Using multiple forms of cues and signals to predict crowdfunding success

ZHANG ZiJian
IS Department
30.03.2023

Motivation

Crowdfunding has become an increasingly popular way for entrepreneurs.

- **$900** million,in 2022 U.S. **alone**
- JD Crowdfunding(China), Babyloan(France), Crowdprop(South Africa)

However, the success of a crowdfunding campaign is not guaranteed, and many campaigns fail to reach their funding goals.

- by December 31,2022, the success rate was **37.66%** in kickstarter,**17%-18%** in Indiegogo, **15.29%** in JD Crowdfunding

One potential approach is to use multiple forms of cues and signals to predict crowdfunding success.

- social media activity, campaign description, video content, and other factors

Related work

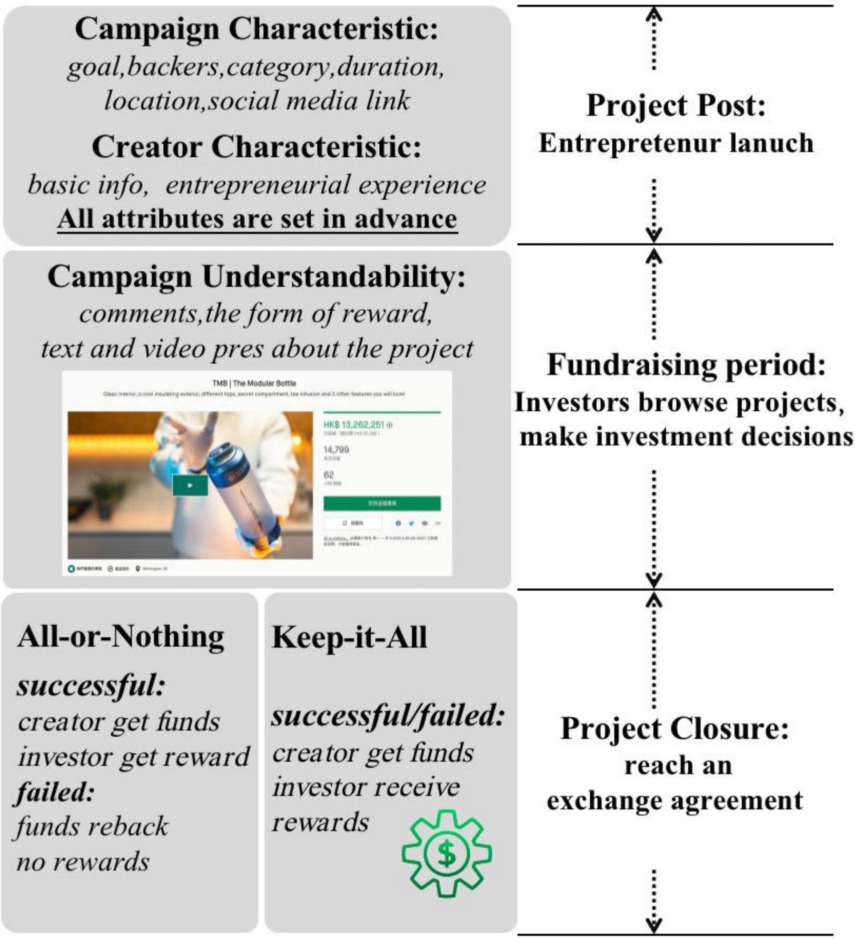
There are previous studies have explored the factors that affect the success of crowdfunding, including **structured factors** such as **funding goal, project category,duration...** as well as **unstructured information** such as **text** and **images** about the project.

Table 1 Summary of studies for prediction in crowdfunding markets

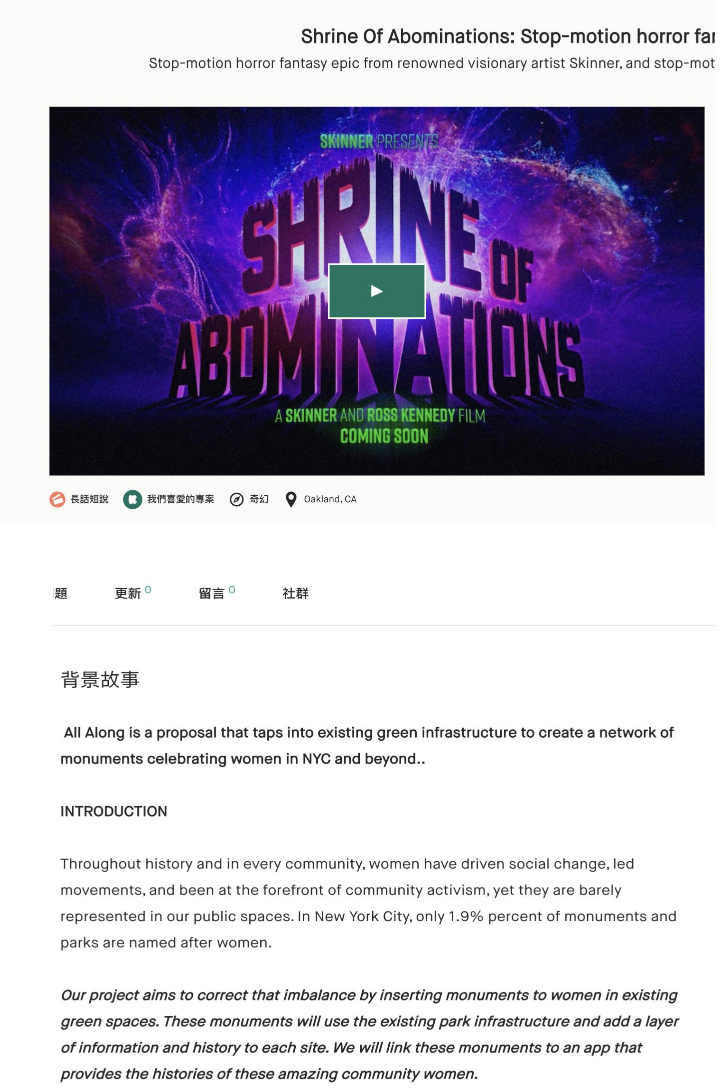
Scholar	Data	Sample Size	Method	Accuracy	Features Form
Greenberg et al. (2013)	Kickstarter	13,000	Random Forest	67.53%	numerical
Mitra and Gilbert (2014)	Kickstarter	45,815	Logistic Regression	58.56%	textual
Desai et al. (2015)	Kickstarter	26,000	Logistic Regression	74.02%	numerical&textual
Etter et al. (2013)	Kickstarter	16,042	SVM	76.22%	numerical
Jermain Kaminski (2020)	Kickstarter	20,188	Neural Network	72.65%	numerical&textual
Zecong Ma (2021)	Kickstarter	652	Logistic Regression	-	Image
Chaoran Cheng (2019)	Kickstarter	18,511	Neural Network	83.26%	Image&textual
Simon J. Blanchard (2023)	Indiegogo	10,487	Bayesian additive tree	**88.80%**	**Image&textual**

Data:cues and signals

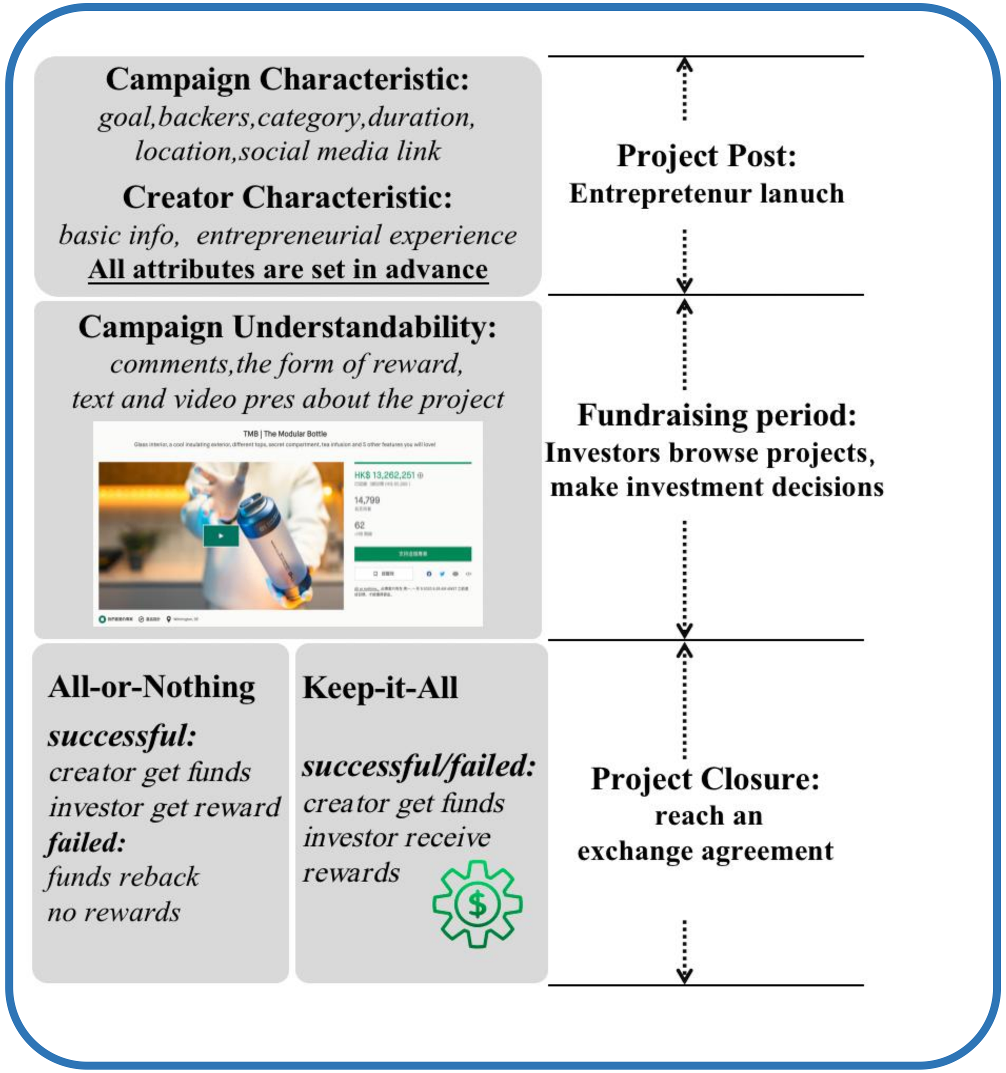
Structured:numberic



Unstructured:textual,video,speech

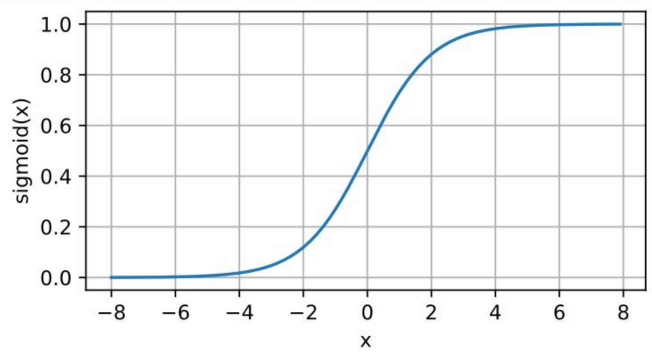


Task 1



The **sigmoid function** g,also known as the logstic function,is defined as follows:

$$\forall z \in R, g(z)=\frac{1}{1+e^{-z}} \in (0,1)$$



For the **logstic regression**,have the folling forms:

$$p(y=1|x;\theta)=\frac{1}{1+exp(-\theta^T x)}=g(\theta^T x)$$

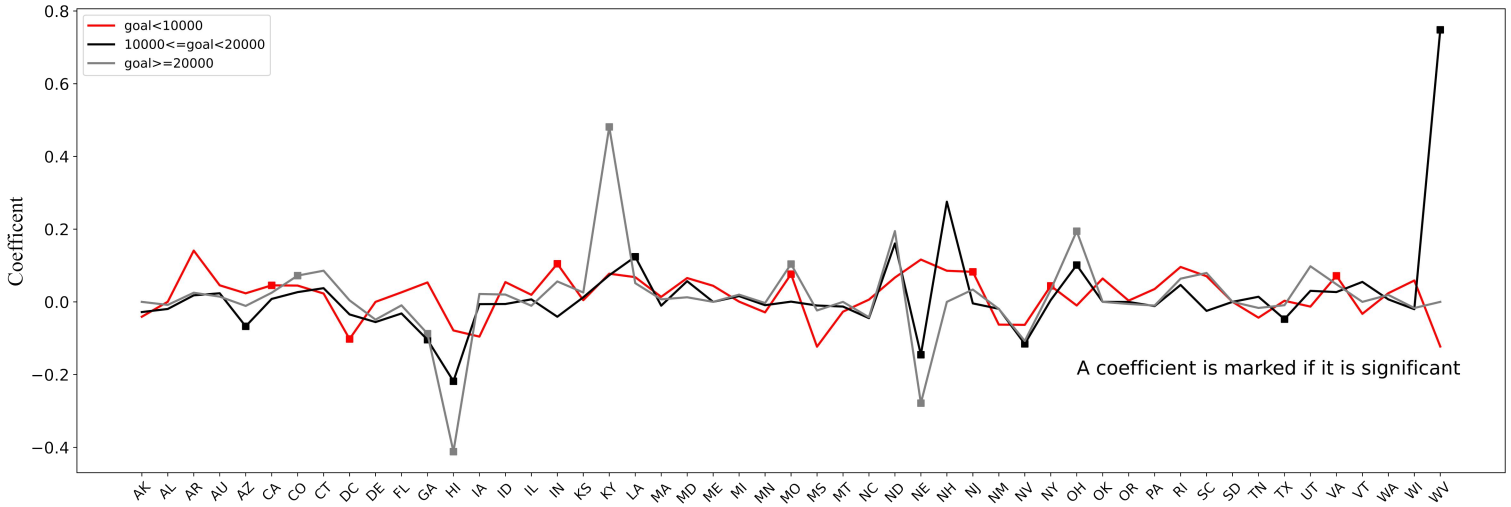
Inputs and regression model

$$Success_i = \alpha + \Sigma \beta_i \times State_i + \gamma \times Control_i + \delta_i$$

Table 2 Summary of variables

Variabletype		Variable	Variabledescription	Mean
Dependent		Success	the outcome of campaign's fundraising	0.727
Independent		State	campaign located	-
Control	Campaign Characteristic	Goal	the target amount of campaign	10013
		Duration	the time from start to finish of the campaign	32.704
		Backers	thenumber of people supporting the project	618.673
		Category	thetype of platform definition belonged	-
	Campaign Understandability	CommentNum	thenumber of commentsaboutthe project	317.843
		PictureNum	thenumber of picturein the overview	11.269
		Video	whethera video in the promotion	0.623
	EntrepreneurCharacteristic	BackerNum	thenumber of projects creatorsupported	57.446
		CreatNum	thenumber of projects creatorcreated	3.759

Regression Results



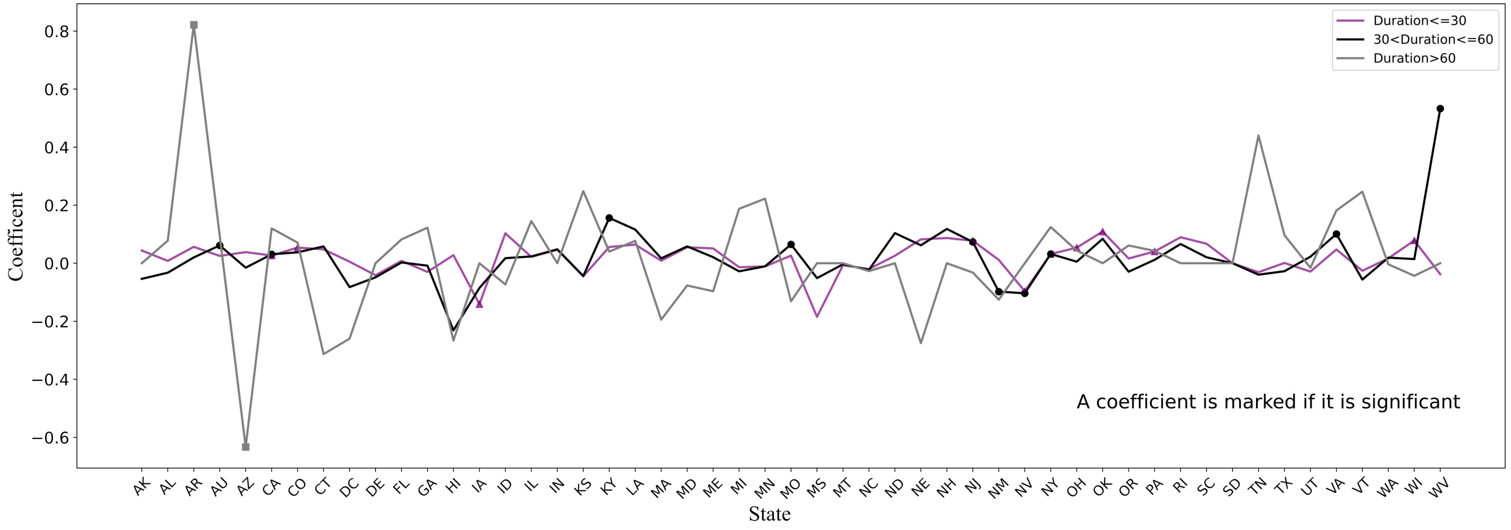
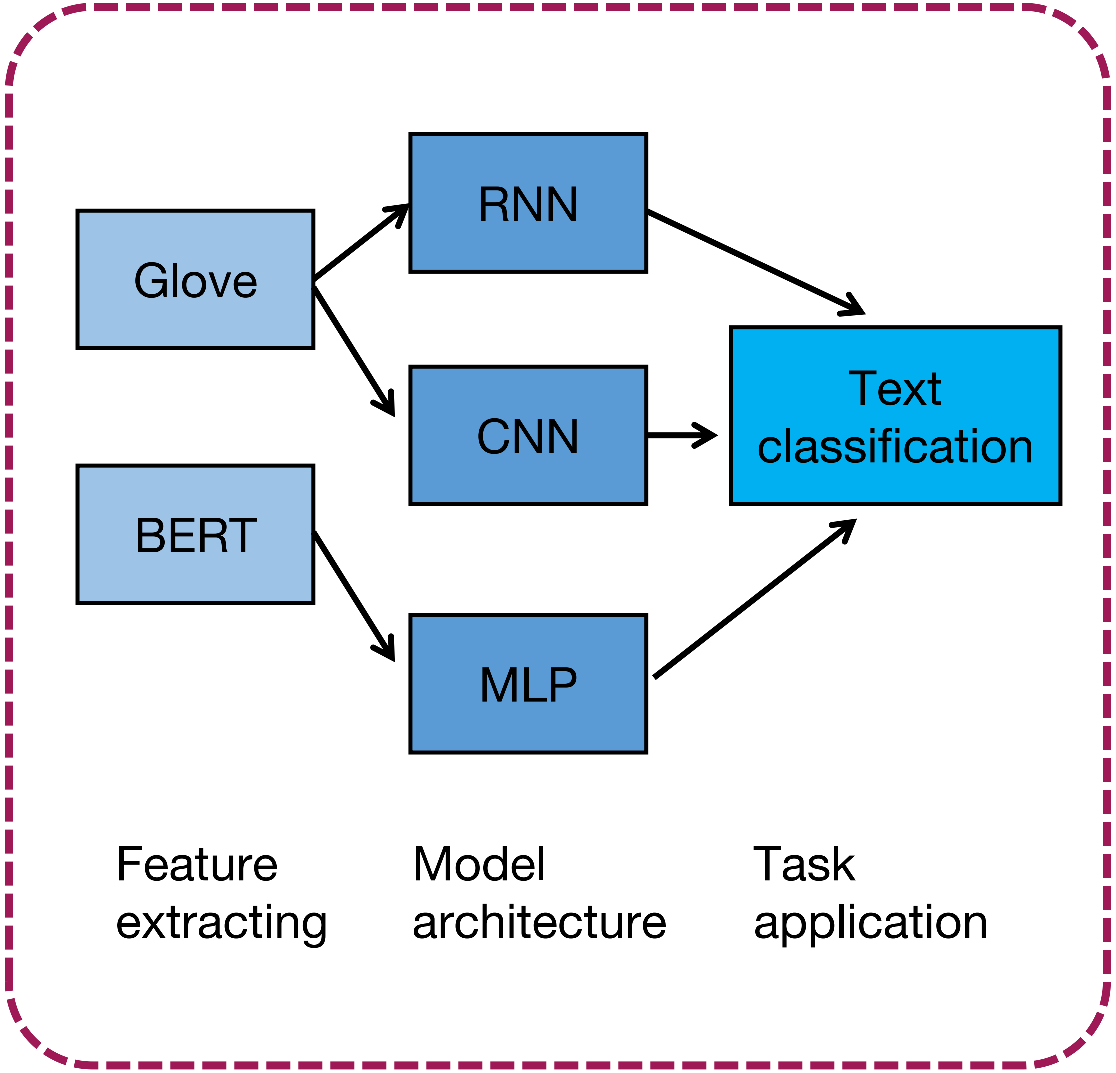
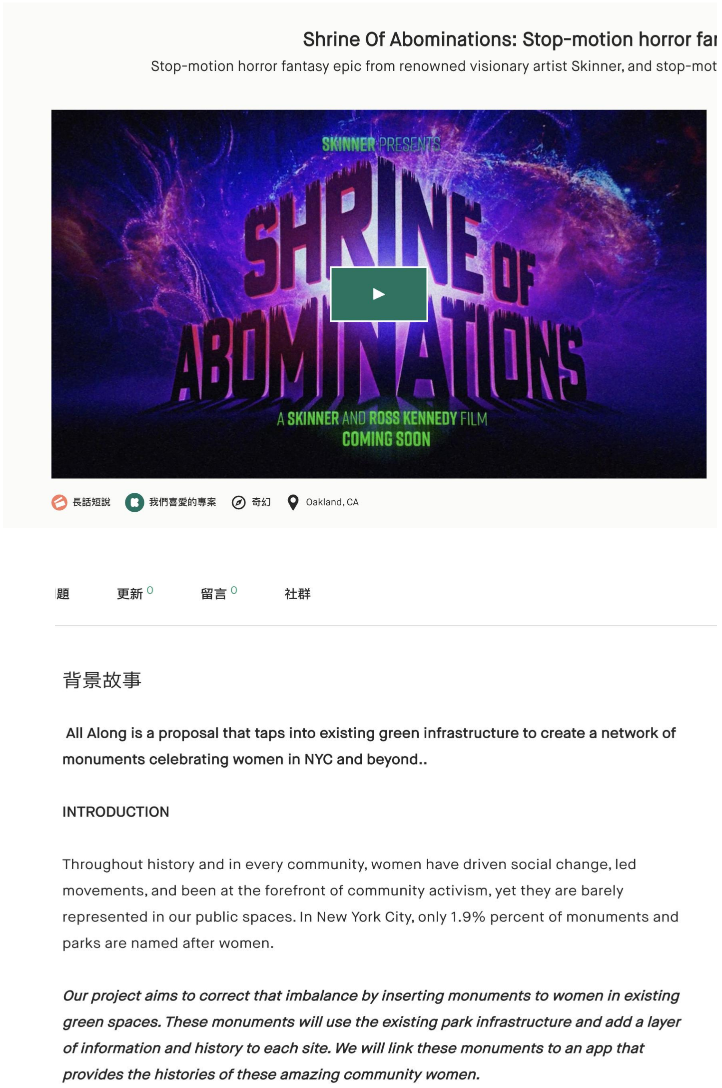


Table 3 Summary of Coef

	Coef	Std Error	T-value
Intercept	**0.9033**	0.014	66.589
Duration	**-0.0023**	0.0001	-7.49
Goal	**-0.0001**	0.0001	-12.258
Backers	**0.0001**	0.0001	1.644
CommentNum	-0.001	0.0002	-0.666
PictureNum	**0.0004**	0.0001	3.009
Video	**0.2023**	0.0008	25.012
BackerNum	**0.001**	-0.001	2.01
CreatNum	**0.0015**	0	4.468
Category	-		
State	-		

Task 2


Glove
BERT
RNN
CNN
MLP
Text classification
Feature extracting
Model architecture
Task application


Shrine Of Abominations: Stop-motion horror fai
Stop-motion horror fantasy epic from renowned visionary artist Skinner, and stop-mot
SKINNER PRESENTS
SHRINE OF ABOMINATIONS
A SKINNER AND ROSS KENNEDY FILM
COMING SOON
長話短說
我們喜愛的專案
奇幻
Oakland, CA
題
更新 0
留言 0
社群
背景故事
All Along is a proposal that taps into existing green infrastructure to create a network of monuments celebrating women in NYC and beyond..
INTRODUCTION
Throughout history and in every community, women have driven social change, led movements, and been at the forefront of community activism, yet they are barely represented in our public spaces. In New York City, only 1.9% percent of monuments and parks are named after women.
Our project aims to correct that imbalance by inserting monuments to women in existing green spaces. These monuments will use the existing park infrastructure and add a layer of information and history to each site. We will link these monuments to an app that provides the histories of these amazing community women.

Inpus: text, speech, video

Table 4 Descriptive statistics of sample

Source	count	positive share
text	10396	55.18%
speech	160	62.5%
video	160	62.5%

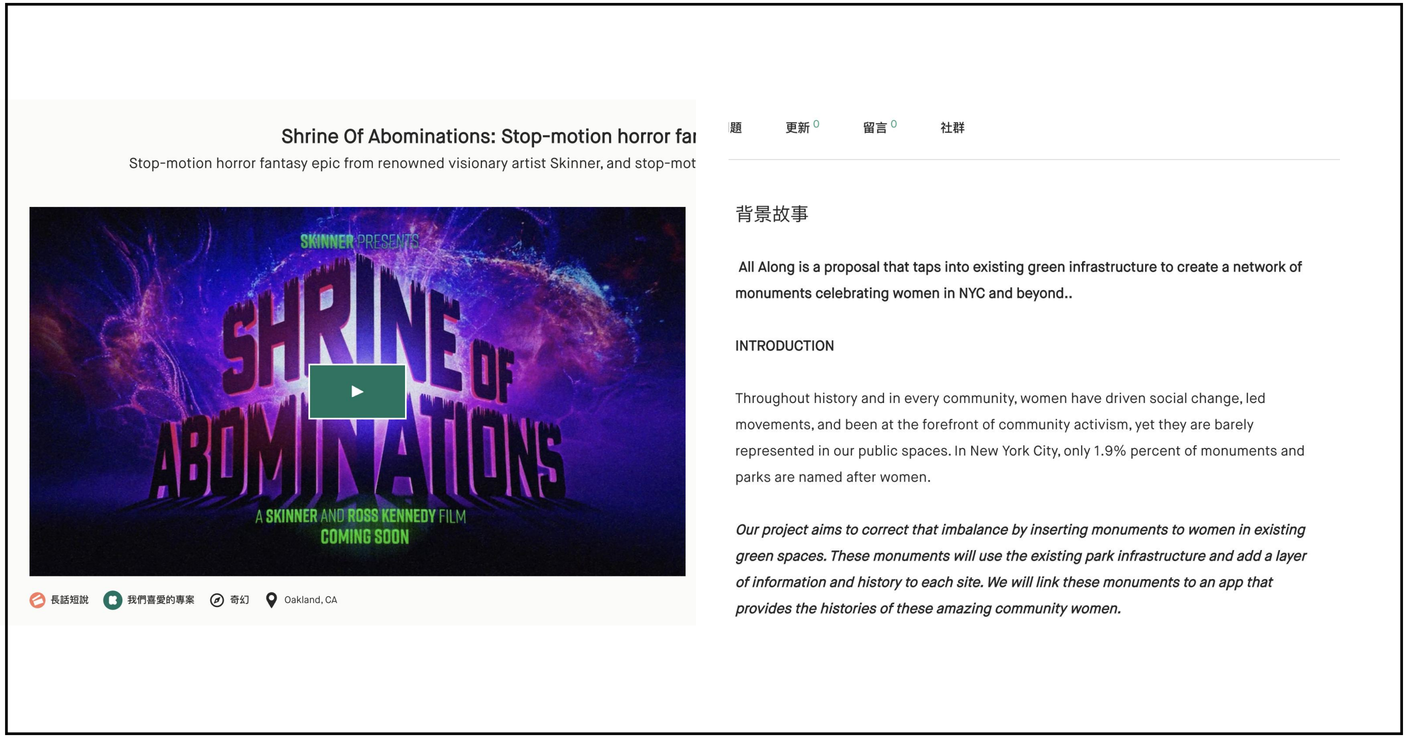


Table 5 Sample content

(A)Text input

"Hello Kickstarter Family. We are excited to bring you our story of Do Dah Dolls™We have setup this Kickstarter because we are so close! We have financed 90% of this project ourselves and only need the last 10% from you.(...)"

(B)Speech input

["We will match books from authors to Industry professionals, based on the obvious elements as well as some of the unspoken rules that get most authors rejected. It's not enough anymore(...) "] [0-26.2s] [Confidence: 0.78]

(C)Video input

['black and white,0-2.2s','style,0-10.2s','clip art,3.2s-7.8s','graphics,4.7s-20.2s','display device5.2s-22.5s', 'technology,6.2s-33.8s','food,12.3s-27.9s', 'car,18.3s-29.8s', (...)]

custom-build python

google cloud video Intelligence API

Preprocessing

Table 6 Descriptive statistics of preprocessed-sample

Source	Mean token	Vocab	Seq length
text	734.52	56246	**1024**
speech	205.46	6002	512
video	100.34	1326	256

Table 7 Sample content

(1) Tokenize('word')

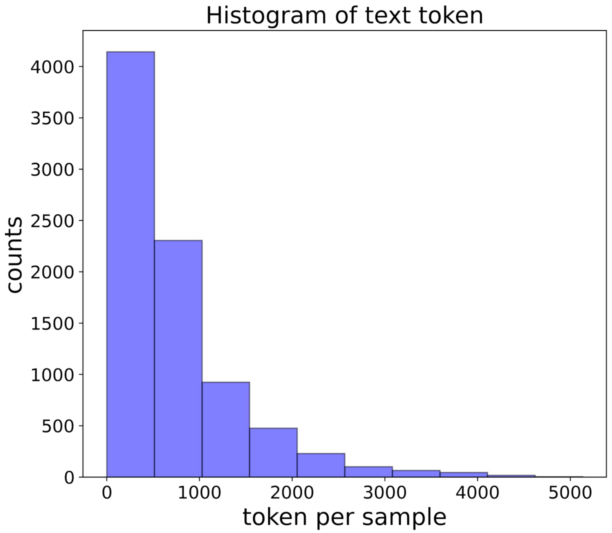
“['Hello','Kickstarter','Family','We','are','excited','to','bring','you','our','story','Do','Dah','Dolls™We','have','setup','this',Kickstarter',(...)”

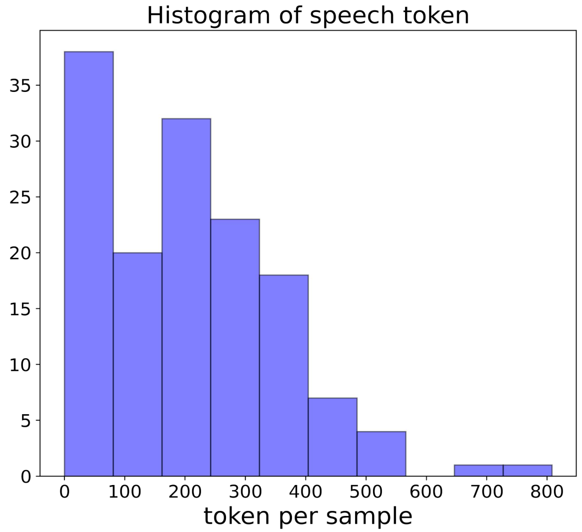
(2)Build a vocab(filter<5,stopwords)

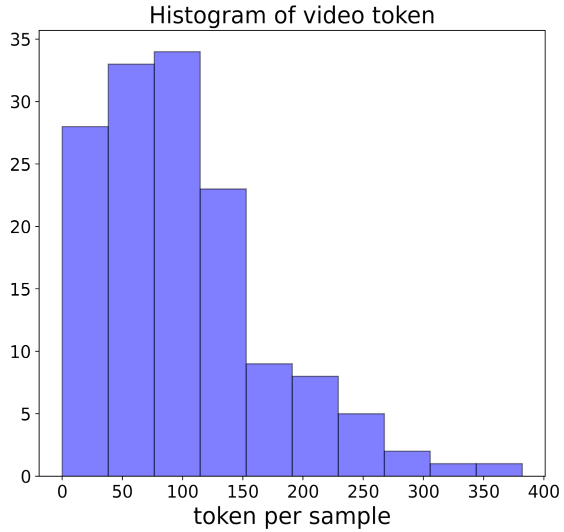
[3071, 70, 42877, 38, 17, 499, 4, 165, 11, 21, 167, 5, 1262, 0, 0, 24, 3460, 22, 70, 178, 18, 17, 49,(...)]

(3)padding&truncate()

[[3071, 70, 42877, 38, 17, 499, 4, 165, 11, 21, 167, 5, 1262, 0, 0, 24, 3460, 22, 70, 178, 18, 17, 49, ..., 1, 1, 1]]







Vectorization

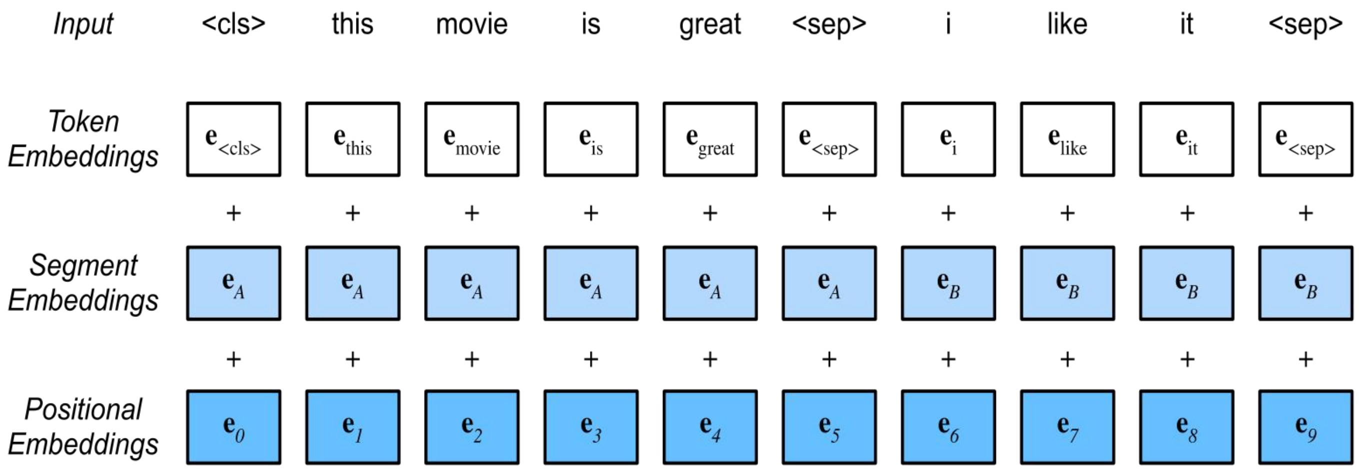
Table 8 Descriptive statistics of preprocessed-sample

Methods	embedding	feature- extraction
TextCNN	100+100	Glove
BiRNN	100	Glove
BERT+MLP	768	Bert

$$w_i \cdot w_j = \log P(i|j)$$

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

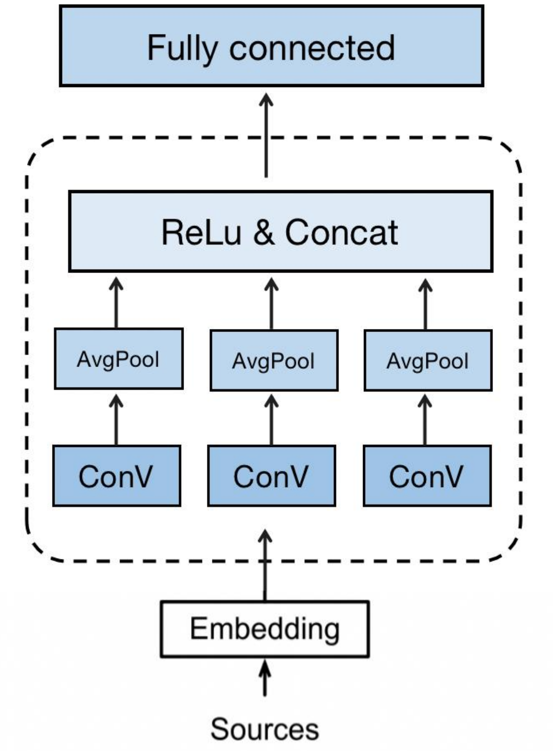
Glove embedding



Bert embedding

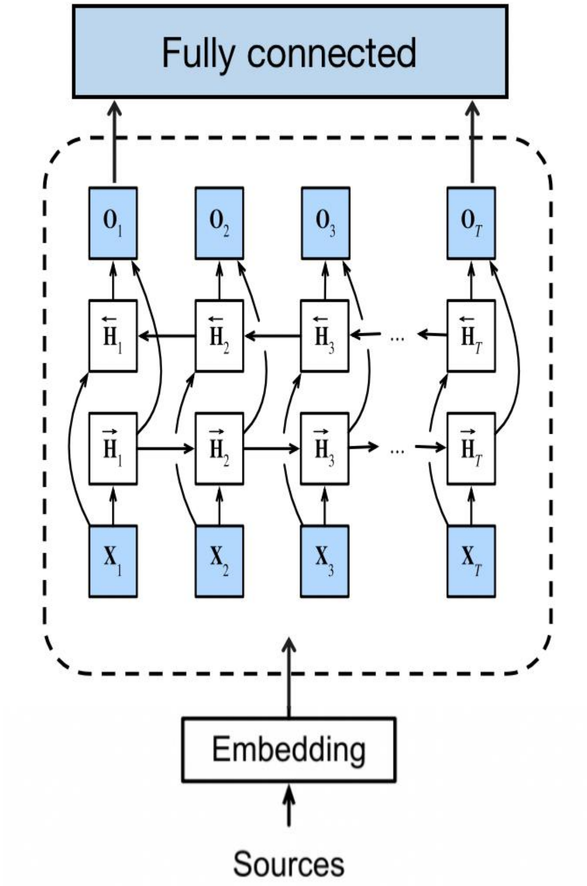
Models

TextCNN



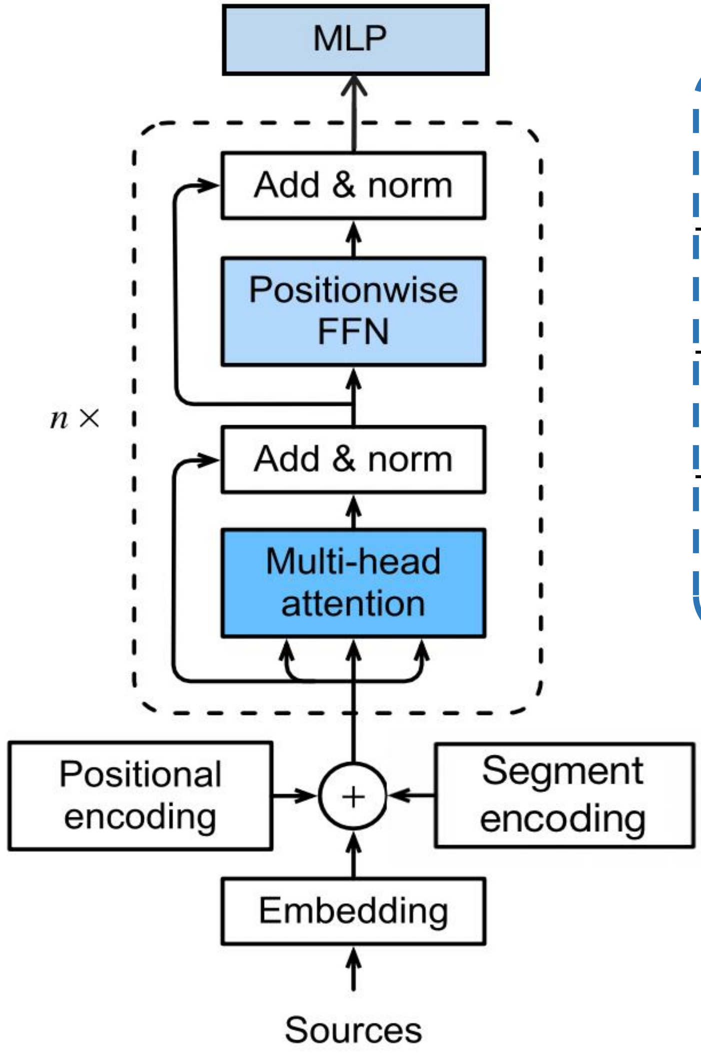
kernel_ sizes	num_ channels
5,6,7	100,100,100

BiRNN



num_ hidden	num_ layers
100	2

BERT+MLP



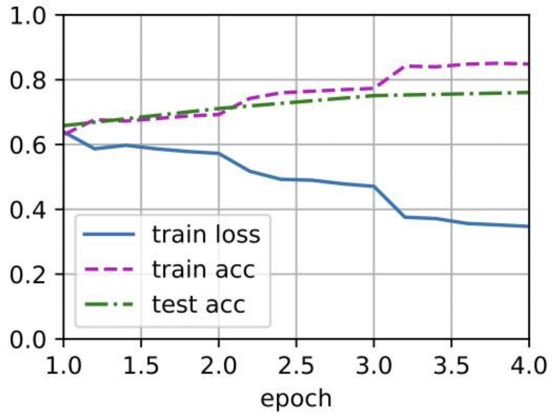
num_ hiddens	768
ffn_num_ hiddens	3072
num_ heads	12
num_ layers	12

Evalution

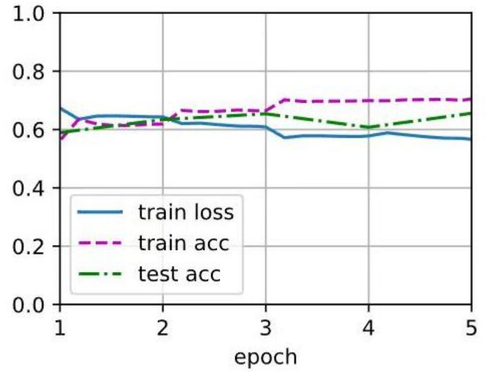
Table 9 Summary of evalution

Inputs	Classifier	Accuracy	support
Text	TextCNN	**0.761**	10147
	BiRNN	0.656	10147
	BERT+MLP	0.735	10147
Speech	TextCNN	0.545	160
	BiRNN	0.532	160
	BERT+MLP	**0.588**	160
Video	TextCNN	0.521	160
	BiRNN	**0.577**	160
	BERT+MLP	0.534	160

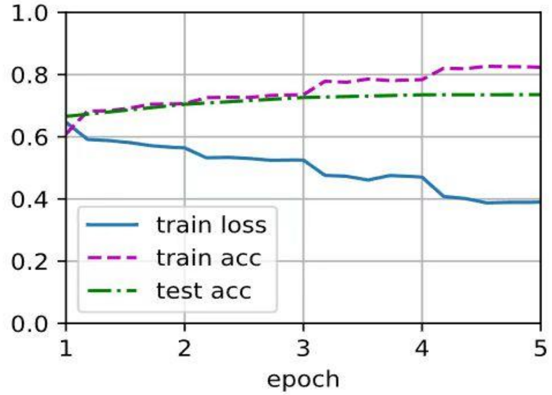
```
loss 0.347, train acc 0.848, test acc 0.761
154.3 examples/sec on [device(type='mps')]
```




```
loss 0.566, train acc 0.705, test acc 0.656
2009.4 examples/sec on [device(type='cuda', i
ice(type='cuda', index=2), device(type='cuda'
device(type='cuda', index=5)]
```




```
loss 0.390, train acc 0.823, test acc 0.735
191.6 examples/sec on [device(type='cuda', in
ce(type='cuda', index=2), device(type='cuda',
evice(type='cuda', index=5)]
```



Thank you

Much thanks to Dr.Raymond and Jiayue,runtao