

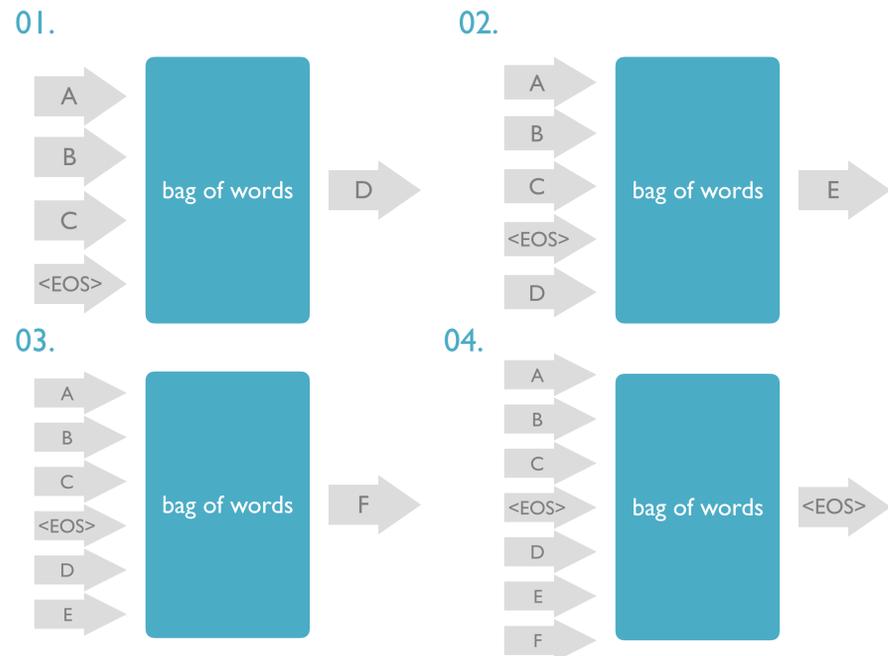
Personal Conversation Model

Ana-Maria Istrate & Lawrence Lin Murata
 CS229 Machine Learning by Andrew Ng, Stanford University

Method: Recurrent Statistical Model

We created the RSM model, which uses the words in the human utterance as inputs in a bag of words (TF*IDF and a mix of uni-grams and bi-grams). With the bag of words, a Support Vector Machine (SVM) with a linear kernel predicts the next word in the machine's response. It then adds the predicted word into the bag of words in order to predict the next word. The process is repeated until the <EOS> tag is predicted, indicating the end of the machine's response.

A purely statistical and recurrent model for building conversations



Other Methods Attempted

Logic Model

This model translates human utterances into first-order logic formulas and enforces logic rules in order to draw inferences and responses from the facts provided by the human.

N-gram Search

Given a starting n-gram, this model finds the next n-gram that is most likely to follow the previous n-gram in the sequence until the response reaches a certain predetermined length.

Examples

H: who did you vote for?
 M: voted for obama
 H: now we'll have a new president
 M: but we have the same old problems
 A: do you want to be president
 M: but think about all the power
 A: what type of power
 B: they forget where they came from maybe you
 H: you should be president
 M: but run for office

H: you had a date
 M: yes it was love at first sight
 H: was it a blind date
 M: you must be joking
 H: what's wrong with blind dates
 M: date someone don't even know
 H: okay then. you said it was love at first sight
 M: yes it was love at first sight
 H: that means you love her
 M: she's cute okay she's not ugly she's
 H: so she is not ugly
 M: she's just fat
 H: how fat
 M: she weighs 98 pounds okay okay

H: what do you want
 M: you me
 H: why you me
 M: don't know it's just so mean to like be who he's not here yet so and soooooo are you going
 H: this will not work
 M: is it out' of your league exits
 H: i can't do this
 M: oh we just delighted so uh you met yes someone huh
 H: you are too awkward
 M: so much
 H: you are so creepy
 M: oh much better than the coffee house

H: do you want to go out
 M: with me
 H: yes
 M: oh my god
 H: let's head out to grab coffee
 M: me yeah
 H: do you want to go somewhere
 M: c'mon
 H: are you having fun today
 M: yeah ok
 H: i have to pee
 M: i'll cancel the sodas

Results

Data:

We had 3 different data sets:

1. Data set 1: a small portion of the Friends script, (data set 1)
2. Data set 2: a script of both Season 1 and 2 together, (data set 2)
3. Data set 3: a script with Seasons 1,2 and 3 (data set 3).

Granular Conversation Test (0-4):

We developed our own test that breaks down the conversation into 4 components:

1. Semantics
2. Structure
3. Context
4. Feeling

