LING 572
Lit Review
Steven Chang, Yushi Zhao and Tianrun Sun

1. Oh and Rudnicky, 2000

query_arrive_city	inform_airport
query_arrive_time	inform_confirm_utterance
query_confirm	inform_flight
query_depart_date	inform_flight_another
query_depart_time	inform_flight_earlier
query_pay_by_card	nform_flight_earliest
query_preferred_airport	inform_flight_later
query_return_date	inform_flight_latest
query_return_time	inform_not_avail
hotel_car_info	inform_num_flights
hotel_hotel_chain	inform_price
hotel_hotel_info	other

Figure 2 : utterance classes

airline	depart_date
arrive_airport	depart_time
arrive_city	flight_num
arrive_date	hotel_city
arrive_time	hotel_price
car_company	name
car_price	num_flights
depart_airport	pm
depart_city	price

Figure 3 : word classes

The goal of the paper is to design a corpus-based approach to natural language generation, specifically designed for spoken dialogue systems. The paper separated the implementation in to three steps, Content Planning, Stochastic Surface Realization, and Evaluation.

First part is content planning, the system will decide which attributes (e.g., depart_city, depart_airport) should be included in an utterance. Within the content planning there are two approaches, tagging old versus new information of the system and the statistic model for content generating (number of attributes from the corpus and the bigram model of the attributes).

Second part, Stochastic Surface Realization use the tagged utterance classes and word classes(Fig. 2, Fig. 3) which are tagged manually by CMU. Using the tagged utterances as described, built an unsmoothed n-gram language model for each utterance class. Tokens that belong in word classes (e.g., “U.S. Airways” in class “airline”) were replaced by the word classes before building the language models. By selecting 5-gram to introduce some variability in the output utterances while preventing nonsense utterances. Then is the generation of the utterances.

Last, is to evaluate the generation system by conducted a comparative evaluation by running two identical systems varying only the generation component. For the content planning part, old/new and bigrams algorithm were evaluating by usability survey with a score of 1 to 3 (1 = good; 2 =

okay; 3 = bad). For Surface Realization, conducted a batch-mode evaluation. Picked six recent calls to the system and ran two generation algorithms.

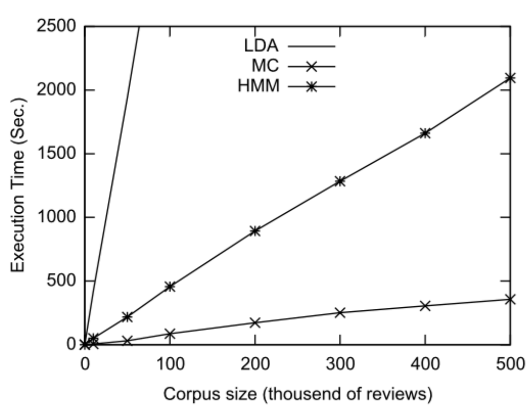
The stochastic generation system has several advantages, first is the response time which is important for a spoken dialogue system; second, by using a corpus-based approach, the system are directly mimicking the language of a real domain expert, rather than attempting to model it by rule.

2. Maqsud, 2015

The paper aims at comparing and evaluating the ability to generate texts with sentiments using three generative models, latent dirichlet allocation, Markov chains and hidden Markov Model. The experiment shows that MC achieves higher accuracy while HMM could generate a wider range of texts.

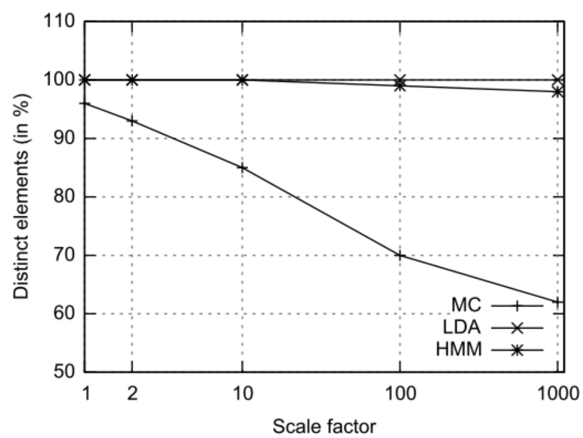
Before introducing the experiments, the author briefly presents the theoretical basis and the practical applications of the three models in text generation. The paper employs three approaches for sentiment analysis. The first approach is lexicon based in which the sentiment of a word is computed by averaging the scores of all related synsets of this given word using SentiWordNet. Numeric polarity scores will then be assigned to the word. The second method is the application of a binary classification model. SVMs will be adopted to separate the vectors, but since this machine learning approach takes advantage of annotated texts with sentiments provided (if I understand this part correctly), we won't consider this method for the purpose of the project. The last approach is by using Stanford Sentiment Treebank. The sentiment score of a sentence at its root node is calculated based on the score at each node. The author then takes the average of the scores over some sentences.

The paper runs three experiments to evaluate the models. Scalability is first assessed on food reviews datasets, which further contain seven sub datasets with different sizes.



The graph shows that MC needs the least execution time due to lower computational complexity.

The second experiment measures cardinality which a feature that assesses the diversity of a generated text. The dataset is a selection of 10,662 movie reviews with half positive and half negative reviews.



LDA and HMM generates a larger number of distinct elements due to the fact that they are not based on previous words.

The third experiment demonstrates the ability to generate texts with sentiments with the three models. An SVM is applied to check if the generated text contains the given sentiments.

	positive			negative		
	(a)	**(b)**	**(c)**	**(a)**	**(b)**	**(c)**
Original	63	75	**79**	57	75	**79**
LDA (1x)	60	73	68	52	71	58
LDA (2x)	62	70	68	52	69	59
LDA (10x)	63	70	69	55	67	59
MC (1x)	62	72	75	54	72	70
MC (2x)	62	73	75	55	**73**	72
MC (10x)	63	74	**76**	56	**73**	72
HMM (1x)	61	69	73	54	68	68
HMM (2x)	61	71	73	54	70	67
HMM (10x)	62	71	73	54	70	67

Again, MC achieves the highest F-measures.

3. Watanabe et al. 2014

Currently, music and melody can be generated by computer and smart devices. So, authors of this paper are eager to build a lyrics generator. They present novel generation models that capture the topic transitions between units peculiar to the lyrics, such as verse, chorus and line. These are modeled by Hidden Markov Model. Besides, they create 3 language models. In order to verify that their models generate suitable lyrics, they evaluate the models using a log probability of lyrics generation and fill-in-the blanks-type-test.

For first model, they utilize an tri-gram model assuming lyrics are characterized by fluent, easily sung word orderings. The model is conditioned by the appropriate mora length. For the second model, they use a state-transition model assuming that the line and verse are generated from a consistent, context-dependent word set. For the third, they change tri-gram into bi-gram to avoid data-sparsity due to the additional conditional parameter, the hidden state.

They evaluate the result by log probability of the original line and fill-in-the-blanks-type test.

From this paper, the method of building three models is helpful and can also be used in our poem generator. Besides, we can use authors' evaluating strategy to quantify the result.

References

Maqsud, Umar. (2015). Synthetic Text Generation for Sentiment Analysis. 156-161. 10.18653/v1/W15-2922.

Oh, A. H., & Rudnicky, A. I. (2000). Stochastic language generation for spoken dialogue systems. *ANLP/NAACL 2000 Workshop on Conversational Systems -*. doi:10.3115/1117562.1117568

Watanabe, K., Matsubayashi, Y., Inui, K., & Goto, M. (2014). Modeling structural topic transitions for automatic lyrics generation. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*.