Generating NFL Game Headlines from Box Score Statistics

Parker Greene, Itamar Belson, John Clarke
{pwgreene, ibelson, jaclarke}@mit.edu

I. Abstract

II. Introduction

With the large number of users and diverse use of applications within the MIT wireless network, maintaining performance becomes critical. When devices attempt to connect to the MIT network, it is common for multiple Access Points to be within range so devices will choose to connect to the strongest signal AP. However, this may not be optimal for performance, since signal strength indicates nothing about traffic congestion. Depending on the application, it could be better to connect to an AP whose signal strength is weaker, but has less congestion and thus higher throughput.

A. Motivation

In looking to address this issue and optimize performance, MIT facilities seeks for its wireless internet users to be able to connect to Access Points (APs) with acceptable performance, either by connecting to an AP within range or by connecting to a nearby recommended AP, all while still maintaining maximum network utilization across the whole MIT network. In addition, for the purposes of network management, MIT facilities needs to collect data on each AP's amount of data transfer as well as its approximate number of unique users. To accomplish this, we propose Latch, a simple, reliable, and scalable system for optimizing network performance for users and optimizing network utilization.

III. Related Work

Generation of somewhat short sentences using Markov chains and stochastic methods have been shown to be quite simple and effective in generating poetry [1, 2], so we focused on applying these same principles in synthesizing valid headlines. This method yields a good way for text generation learned from examples; however, there is relatively much less research in the topic of generation solely from a combination of statistics and their corresponding text examples. Other, more syntactically-based models geared toward dialogue systems [3, 4], were shown to be inapplicable due to the lack of grammatical structure of the headlines in our data. Wen et. al. [5] presented an empirically-tested LSTM model capable of generating linguistically varied responses, but applying a deep, recurrent model to our problem proved hard due to the lack of data and difficulty of integration of the statistics in training.

Football game summaries have been explored by Nichols et. al. [6] and Chakrabarti & Punera [7], but both methods relied on learning data directly from Twitter and were uninformed by the statistics of games, as our model was.

IV. Data Collection

V. Approach

1.

VI. Results

VII. Contribution

A. Parker Greene

B. Itamar Belson

C. John Clarke

VIII. References

[1] Kenner, Hugh; O'Rourke, Joseph (November 1984). "A Travesty Generator for Micros". BYTE. 9 (12): 129-131, 449-469.

[2] Hartman, Charles (1996). *Virtual Muse: Experiments in Computer Poetry*. Hanover, NH: Wesleyan University Press.

[3] Alice H. Oh and Alexander I. Rudnicky. 2000. "Stochastic language generation for spoken dialogue systems". In Proceedings of the 2000 ANLP/NAACL Workshop on Conversational Systems - Volume 3, ANLP/NAACL-ConvSyst '00.

[4] Adwait Ratnaparkhi. 2002. "Trainable approaches to surface natural language generation and their application to conversational dialog systems". Computer Speech and Language.

[5] Tsung-Hsien Wen, Milica Gasic, Dongho Kim, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. 2015. "Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking". In Proceedings of SIG-dial. Association for Computational Linguistics.

[6] Jeffrey Nichols, Jalal Mahmud, Clemens Drews. 2012. "Summarizing Sporting Events Using Twitter". IBM Research - Almaden

[7] Chakrabarti, D., and Punera, K. 2011. "Event summarization using tweets". Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

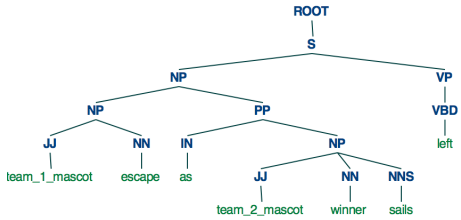

Fig. 1. PCFG example