

Text Analysis of Bilibili's Video Headline



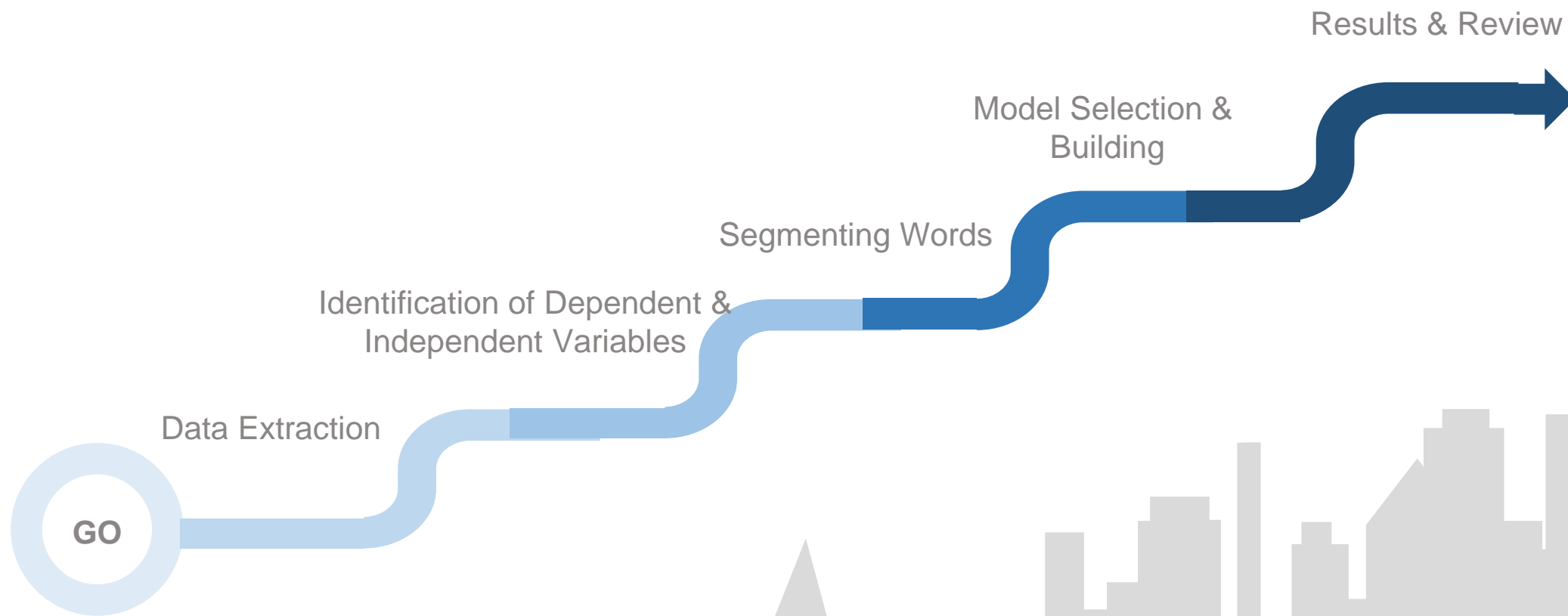
- How to compose a headline that makes it more attractive to video audience?

Group B3

Binxi Wu, Hsieh Chin Hui, Jiantong Yang, Lan Zhong, Linzha Zou, Ruidong Feng, Yiyan Zhu

January 14 2020

Project Timeline



Data Information



400块钱在韩国化妆室能画什么样的妆? |带着妈妈

10.0万 493

阿诺_anuo



我一口吃掉177万播放量的激励金! 世界顶级日本A5神

107.6万 2019-12-08

Terryhimself



史上三观最扭曲的电视剧, 这真的涉及到我的知识

34.3万 2.2万

阿银哦哦



力荐!! 【中英双语】Ballet Beautiful 美丽芭蕾

732.2万 2017-07-20

pll25586

Extract 20k+ data from MAKEUP, FOOD, FITNESS and Funny Videos on bilibili

Extracted information includes videos' titles, creators, published time, # creators' followers, # likes, # views, # collections, # repost, # reposts, etc.

Introduction



Data

Model

Result

Limitation

Models

- **Linear Regression**

Finding out the influences of numbers of sentimental words, numbers of special punctuations, numbers of characters, whether containing question in the titles on view numbers.

- **Neural Network Method**

Using neural network to find out the important elements in each title and then find out the significant elements in the categories.



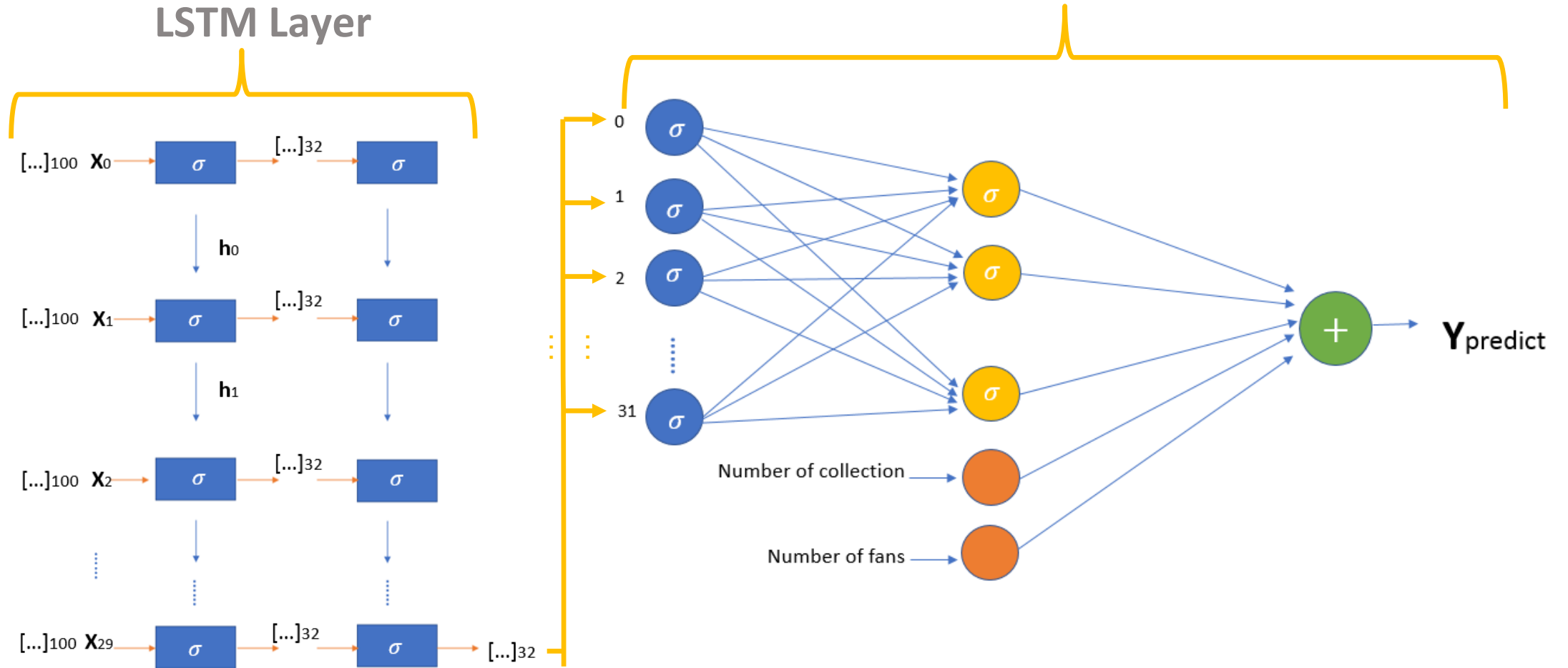
Embedding Model

- Embedding: Embedding unstructured text data to structural vector
- Method: Wor2Vec model: words that share common contexts are located close to one another
- Hyper-parameter:
 - vector dimension: 100
 - context window: 5
 - assign all words a word vector

Source: Mikolov, Tomas; et al. (2013). "Efficient Estimation of Word Representations in Vector Space". [arXiv:1301.3781](https://arxiv.org/abs/1301.3781)



LSTM +



Introduction

Data



Model

Result

Limitation

Neural Network Visualization

Attribution Method: Integrated Gradient

The deeper the color, the higher the word weights

60 元买 108 个肯德基鸡翅尖再不吃这个产品又要下线吧过瘾

E 锅·一起化妆~J 姐蓝血上脸试用+使用心得

【PONY】|2019 年 Coachella 音乐节妆容!

当两个不同信仰信徒在地铁上相遇

空瓶记不是大 up 推荐都好用看我踩哪些雷!

日系女孩|乌龙蜜茶清透妆容分享

【大盒】改变画风! ins 画风灵感妆容日常换头操作

【东北汤唯】超长下饭购物分享家居/彩妆/香水/收纳/酒/日用/锅

如果爸爸态度能 (4) ~ 多样活泼些就好

【胡辣汤】味美汤浓真材实料天天喝都不过瘾

如果不点鸡排在“正新鸡排”里还能买到什么?

牌牌奇——WalkitTalkit

【爱普罗西】自己在家化妆踩大雷|近期晚间护肤分享

【李吉轩】素男变野鸡? 一手画三人! 时尚三姐妹大跳 A 妹 7rings

熊学生画张凳子老师没注意就坐下去结果尴尬!

Introduction

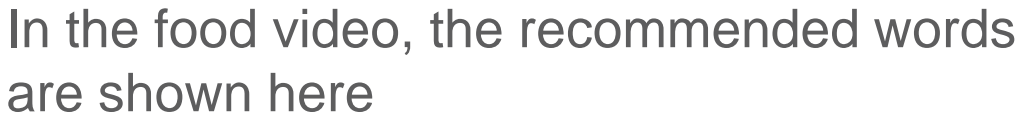
Data

Model



Result

Limitation



Limitations & Future Works

Limitations

- Difficulty in eliminating impacts of other crucial factors on number of views:
—— *Front page, videos duration, system recommend, target audience on different categories, etc.*
- Unable to acquire dynamic data:
—— *Unable to track the trends of data.*
- Invalid linear regression model:
—— *Multilinearity and Endogeneity.*

Future works

- Extracting front page's influence on number of views:
—— *Algorithms like aesthetics scoring, object detection, sentiment analysis, emotion analysis.*
- Topic modeling:
—— *Using topic modeling to find out the popular topics in each categories and giving relevant recommendations to video creators.*
- Applying models to more video categories:
—— *Finding out more popular videos.*

Introduction

Data

Model

Result



Limitation