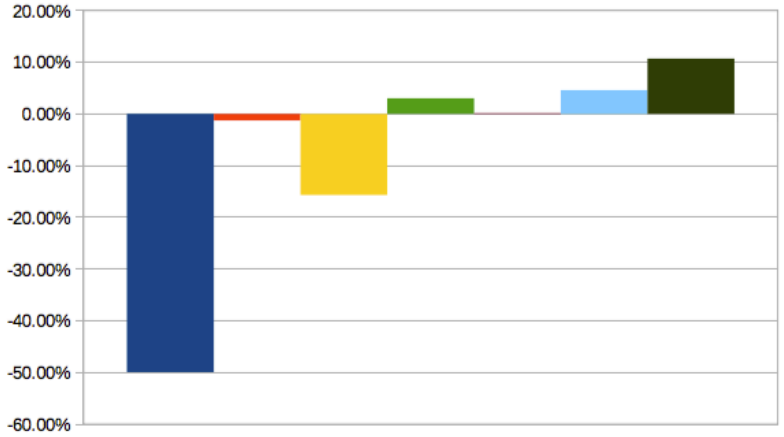
(b) n-grams TFIDF



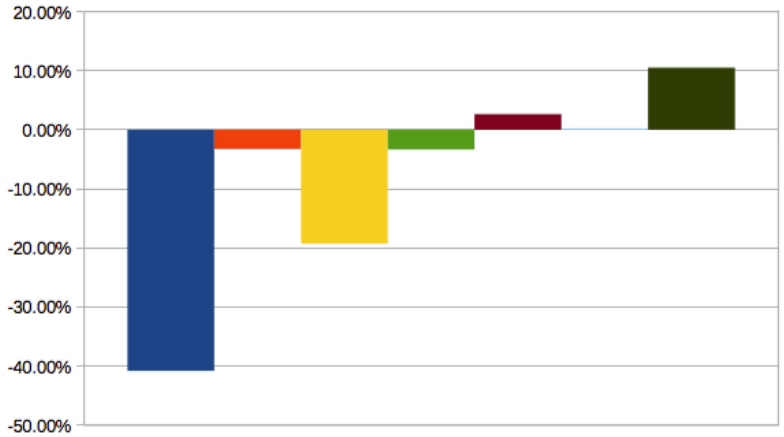
(c) LSTM



(d) word2vec ConvNet



(e) Lookup table ConvNet



(f) Full alphabet ConvNet

AG News DBPedia Yelp P. Yelp F. Yahoo A. Amazon F. Amazon P.

Question?

- Thank you