Love Advice Bot
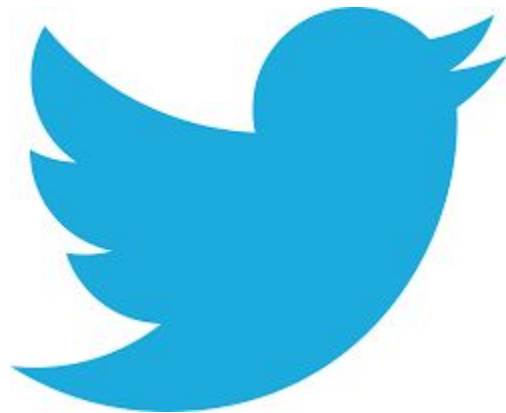
Nothing says romantic expert
like a bot relying on Reddit

Terry S.

Starting Off- FirstDateBot

- Generating questions to ask on first date.
- Simple markov chains on a corpus of conversation starting questions

Love Bot @FirstDateBot · Oct 24
What do you have? #FirstDate #ImportantQuestions #MustKnow

Next Steps

- Have the user provide the questions instead of the bot
- Have the bot provide the answers for the questions from the users.

Love Advice Bot- Tech Overview

- PRAW - Python Reddit API Wrapper
  - r/relationships - corpus for algorithms
- LDA - Latent Dirichlet Allocation (Classification Algorithm)
- Markovify - Text Generation
- Gensim- provides LDA, stemming, lemmatizing

Corpus

- Taken from r/relationship
- Utilized 375 of top 1000 posts of all time
- Main filter was removing update posts from corpus
- Grabbed top 5 comments from each of the posts

- Fair Warning: Some of the output of project is kinda explicit, result of relying on reddit for information

Some examples

Q: There's this person in my school that's really sweet. They always say hi to me and ask me for help on work. I think maybe there's something there. How should I ask this person out without being weird?

A: It can imply a life of leisure, or a statement, that makes me uncomfortable. It can imply a life we lived, that someone loves you, and do whats right for you. Then he tries to build that bridge for him.

More Samples (this one's taken from r/relationships)

Q: Title: I [19F] feel numb in my relationship. He's [19M] the sweetest, best guy I know, and I don't feel anything

A: Tell him marriage and how to move on. I had before turned into an apartment, after you. This likely isnt going to do the deed so he has pre warned you that he go to college and coming out with her.

A: Then he went on to read too much from you for not going to baptise her child. especially when he was a lot right now. Let him work on creating the causes and conditions for you or his son, never go away.

How it Works - Preprocess

- After getting data from reddit, need to preprocess it for LDA.
- Removing links, tl;dr, and edit messages.
- Lemmatize and Stem
- Tokenize - turning passage into list of individual words

Playing ⟶ Play
Plays ⟶ Play   } Common root form 'play'
Played ⟶ Play

am, are, is ⟶ be

Car cars, car's, cars' ⟶ car

Using above mapping a sentence could be normalized as follows:

the boy's cars are different colors ⟶ the boy car be differ color
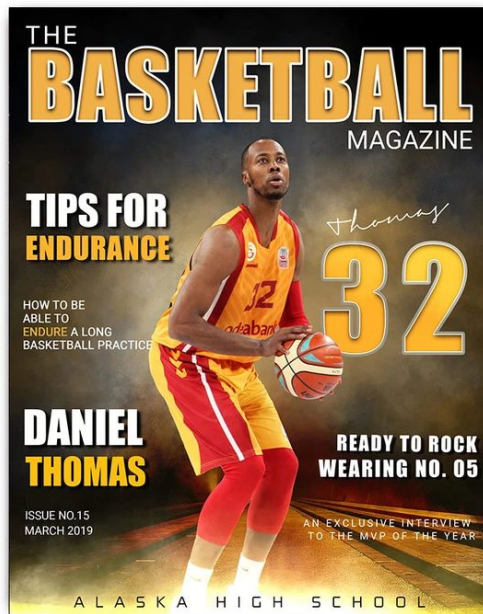
Preprocess sample

Original:

Stemming is the process of reducing inflection in words to their root forms such as mapping a group of words to the same stem even if the stem itself is not a valid word in the Language.

Processed:

['stem', 'process', 'reduc', 'inflect', 'word', 'root', 'form', 'map', 'group', 'word', 'stem', 'stem', 'valid', 'word', 'languag']

How it Works - LDA

- Topics
- Documents
- Terms
- Magazine example

How it Works - LDA

- Using the processed data, make dictionary of entire training data
  - Number of tokens = 14826735
  - Unique tokens = 5446
- Dictionary → Bag of Words
- Bag of Words → LDA Model
- Set parameters, like num_topics (10 in this case)
- Example:

Ori: We are as solid as I guess anyone thinks they are before something like this blindsides them.

Pro: ['solid', 'guess', 'think', 'like', 'blindsid']

BoW: [(22, 1), (62, 1), (99, 1), (177, 1), (193, 1)]

LDA: [(0, 0.8498824), (1, 0.016679173), (2, 0.016678764), (3, 0.016678385), (4, 0.016680282), (5, 0.016679173), (6, 0.016680297), (7, 0.016680004), (8, 0.01668089), (9, 0.016680598)]

LDA Topics

```
Topic: 0
Words: 0.017*"know" + 0.014*"go" + 0.012*"tell" + 0.012*"like" + 0.012*"want" + 0.011*"say" + 0.010*"think" + 0.010*"year" + 0.009*"time" + 0.009*"friend"
Topic: 1
Words: 0.016*"want" + 0.012*"like" + 0.010*"thing" + 0.010*"time" + 0.010*"say" + 0.010*"wife" + 0.010*"tell" + 0.010*"know" + 0.010*"go" + 0.009*"feel"
Topic: 2
Words: 0.013*"know" + 0.011*"time" + 0.011*"come" + 0.010*"like" + 0.009*"go" + 0.008*"year" + 0.008*"tell" + 0.008*"think" + 0.007*"talk" + 0.007*"friend"
Topic: 3
Words: 0.009*"go" + 0.009*"think" + 0.008*"year" + 0.007*"know" + 0.007*"tell" + 0.006*"like" + 0.006*"birthday" + 0.006*"time" + 0.005*"thing" + 0.005*"say"
Topic: 4
Words: 0.015*"work" + 0.013*"say" + 0.013*"go" + 0.012*"know" + 0.012*"time" + 0.011*"year" + 0.011*"tell" + 0.011*"friend" + 0.011*"want" + 0.010*"think"
Topic: 5
Words: 0.015*"tell" + 0.014*"go" + 0.012*"want" + 0.012*"come" + 0.011*"say" + 0.011*"like" + 0.010*"work" + 0.010*"know" + 0.009*"friend" + 0.009*"sister"
Topic: 6
Words: 0.019*"tell" + 0.017*"like" + 0.017*"say" + 0.016*"want" + 0.015*"know" + 0.012*"think" + 0.011*"time" + 0.011*"go" + 0.010*"year" + 0.010*"feel"
Topic: 7
Words: 0.017*"say" + 0.017*"like" + 0.011*"know" + 0.011*"want" + 0.009*"tell" + 0.008*"go" + 0.008*"ask" + 0.008*"think" + 0.007*"friend" + 0.007*"thing"
Topic: 8
Words: 0.018*"want" + 0.013*"say" + 0.013*"know" + 0.013*"famili" + 0.011*"year" + 0.011*"tell" + 0.011*"sister" + 0.010*"like" + 0.010*"time" + 0.010*"think"
Topic: 9
Words: 0.019*"friend" + 0.018*"tell" + 0.013*"like" + 0.011*"go" + 0.010*"say" + 0.010*"want" + 0.009*"talk" + 0.008*"know" + 0.008*"think" + 0.008*"relationship"
```

How it Works - Markovify

- Markov Chain for each topic of the LDA model

- Corpus of each chain is comments from reddit posts

- Comments sorted into the topics based on percentages given by LDA

How it Works- Generating Advice

- Accept user story/question.
  - Process same way as before
  - Run it through LDA to get topic spread
- Combine topic Markov Chains together with weights given by LDA
- New Markov Chain → 3 sentences
- Output sentences

Demo Time

so I met this guy at a party and we hit it off pretty well, talked the whole night. We went on a date a few weeks ago and it went well, lasted quite a while and I think we both had a good time.

He's been really busy at work but I've also since dropped multiple hints -- told him I had a good time, asked him to hang out again, said we should get food... the first one, he didn't respond to and the second two he responded and said he was down but never set a date/initiated anything.

We're still talking everyday via message and are having quite long conversations, but I'm getting to the point where I feel like he's not interested/don't know where he's at so I don't want to put in the effort anymore.
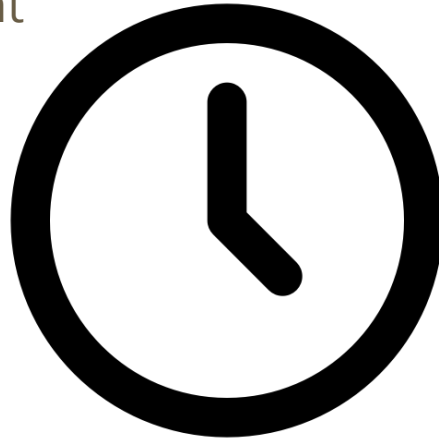
We have a mutual friend between us, and the mutual friend talked to him and apparently he said he's still "thinking about us"... which I'm not sure what that means.

Should I wait for him to initiate or ask him how he feels so I can get this cleared up?

TL;DR: went on a date with a guy, it went well, but he hasn't initiated anything after the first date after I've dropped multiple hints. Getting to the point where I kind of want to give up because it feels like he doesn't care. Should I stop trying/wait for him to initiate, or ask him how he feels?

Problem Areas

- Spellchecking

- Processing Data

- Time Management

Where to next?

- Explore the parameters of LDA more, see how more topics affects model

- Find a bigger corpus

- Grammar Scorer for output text

- Explore other text generation algorithms that benefit from topic analysis

Questions?

Resources

- https://www.datacamp.com/community/tutorials/stemming-lemmatization-python

- https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24

- https://praw.readthedocs.io/en/latest/