

SEINFELD

SCRIPT GENERATOR

Cloudy Liu

Problem Statement

Audience:
Marketing & Operation Professional

Build a script generator
to showcase how text
generation can create
content for marketing
initiatives faster, better
and cheaper.

Questions Breakdown:

- WHAT to use?
- WHEN to use?
- HOW MUCH time does it take to use?
- HOW MUCH money does it cost to use?
- HOW WELL it performs?

Data

- Scraped: 54,211 lines
- After data cleaning: 44,661 lines
- Data Size: 4.2MB, .txt
- 3.39 Million characters

GEORGE: We can't just let him get away with this.

JERRY: Do you realize how crazy he had to be to do something like this? He knew I was gonna know it's missing, and he took it! And of all things to take - I left my watch, tape recorder, stereo. He's crazy.

KRAMER: You want to go get him?

ELAINE: Well, then, if he's crazy you should just forget it.

GEORGE: Forget it? I already called my parents. I told them to expect the surprise of a lifetime. My mother's making her roasted potatoes!

ELAINE: George, do you realize that Rava's asked me to edit her book?

GEORGE: Who is this Rava?

KRAMER: I say we get him.

ELAINE: No!

GEORGE: Let me just call him.

(Jerry's apartment, Jerry is at the table and Elaine is on the phone)

ELAINE


(to the phone) Well did he bring it up in the meeting?

(Jerry picks up a yellow shirt and walks to Elaine)

JERRY

Elaine, see this T-shirts, six years I've had this T-shirts, it's my

best one, I call him...Golden Boy

**IMSDb**
The web's largest
movie script resource!

Search IMSDbGo!

AlphabeticalA B C D E F G H I J K L M N O P Q R S T U V W X Y Z

The Internet Movie Script Database (IMSDb)

Seinfeld Transcripts

Series 1

[Good News, Bad News](#) (1989-07-05)
Written by Larry David, Jerry Seinfeld

[The Stakeout](#) (1990-05-31)
Written by Larry David, Jerry Seinfeld

[The Robbery](#) (1990-06-07)
Written by Matt Goldman

[Male Unbonding](#) (1990-06-14)
Written by Larry David, Jerry Seinfeld

[The Stock Tip](#) (1990-06-21)
Written by Larry David, Jerry Seinfeld

Series 2

[The Ex-Girlfriend](#) (1991-01-23)
Written by Larry David, Jerry Seinfeld

[The Pony Remark](#) (1991-01-30)
Written by Larry David, Jerry Seinfeld

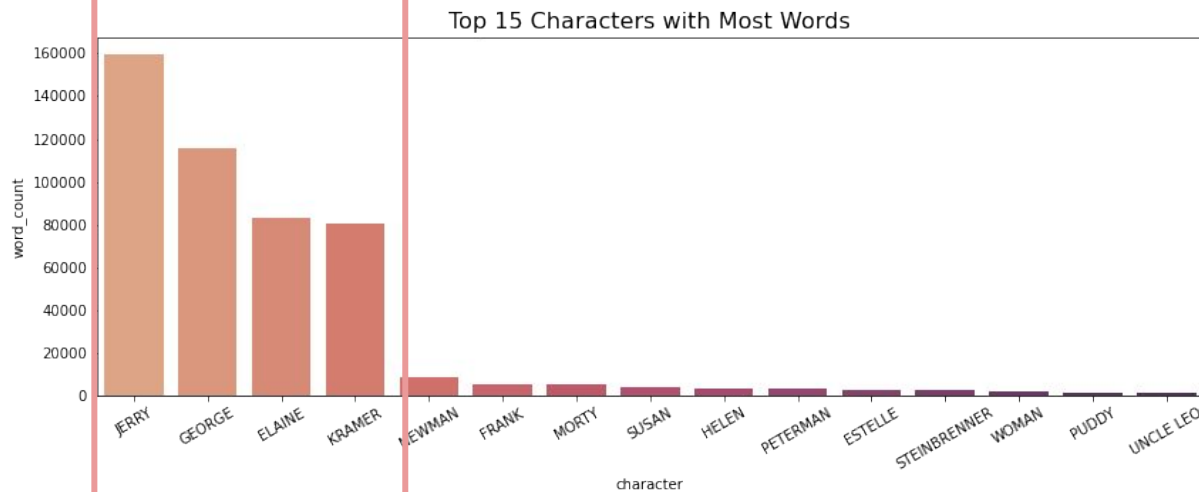
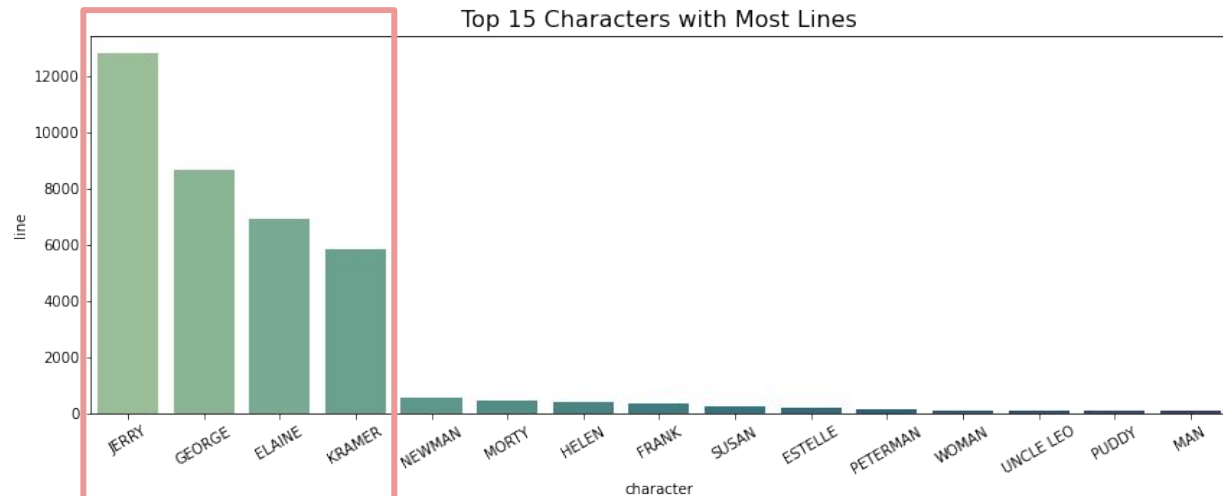
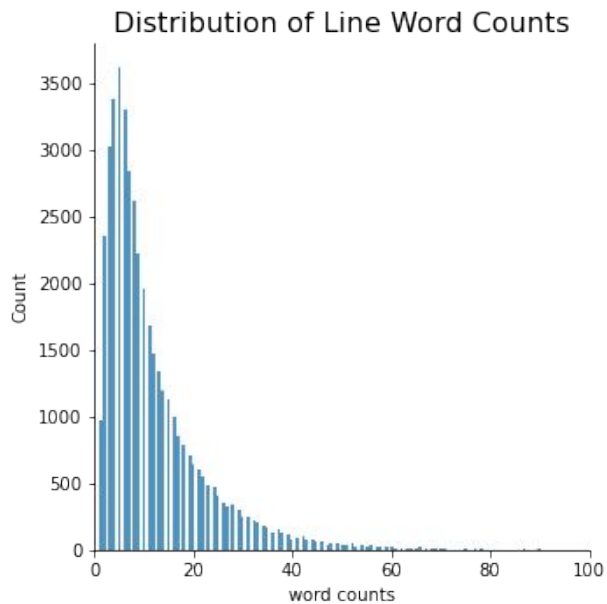
[The Jacket](#) (1991-02-06)
Written by Larry David, Jerry Seinfeld

[The Phone Message](#) (1991-02-13)
Written by Larry David, Jerry Seinfeld

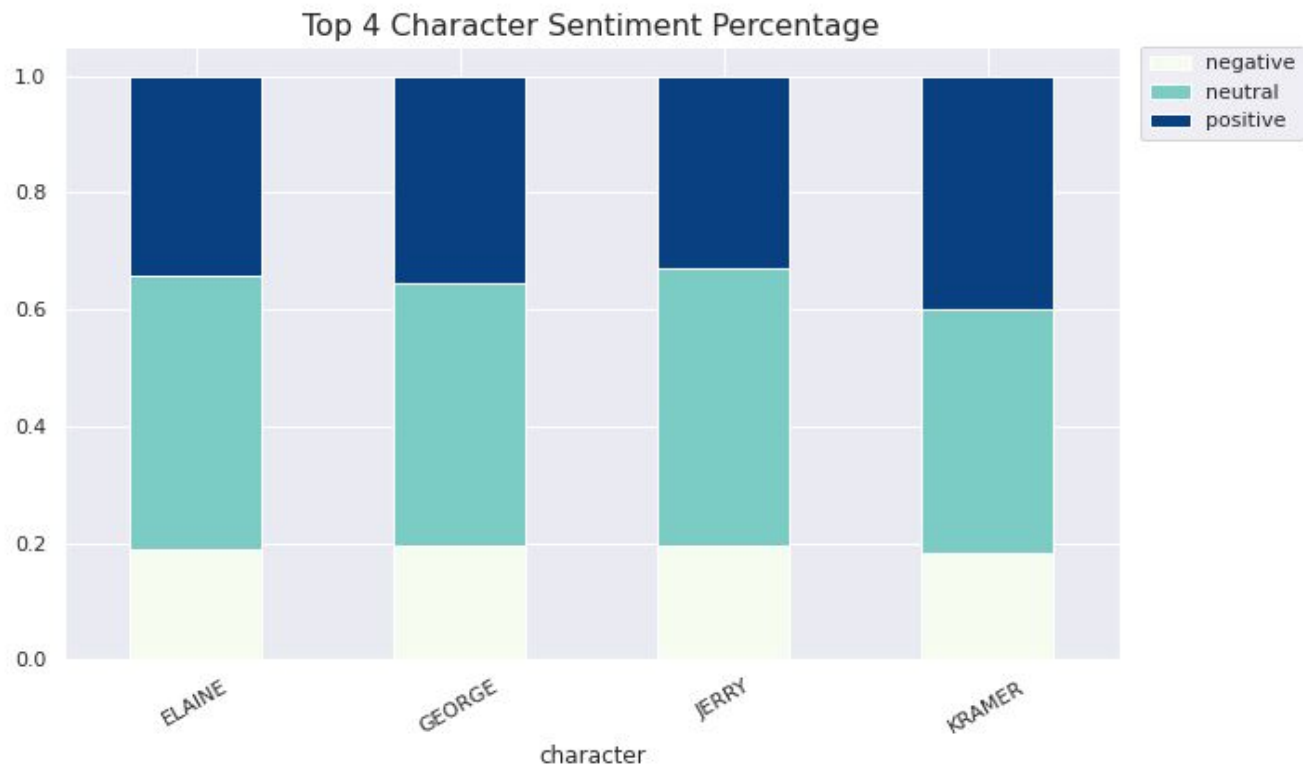
[The Apartment](#) (1991-04-04)
Written by Peter Mehlman

[The Statue](#) (1991-04-11)
Written by Larry Charles

EDA



Sentiment Analysis



Word Clouds



Jerry



George







Elaine



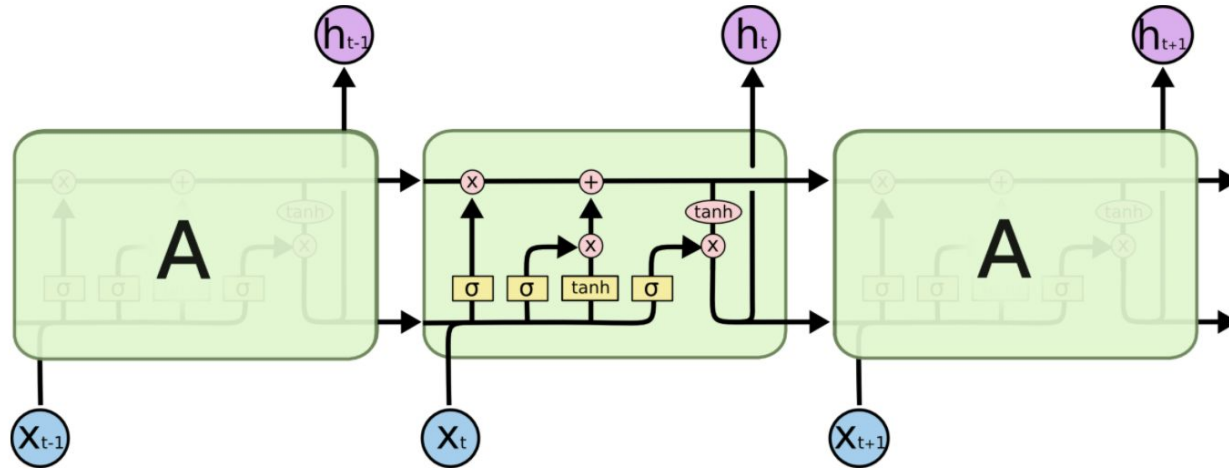
Kramer

Most Common Words/ Phrases

	Jerry 	George 	Elaine 	Kramer 
Interjection yeah, hey, oh, oh yeah, Hey hey, oh hi, yeah yeah, yeah right, yada yada yada		ha ha	ha ha uh uh	uh uh ho ho
Negation don know, don want, don think	don understand don care	don understand don like	don understand	don like
Character Name Jerry, George, Elaine, Kreamer	Uncle Leo		Mr. Peterman Mr. Pitt	Newman
Meaningful get out	coffee shop wait second	coffee shop wait minute	coffee shop wait minute	coffee table

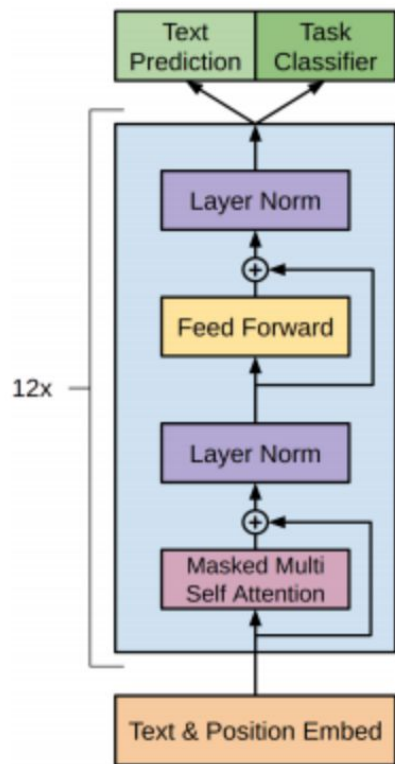
Nope. One size does NOT fit all.

Text Generation Model -- RNN-LSTM

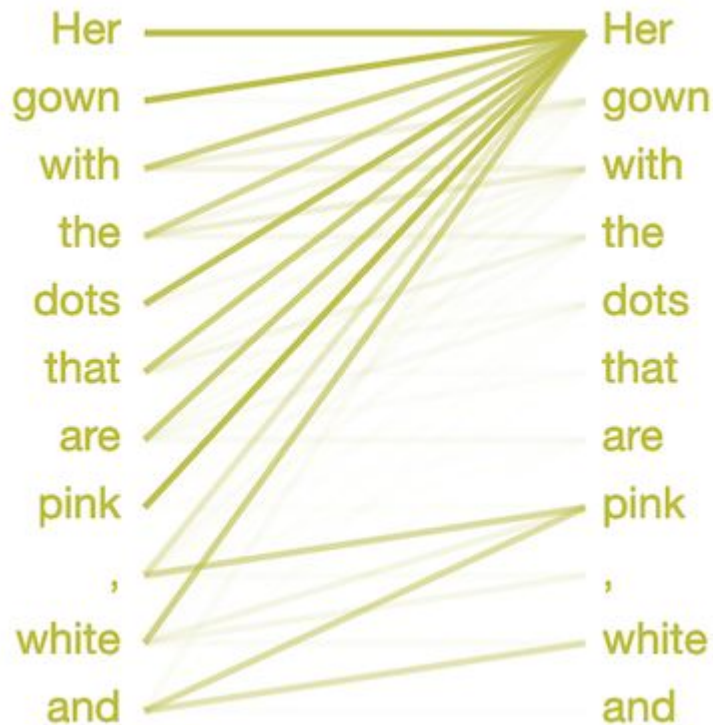


Memory mechanism
Word-by-word

Text Generation Model -- GPT-2



Attention mechanism
Chunk-by-chunk
Transfer learning



RNN-LSTM Model



- Generator: resolve hardware limitation



- Diversity: specify creativity level of the script



- Seed text: random or specified character



- Hours trained: 3.6h



- Speed to generate: 12s/200 words



- Cost: <\$150 with nearly non-stop GCloud usage for 2 weeks

RNN Generated Script

“

JERRY: I'm comfortable with your friend Kramer.

KRAMER: Oh, he's a good friend.

JERRY: This is an emergency.

GEORGE: I think he's a friend of mine. He's not at the party.

JERRY: You're not getting the good job?

GEORGE: No, I'm fine.

JERRY: How much is it?

GEORGE: Six.

JERRY: The little black?

GEORGE: Yeah, sure. It's a. It's a big thing. Can you believe it?

JERRY: I'll pay you later.

GEORGE: (laughs) Oh, really, it's fantastic.

”



Sometimes friendly

RNN Generated Script

“

GEORGE: You have a present?

ESTELLE: It's a hundred dollars.

ESTELLE: The Cadillac with the Seinfelds.

GEORGE: A little!

ESTELLE: (in front of the the-the white) Oh, the that's the!

ESTELLE: What is that?

FRANK: That's with the it! The key comes from the living room. It just happens to it! It's just for me!

ESTELLE: The Seinfelds couldn't hide it! This was the elevator!

ESTELLE: All right, I'll call you.

FRANK: (shouting) Elaine, where the hell is your son?



Mostly dramatic

”

RNN Generated Script

“

GEORGE: Man they ordered the **Mets!**
who's your offence?

KRAMER: Well he was it. You know he has
the **Yankees!**

JERRY: Really? Most bucks?

KRAMER: (impressed) Oh, sure! They put
some nuts. I told you. **They're out..**



Sometimes (tries to) talk
about sports

”

RNN Generated Script

“

GEORGE: Indeed. Oh, this is the fortress of solitude. Can i borrow an ice form? Uh huh. Really? Uh, excuse me. Are you coming with me? Over fifty. Four hundred the looks for a big home. I'll see ya guys. Doesn't you see someone else together. Six times together. He was in a ruining love. They want to be Steinbrenner.



”

“



KRAMER: Or one day Jerry has dropping. They're you ladies, as they are still eighty. With one of your friends at Jerry's couch, Look at him. Kramer is down the road with the racquet. He gives out the lot of choice.

”

Sometimes poetic

GPT-2



- Pre-trained: with 124M and 500MB in size



- Hours trained: 2h



- Speed to generate: 40s/200 words



- Cost: \$10 (basically free)

Seinfeld Script Generator

Who to start

How Long
200

50 1000

how creative
0.90

0.10 2.00

Generate!

See some past examples?

Jerry ▼

Go!

Seinfeld Script Generator

by Cloudy Liu



Conclusion

WHAT to use?

WHEN to use?

HOW MUCH time does it take to use?

HOW MUCH money does it cost to use?

HOW WELL it performs?

Data size is the key

- <1MB → GPT-2
- 1-6MB → GPT-2/RNN
- >6MB → RNN

Know what you want

- Consistency with original data → RNN
- Ready to use → GPT-2
- Inspiration → GPT-2/RNN
- Long → GPT-2 / Short & abstract → RNN

Time

- Faster train → GPT-2
- Faster generate → RNN

Cost

- Both highly dependant on GPU but the cost is significantly lower than other options

Performance

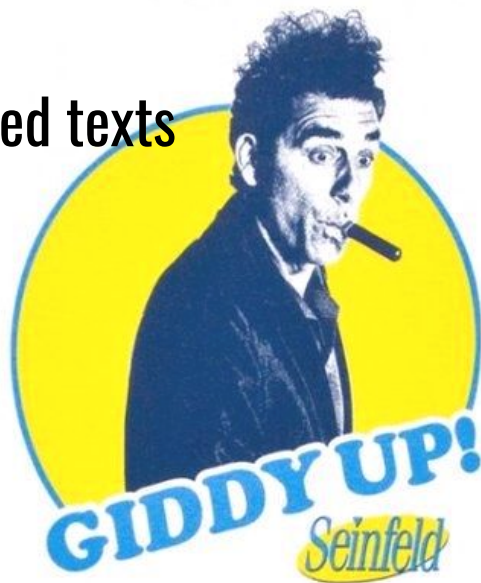
- RNN → Flexibility/ Room to improve
- GPT-2 → As is

Limitations

- Data not clean enough, adding noises to the modeling
- RNN-LSTM model had a hard time reducing loss function
- Additional time/efforts on fine-tuning RNN model needed
- Highly dependant on cloud computing
- Slow and subjective evaluation
- Internet-trained models have internet-scale biases.
- Unable to deploy API due to dependency version conflicts

Next Steps

- Incorporate attention mechanism to RNN-LSTM model
- Train the existing model with jokes datasets
- Improve the API to be more user-friendly
- Build an automated way to evaluate the generated texts



THANK YOU!