Popularity Analysis for Youtube Makeup Videos

Shiyi Cheng, Kexin Wang, Ruchen Zhang, Zhuocheng Shang, Chiyuan Ma

Motivation & challenges

- **Motivation:**
- Makeup videos are popular
- What are collative concepts among trending videos

- **Challenges:**
- Crawling data from Youtube Channels
- Select algorithms to train dataset

Data preprocessing

- **Data Collection:**
- Crawling data from Youtube
- Data Attributes
- [ID, Title, View Count, Like Count, Dislike Count, Publish Date, Description]

- **Data Cleaning :**
- (Title & Description)
- NLTK
- Bag of Word
- Tf-idf Vectorizer

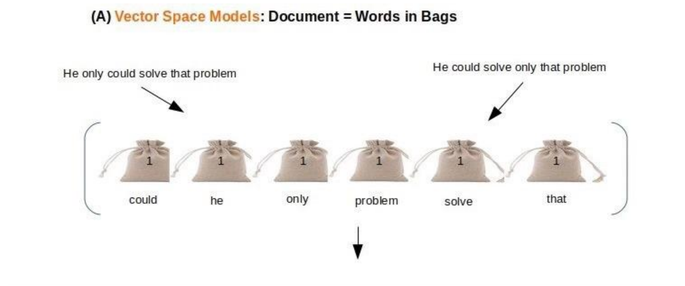
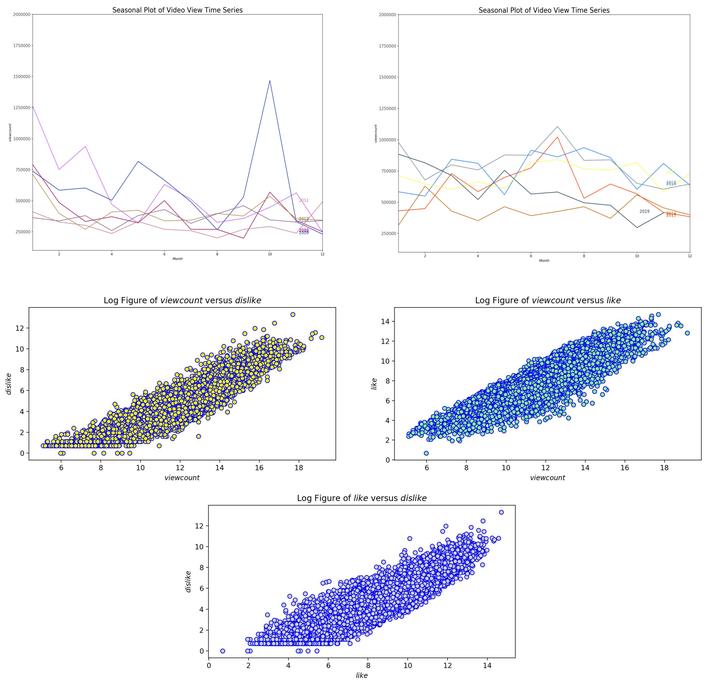
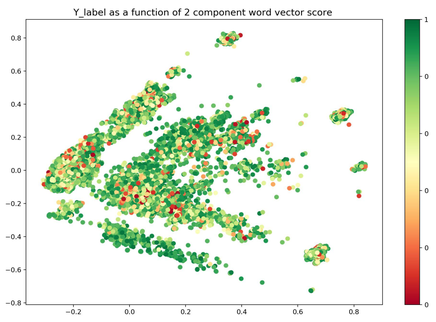


Figure . Bag of word





Data Describe

count	38899.000000
mean	0.973436
std	0.019546
min	0.900000
25%	0.965217
50%	0.979620
75%	0.987576
max	1.000000

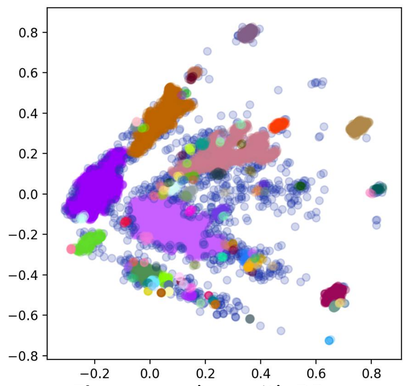


Figure . analyze with DBscan

Top Cluster	TERM	WEIGHT In each cluster
1	video	0.021817
	makeup tutorial	0.012237
	use code	0.011442
	smokey eye	0.007478
	eye makeup	0.007064
2	use code	0.010647
	affiliate link	0.008296
	Code lauralee	0.007519
	makeup tutorial	0.0067
	video like	0.006287

Top Cluster	TERM	WEIGHT In each cluster
3	affiliate link	0.022977
	use code	0.020865
	coupon code	0.019308
	video sponsored	0.017234
	makeup tutorial	0.01647
4	use code	0.014387
	code lauralee	0.0094
	affiliate link	0.00834
	code james	0.008169
	coupon code	0.00692

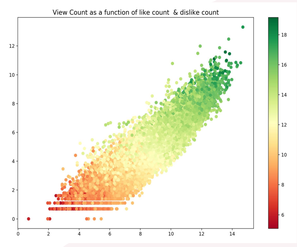
Table . analyze with DBscan

Algorithms

- **Variables:**
- X: [word vectors]
- Y: [like / (dislike + like)]
- Y_label:
- [“popular”] if Y(score) > mean
- [“unpopular”] if Y(score) < mean

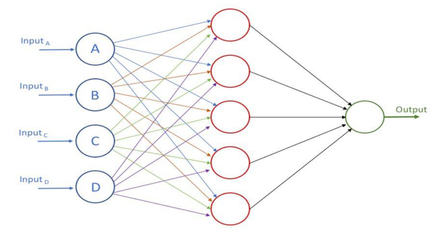
- **Models:**
- Neural Network
- Linear Regression
- Logistic Regression
- K Nearest Neighbors

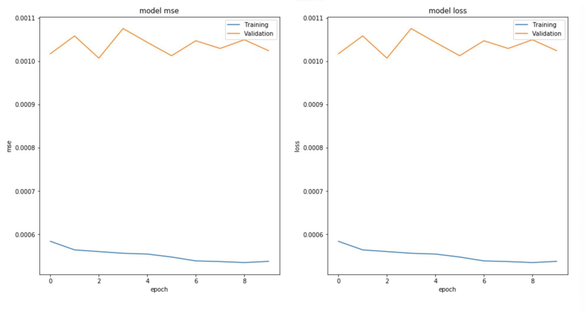
Linear Regression&Output



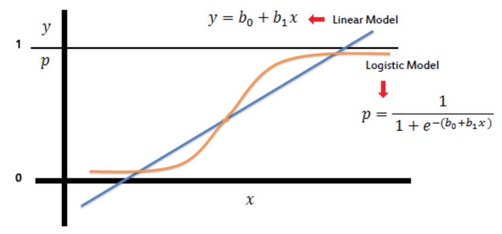
Linear Regression Summary:
# of training observations: 30282
Residual Standard Error: 0.03107
r-Squared: 0.1860813189038045
Log-Likelihood: 62153.9163681
Test score: 0.17740594576943547

NeuralNetwork&Output





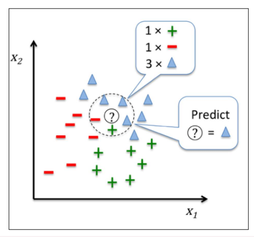
Logistic Regression&Output

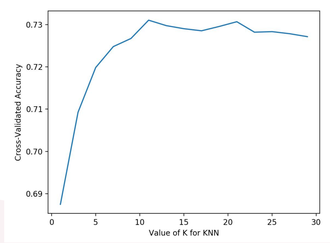


Class 0: popular Class 1: unpopular
Train Accuracy: 0.7554322699953768
Test Accuracy: 0.7494551218545671

	Pred_0	Pred_1
True_0	18528	2082
True_1	5324	4348

KNN&Output





KNN model get best accuracy 0.73 when K’s value is around 11.

Conclusions

- The public pays less attention to makeup videos.
- ‘View count’ and ‘like’, 'dislike' show a linear increasing relationship
- Word choice in ‘Title’ and ‘Description’ is collated with popularity of a video. Audience pay more attention to key words such as “makeup tutorial”, “eye makeup” or “coupon code”.
- Neural Network works well on our dataset.
- Compared with NN, Linear Regression' accuracy is not high.
- KNN 's accuracy is almost the same as LR.

Figure . Keyword Changed in Years

References

[1] Gloria Chatzopoulou, Cheng Sheng and Michalis Faloutsos. (2010). A first step towards understanding popularity in YouTube. *IEEE.*
[2] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl and Jose San Pedro. (2010). How Useful are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings. *Proceedings of the 19th international conference on World wide web (WWW '10).*
[3] Fiktor Imanuel Tanesab1 , Irwan Sembiring2 and Hindriyanto Dwi Purnomo3. (2017). Sentiment Analysis Model Based On Youtube Comment Using Support - Vector Machine. *International Journal of Computer Science and Software Engineering (IJCSSE)*
[4] Muhammad Zubair Asghar, Shakeel Ahmad, Afsana Marwat, Fazal Masud Kundi. Sentiment Analysis on YouTube: A Brief Survey.
[5] Soujanya Poriaa, Erik Cambriab, Newton Howardc, Guang-BinHuangd, Amir Hussain. (2015). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *NueroComputing Volume 174, Part A, 22 January 2016, Pages 50-59.*
[6] Ming-Te Chi, Shih-Syun Lin, Shiang-Yi Chen, Chao-Hung Lin, Tong-Yee Lee. Morphable Word Clouds for Time-Varying Text Data Visualization. *IEEE.*
[7] Aliaksei Severyn, Alessandro Moschitti, Olga Uryupina, Barbara Plank, Katja Filippova. Opinion Mining on YouTube.
[8] Stephanie Orme, Jared LaGroue, Arienne Ferchaud, Jenna Grzeslo. (2018). Parasocial attributes and YouTube personalities: Exploring content trends across the most subscribed YouTube channels. *Computers in Human Behavior*
[9] Lavanya Sunder. (2016). Predictive Model for Views In YouTube Beauty Community.
[10] Phakhawat Sarakit, Thanaruk Theeramunkong, Choochart Haruechaiyasak, Manabu Okumura. (2015). Classifying emotion in Thai youtube comments. *International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)978-1-4799-8565-4/15/$31.00 ©2015 IEEE*