Sentence Generation Using User-Defined Predictive Text

a Special Problems proposal by
John Alvin L. Sayson

Significance of the Study

- Previous ventures into text generation allowed for applications in multiple fields.

- Writing mediums with embedded text prediction are now commonplace in mobile phones, with the goal of minimizing the general writing time in communication.

- The study aims to achieve a similar goal, focusing more on long-form outputs.

- Text options in the editor can provide ideas visually when using the proposed application for creative writing.

Objectives of the Study

GENERAL OBJECTIVE

The study aims to improve user writing speeds by allowing them to create long-form outputs using a predictive text system embedded in a text editor application.

SPECIFIC OBJECTIVES

1. Categorize input source material into genres and convert it into an interpretable representation;
2. Use a word-level language model to assess word relation in the sources provided;
3. Present word relation to the end user in the form of a predictive text writer interface;
4. Allow user text selection to influence future predictive outputs shown in the interface; and
5. Assess user experience and program outputs using surveys.

Proposed Methodology

Program Flow

- The predictive text system shall be implemented in Python, to make use of the multiple natural language processing libraries in the programming language.

- Input: Plaintext corpora that shall be tagged with its genre/type and converted into a bag-of-words equivalent and a chain. The application shall be able to accept multiple inputs.

- Processing: All possible first words shall be generated. Upon user selection of the first word, the suggestions shall change into the next possible words that will follow the previous one.

- Output: Ending the writing process shall produce a tangible output for the user to review, allowing them to fine-tune their output as they wish.

Evaluation

- Evaluation shall be done by conducting surveys to multiple respondents with experience in using the application.

- Aspects of the application to be evaluated shall be the intuitiveness, helpfulness and effectiveness of the application for the user's experience.

- Effects on the user such as their productivity and creativity shall also be considered in the evaluation.

Preliminary Results

Prototype Application

- The prototype application implements bag-of-words conversion, chain generation, and probability ranking of the next possible words given one word.

- The prototype included dynamic Hidden Markov Model implementations to create the probabilities, stored in an adjacency matrix. The converted chain was used in counting word occurrences.

- The formula used in Bayes' theorem is implemented to calculate these probabilities.

```
  1  Harry Potter and the Sorcerer's Stone
  2
  3  CHAPTER ONE
  4
  5  THE BOY WHO LIVED
  6
  7  Mr. and Mrs. Dursley, of number four, Privet Drive, were proud to say that they were perfectly normal, thank you very much.
  •  They were the last people you'd expect to be involved in anything strange or mysterious, because they just didn't hold with
  •  such nonsense.
  8
  9  Mr. Dursley was the director of a firm called Grunnings, which made drills. He was a big, beefy man with hardly any neck,
  •  although he did have a very large mustache. Mrs. Dursley was thin and blonde and had nearly twice the usual amount of neck,
  •  which came in very useful as she spent so much of her time craning over garden fences, spying on the neighbors. The
  •  Dursleys had a small son called Dudley and in their opinion there was no finer boy anywhere.
 10
 11  The Dursleys had everything they wanted, but they also had a secret, and their greatest fear was that somebody would
  •  discover it. They didn't think they could bear it if anyone found out about the Potters. Mrs. Potter was Mrs. Dursley's
  •  sister, but they hadn't met for several years; in fact, Mrs. Dursley pretended she didn't have a sister, because her sister
  •  and her good-for-nothing husband were as unDursleyish as it was possible to be. The Dursleys shuddered to think what the
  •  neighbors would say if the Potters arrived in the street. The Dursleys knew that the Potters had a small son, too, but they
  •  had never even seen him. This boy was another good reason for keeping the Potters away; they didn't want Dudley mixing with
  •  a child like that.
 12
 13  When Mr. and Mrs. Dursley woke up on the dull, gray Tuesday our story starts, there was nothing about the cloudy sky
  •  outside to suggest that strange and mysterious things would soon be happening all over the country. Mr. Dursley hummed as
  •  he picked out his most boring tie for work, and Mrs. Dursley gossiped away happily as she wrestled a screaming Dudley into
  •  his high chair.
 14
 15  None of them noticed a large, tawny owl flutter past the window.
 16
 17  At half past eight, Mr. Dursley picked up his briefcase, pecked Mrs. Dursley on the cheek, and tried to kiss Dudley
  •  good-bye but missed, because Dudley was now having a tantrum and throwing his cereal at the walls. "Little tyke," chortled
  •  Mr. Dursley as he left the house. He got into his car and backed out of number four's drive.
```

Fig. 1. Text corpus used in the prototype application.

Fig. 2. Results of calling get_top_rankers() using "harry" and 10 as parameters after generating the chain and bag of words.