

# Sentiment Analysis (Movie)

## Example Data Set - Rotten Tomato

PhraseId	SentenceId	Phrase	Senti
537	20	It's what you'd expect	2
538	20	what you'd expect	2
539	20	you'd expect	2
540	20	expect	3

0 – negative

1 – somewhat negative

2 – neutral

3 – somewhat positive

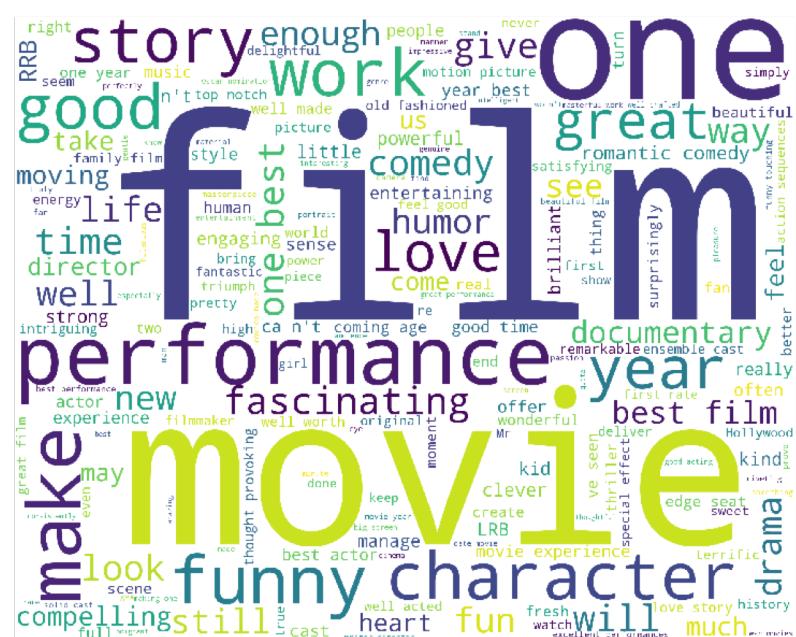
4 – positive



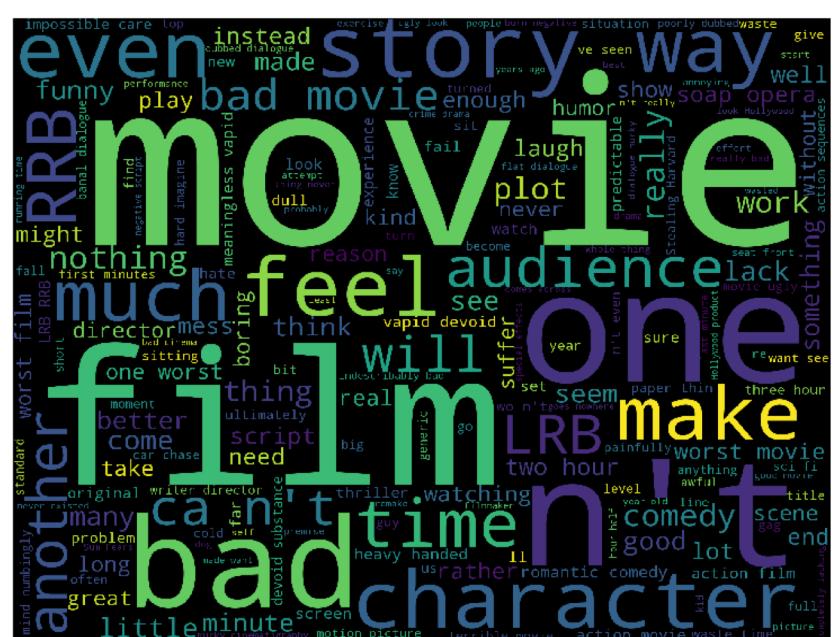
Data, Model, Code: [github.com/MangoHaha/SentimentAnalysis](https://github.com/MangoHaha/SentimentAnalysis)

The goal of the project is to predict the sentiment of phrases using the labeled dataset. The data "Rotten Tomatoes movie review dataset" is a corpus of movie reviews used for sentiment analysis. This Kaggle competition has been inspired by Socher's work on labeling parsed phrases in the corpus.

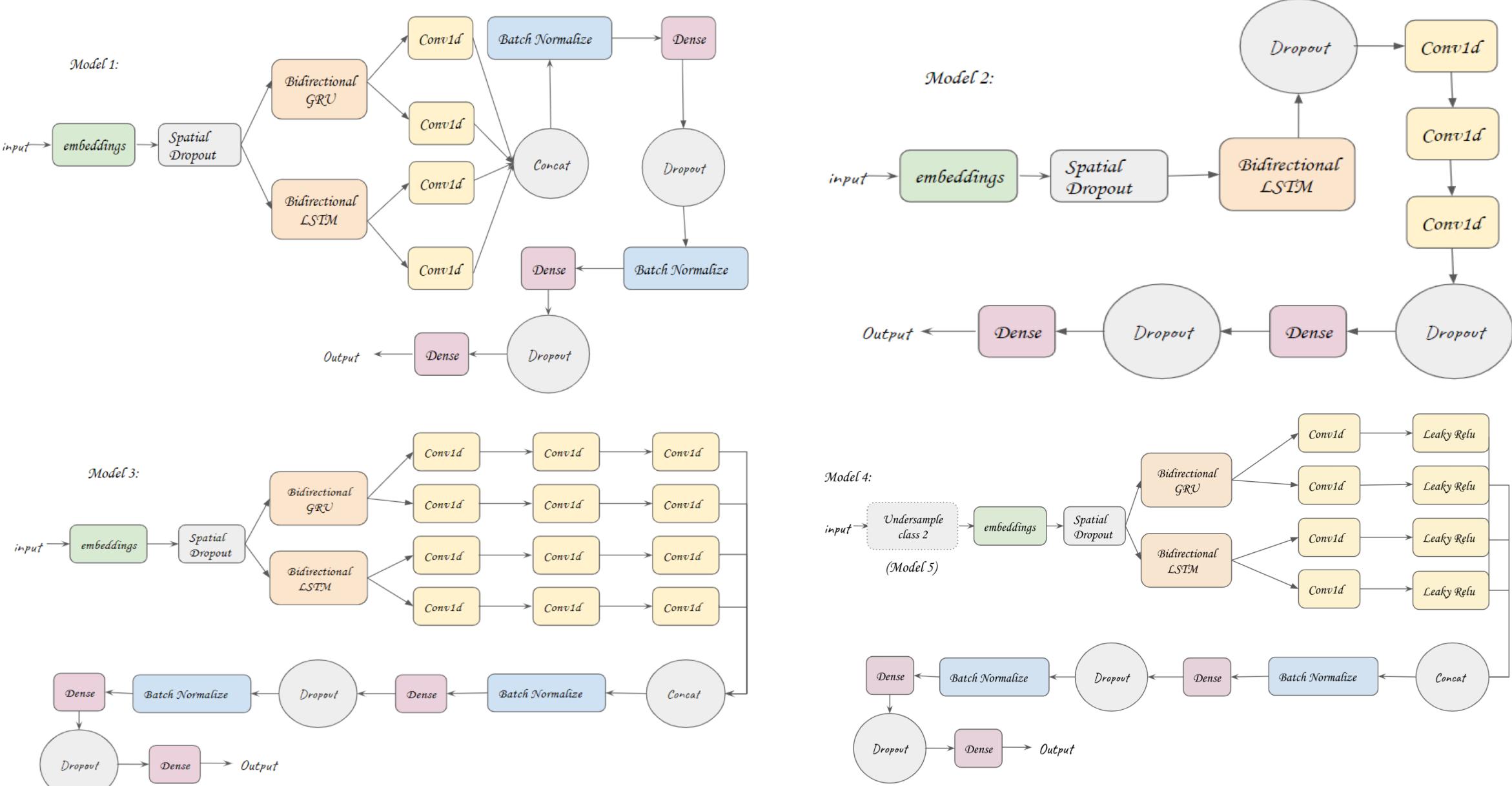
Positive Words:



Negative Words:



## Neural Network Architectures (RNN/CNN)



## Results

Table 3: Original Class Distribution

Sentiment	Count
0	5%
1	17%
2	51%
3	21%
4	6%

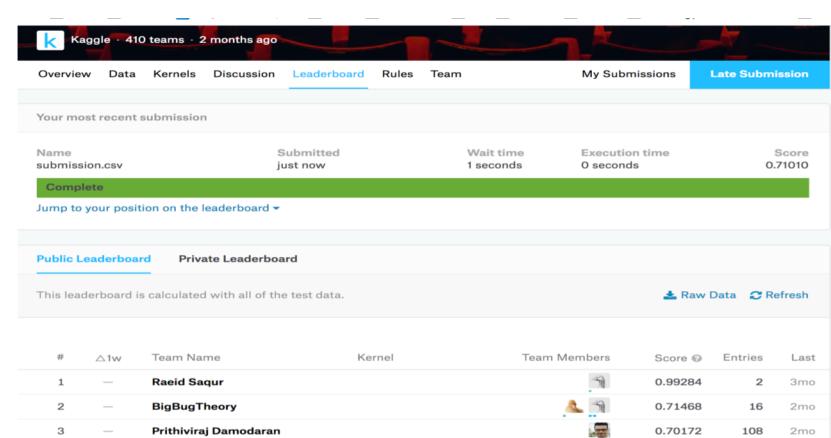
Redistribution  
Of Rare Class

Table 4: New Class Distribution

Sentiment	Count
0	6%
1	23%
2	34%
3	28%
4	8%

For comparison, our competitors on Kaggle achieves best score of 0.71468 while the second bestscore is 0.70172.(Igoring the first place of 0.99 which is unrealistic).

Our achieved 0.71 on Kaggle. This result outperforms the second best score and is very close to the best score.



Kaggle Top  
2 of 410  
teams

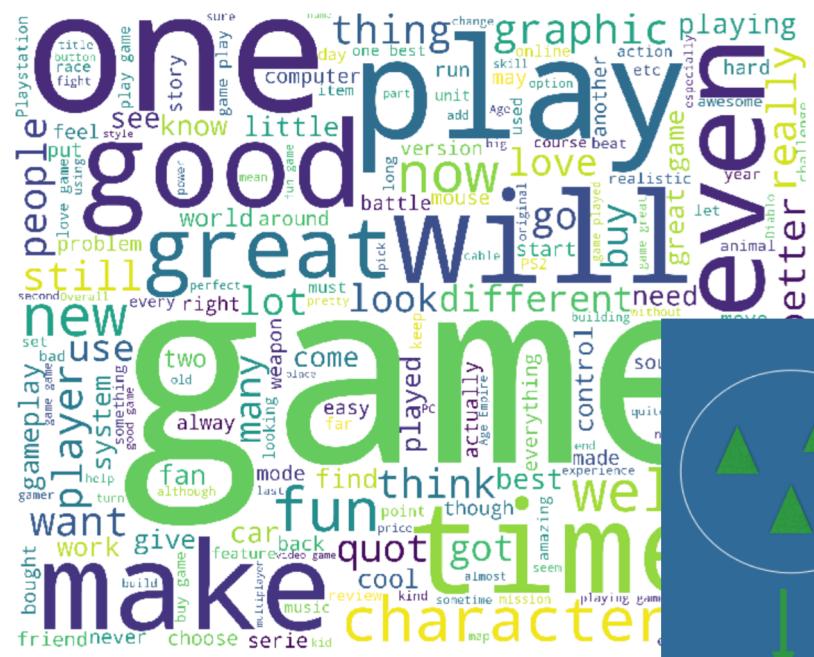
model	loss	acc	val loss	val acc	kaggle
model1	.274	.875	.237	.892	0.710
model2	.679	.717	.608	.749	0.706
model3	.270	.877	.245	.889	0.711
model4	.277	.874	.250	.887	0.701
model5	.252	.873	.230	.891	0.680
ensemble	.265	.881	.240	.891	0.712



# Sentiment Analysis (Game)

# Example Data Set – Transfer Learning

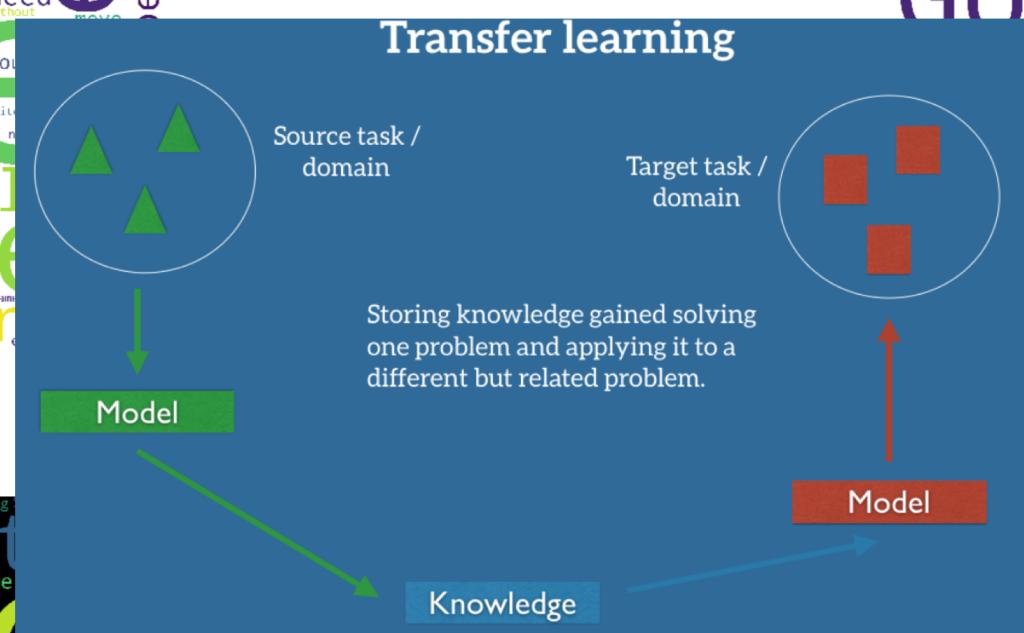
## Positive Words: amazon



## Negative Words: amazon



Model trained by Game review, compared to  
The same one without Game review  
RMSE has been improved by 51.2%

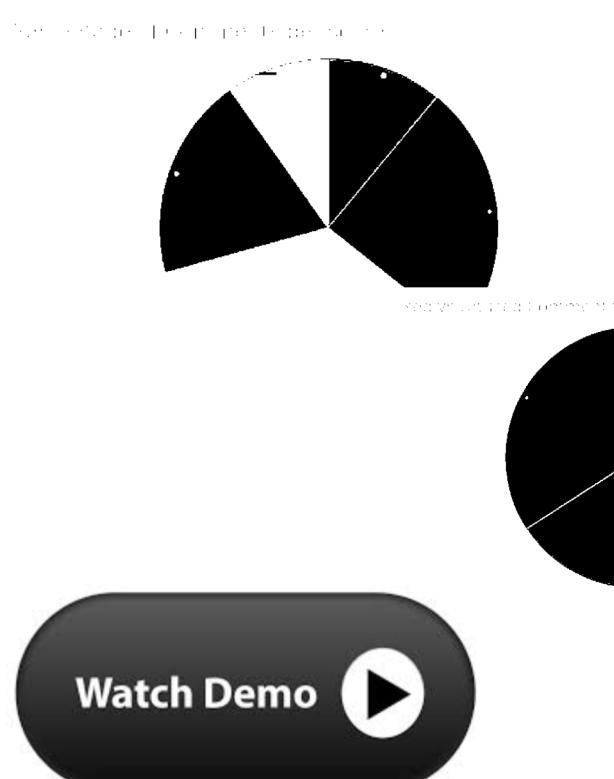


Transfer learning is broadly used for image classification.

For example: knowledge gained while learning to recognize cars could apply when trying to recognize trucks.



## Negative Words: PS



## Linguistic Regularization



- Those special words play a critical role in modifying sentiment of textual expressions. The basic idea is to enforce the model to produce coherent predictions - a new loss term is plugged into the original cross entropy loss:

$$\mathcal{L}(\theta) = -\sum_i y_i \log y_i + \lambda \sum_t \sum_i \mathcal{L}_{t,i} + \beta ||\theta||^2$$

**Who else think this project is good??!**

