Going Deeper into Sentiment Analysis with CNN's

Sam Liu, Nabeel Sidiqqi

Outline

Executive Summary: Project goals

Background and motivation

Design & Implementation

Results & conclusions

Executive Summary: Project Goals

Goal: Leverage training speed and CNN architecture to predict sentiment of text Yelp reviews

Why: Faster than state of the art (RNN), more accurate than traditional ML

Result: Our CNN model outperformed SVM by 4 percentage points

- **Metrics (F1 Macro):**
  - CNN 94%
  - SVM 90%

Outline

Executive Summary: Project goals

Background and motivation

Design & Implementation

Results & conclusions

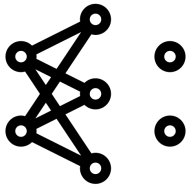
Sentiment analysis a hot topic for companies

Use cases:

- Better understanding of their products
- Timely customer support to unhappy customers
- Politicians understand the online sentiment of public opinion

CNN is a practical tool for text sentiment analysis

CNN: relatively new for language, traditionally used for image



- Can learn embedding directly from text
- Fast to train (optimized for GPU)
- Better than traditional ML (tf–idf+SVM): learns order context

New architecture: CV applications show deeper is better

- Show performance at different levels of depth

Hypothesis: deeper CNN will perform better

Outline

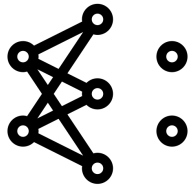
Executive Summary: Project goals

Background and motivation

Design & Implementation

Results & conclusions

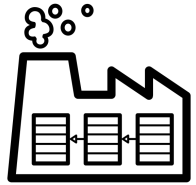
Design & Implementation



Hyperparameter decisions

- **Filter size:** size of the convolutional layer
- **Kernel size:** number of words to include in each pass
- **Activation:** choice of activation to potentially avoid dying ReLU's
- **Embedding Size:** Size of the embedding layer
- **Number of convolutional layers:** depth of model

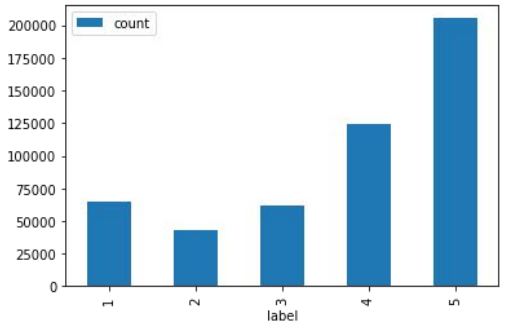
Data exploration and pre-processing



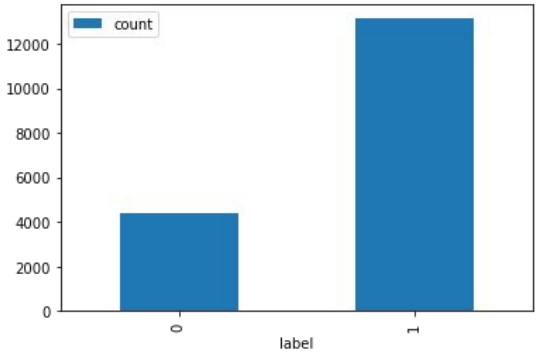
Data processing:

1. Convert multinomial dataset to binary classification
2. Balanced skew by undersampling majority class

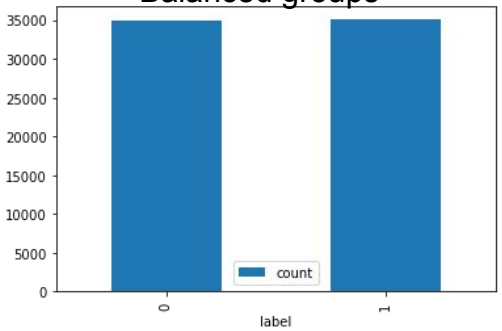
Original Dataset



Grouped (1,2), (4,5)



Balanced groups



Outline

Executive Summary: Project goals

Background and motivation

Design & Implementation

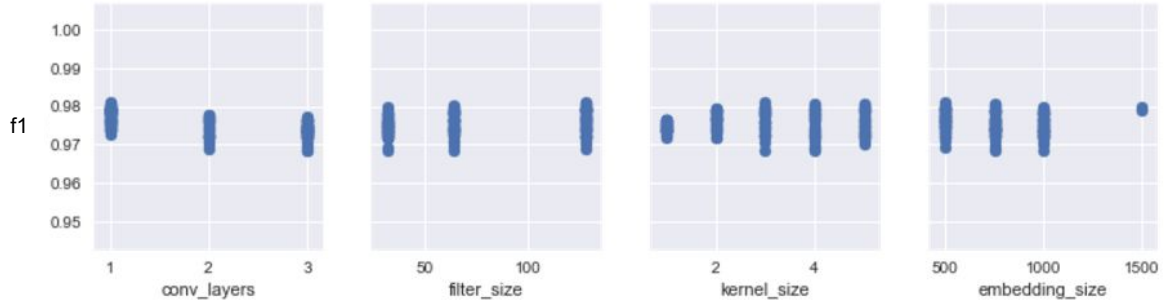
Results & conclusions

Deeper models tend to perform well

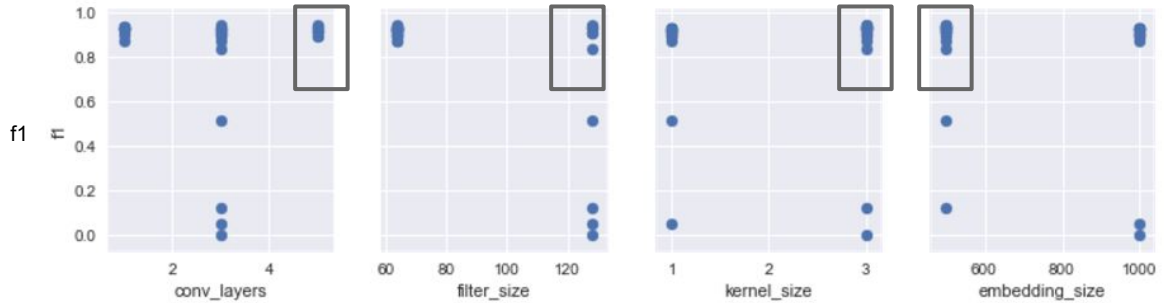
conv_layers	filter_size	kernel_size	embedding_size	activation	f1
5	128	3	500	relu	0.941688
3	64	3	500	leakyrelu	0.939556
1	64	3	500	relu	0.938586
1	128	3	500	relu	0.934110
5	64	3	500	leakyrelu	0.931836
3	128	1	500	relu	0.518116
3	128	3	500	relu	0.123827
3	128	1	1000	relu	0.054609

Deeper models tend to perform well

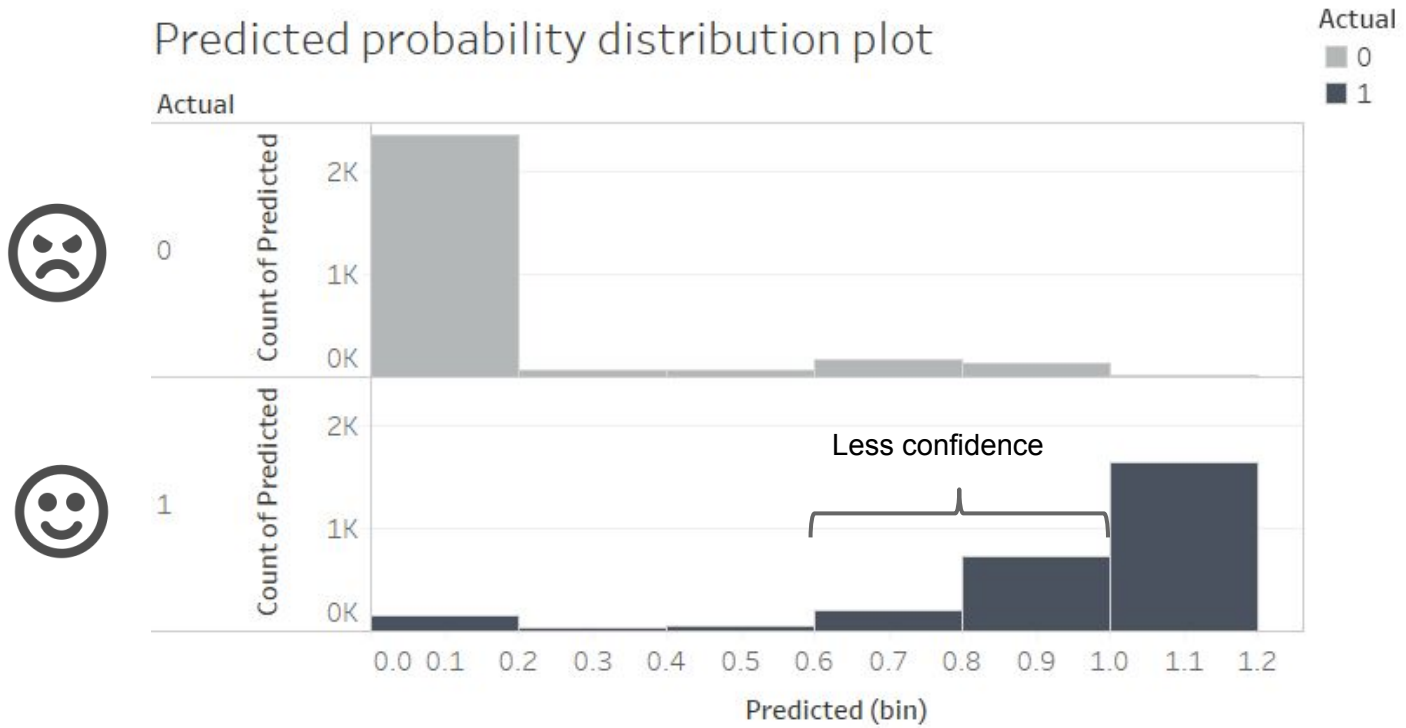
First iteration:



Second iteration:

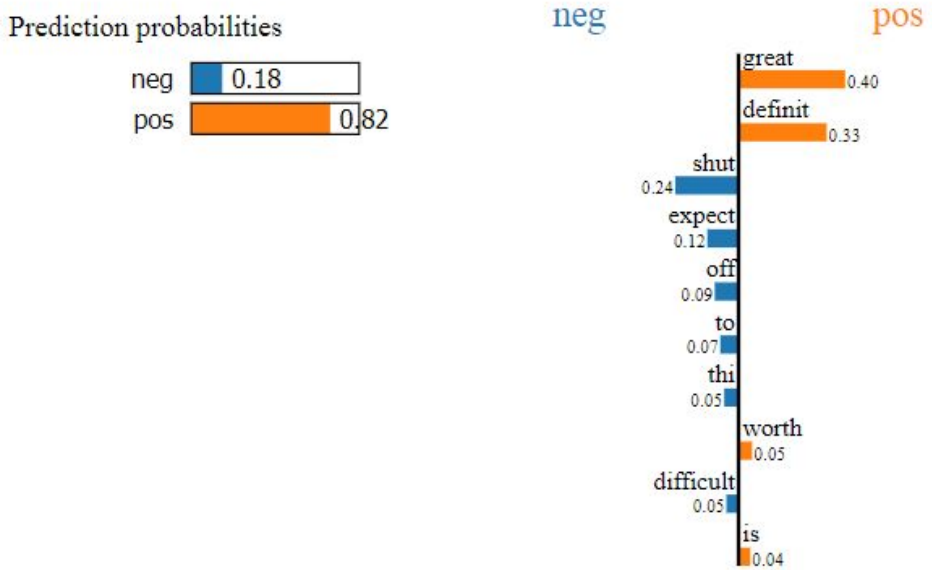


Positive review predictions are less confident than negative reviews



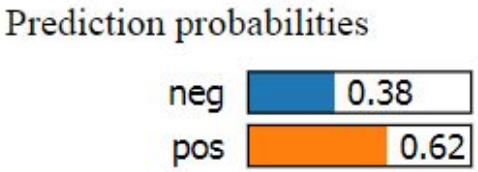
Positive reviews that contains more negative words throws off model confidence

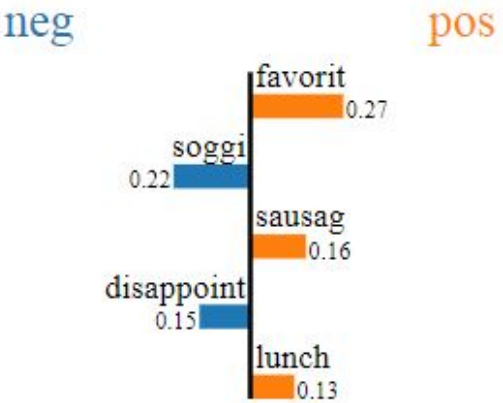
```
Document id: 95
True class: pos
```



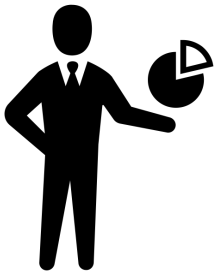
CNN not always able to detect sentiment where strong pos and neg words shows up

True class: neg





Conclusion & future research



- Showed that 5 layer deep models did perform better than shallower models but embedding size matters as well
- What is the relationship between depth and size of data?
- What is the optimal order of convolutional, activation and normalization layers? Ioffe & Szegedy recommended Conv-BN-ReLU but other permutations perform well (He 2016)
- What is the relationship between embedding and data size?