Highlighting Text Regions of Interest with Character-Based LSTM Recurrent Networks

Johannes Knittel* Steffen Koch† Thomas Ertl‡

VIS
University of Stuttgart

Abstract

Exploring large sets of textual documents to find relevant, new insights is a tedious task, particularly if interesting passages are sparsely scattered throughout the collection. In many cases, it is even not feasible to manually inspect every paragraph, either due to the massive amount of data at hand or when there is a need for quick assessments, e.g. journalistic investigations after leaks.

We propose a new visual text analytics approach that relies on character-based Long Short-Term Memory (LSTM) recurrent neural networks to highlight text sections depending on their model probability. Thus, users can quickly dismiss sections that only contain common, redundant information. We trained three different LSTM models on subtitles of TV channels, tweets, and news reports. In this work we present analysis tasks that highlight the applicability and indicate the usefulness of our approach.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Visual analytics;

1 Introduction

It is impossible to analyze by hand the massive amount of textual data published each day through social media posts and news. As a consequence, automatic aggregation and extraction techniques have been proposed and combined with interactive visual approaches to cope with this situation. Prominent examples include topic modeling that aims to provide a general overview of the most important themes, and entity detection which can help understand how people, locations and objects relate to each other. Specific filters can then be applied to extract sources that are relevant to the analyst. However, these strategies typically do not assess whether regions related to entities or topics contain novel information. Hence, analysts still have to skim through all related documents relying on additional metadata and their experience to further narrow down the search results.

We propose a new visual analytics approach to assess the novelty of written text that makes use of complex language models based on recurrent neural networks and which offers a user steerable visual representation of the results. Our core assumption is that regions that are less predictable by the model have a higher chance to contain novel and interesting information than others. If the probability is too low, however, this indicates that the input is noisy or erroneous (e.g. contains spelling errors). Hence, we visually highlight text regions that are rather unlikely under our model. This visualization quickly guides analysts to paragraphs that seem to be most interesting, further reducing the amount of text that has to be read in depth. Our approach is inspired by human reading behavior to some extent. It has been shown that word skipping and fixation durations are heavily influenced by word predictability [5].

*e-mail: johannes.knittel@vis.uni-stuttgart.de
†e-mail: steffen.koch@vis.uni-stuttgart.de
‡e-mail: thomas.ertl@vis.uni-stuttgart.de

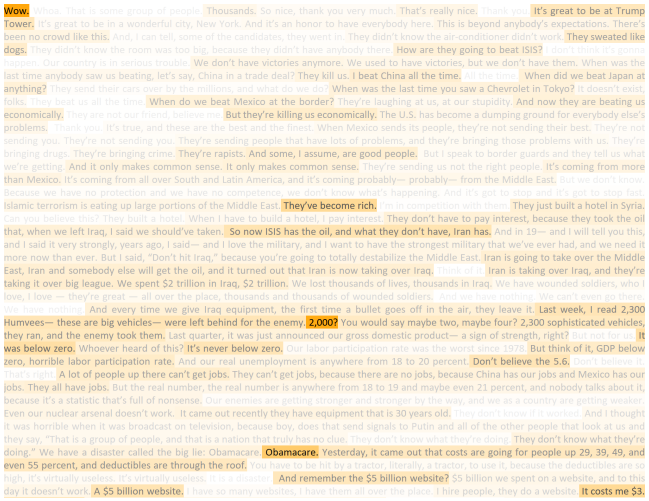

Figure 1: Visualization of extract of Donald J. Trumps presidential announcement speech. Mapped background and foreground colors are averaged per sentence. Darker text and orange highlighting means less likely under the model.

2 Related Work

Several visualization approaches have been developed in the past to better understand the inner workings of machine learning models, particularly neural networks. Karpathy et al. [4] visualized activation patterns and cell states of character-based LSTM networks trained on text sources showing that LSTMs can learn long-range interactions. Strobelt et al. [6] developed LSTMVis for visualizing hidden state dynamics in recurrent networks. Hohman et al. [3] provide a more comprehensive survey on visual analytics in deep learning.

3 LSTM character-level language model

We built several character-level language models using Long Short-term Memory recurrent neural networks to facilitate detection of unlikely and therefore potentially novel and interesting text fragments. LSTMs were introduced by Hochreiter and Schmidhuber [2] to improve the performance of traditional RNNs that have difficulties capturing dependencies over a longer time range when trained with backpropagation. On each timestep the next character is fed to the model using one-hot encoding. The network is trained to correctly predict the following character.

On the one hand, the resulting model can then be used to generate text by sampling character after character from the probability distribution of the softmax output layer. On the other hand, we can evaluate the LSTM network with new text and calculate the average cross-entropy error which is directly related to perplexity in NLP. This indicates how likely the data are under our model.

We trained three LSTMs with 2 layers of 512 hidden nodes over several weeks on different English datasets. We collected about

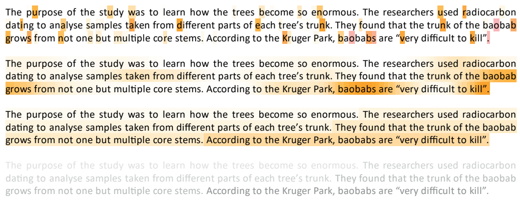

Figure 2: Four different visualization modes for the same text. The first three variants map values to the background color of the characters: raw, with gaussian blur, averaged per sentence. The last paragraph maps values to the foreground color, averaged per sentence.

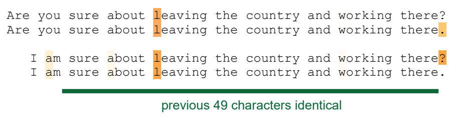

Figure 3: Visualizing probability of each character reveals that this LSTM network can model longer-distance relationships.

200.000 distinct news reports with 700m characters, 8m tweets with about 700m characters, and subtitles from ca. 48.000 tv shows and episodes with about 1 billion characters.

4 System Design

We developed a desktop application that allows to import pre-trained LSTM language models. Users can load text documents and enable the visualization which highlights regions based on how unlikely they are according to the language model.

Let $p(c)$ be the probability of the current character within its context of previous characters according to our language model. We calculate $f_c = (1 - p(c))^k$ which is still between 0 and 1. The result f_c is then linearly mapped to the saturation of the background color or transparency of the text. Intuitively, f_c represents the odds that a character with probability $p(c)$ would not have been sampled by the model in a typical sentence of length k. A high value of f_c therefore indicates that the specific character c is rather unlikely, taking the complete context of preceding text into account. In our examples we set $k = 100$ as default value.

Users can choose between different visualization modes. The probabilities can be independently mapped to the foreground and background color of the text view. Furthermore, the colors can be smoothed with a gaussian blur or averaged sentence-wise to better indicate regions of interest, since a low probability for a specific character means the sentence is continuing differently than the model predicted according to the *complete* context of previous characters. Therefore, the surrounding area is of interest and not just the character itself. Users can change the radius of the blur and choose to completely hide sentences that are very predictable by setting a threshold. Fig. 2 shows a subset of possible modes.

5 Applications of the Approach

Understand and debug trained recurrent networks: Visualizing outputs of neural networks allow researchers to better understand how their trained models work and perform [1]. Fig. 3 shows the resulting visualization of four similar sentences after applying our model trained on subtitles. The first two are formulated as questions and the last two as statements. The 49 characters before the punctuation mark at the end are identical. Nevertheless, the model correctly detects the full stop as being unlikely when the beginning of the sentence indicates a question, and the question mark as an unlikely punctuation for the statements. Thus, the visualization reveals that the trained LSTM in this case can memorize and model longer-distance relationships.

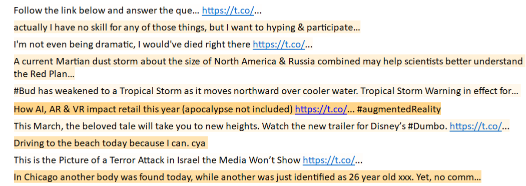

Figure 4: Markup of tweets according to their model probability.

Filter and focus: Highlighting regions that are less predictable enables analysts to focus on the most promising text passages and helps them to decide quickly if investigating a particular text fragment is worth the effort or not. Fig. 1 shows an example where our subtitle model is applied on parts of Donald Trumps presidential announcement speech. Content like "Obamacare" that is new to the model is prominently highlighted while more common expressions are faded out.

Analyzing tweets is especially challenging due to the vast number of posts each day. Relying on community-generated features like no. of retweets is one way to filter important tweets. Our approach offers a new way to quickly assess the novelty of tweets without having to wait for human reactions or favoring users with a big following. Fig. 4 shows the visualization of several tweets after applying our LSTM model trained on 8m tweets. In this example, posts containing rather generic information are clearly downrated.

Linguistic comparison: With the actual text hidden and only the markup visualization enabled, it is easy to compare different text sources structurally, e.g. retrace the arc of tension.

6 Conclusion

Preliminary results of our visual analytics system that applies powerful machine learning models to visually point out text regions of interest are promising. Language models using character-based LSTMs trained on large datasets seem to be able to model reasonably long relationships. While perplexity can be an indicator of novelty, statements that are simple from a language modeling viewpoint may nevertheless contain significant content and vice versa. Further research has to be carried out to properly evaluate how such a visual analytics system can support and speed up the work of analysts handling large amounts of texts, especially in real-time situations.

Acknowledgments

This work was funded by the DFG ER 272/13-1 project Visual Analytics of Online Streaming Text (VAOST).

References

[1] J. Choo and S. Liu. Visual analytics for explainable deep learning. *CoRR*, abs/1804.02527, 2018.

[2] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735

[3] F. M. Hohman, M. Kahng, R. Pienta, and D. H. Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2018. doi: 10.1109/TVCG.2018.2843369

[4] A. Karpathy, J. Johnson, and F. Li. Visualizing and understanding recurrent networks. *CoRR*, abs/1506.02078, 2015.

[5] K. Rayner, T. J. Slattery, D. Drieghe, and S. P. Liversedge. Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of experimental psychology. Human perception and performance*, 37(2):514–528, Apr. 2011. doi: 10.1037/a0020990

[6] H. Strobelt, S. Gehrmann, H. Pfister, and A. M. Rush. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):667–676, Jan 2018. doi: 10.1109/TVCG.2017.2744158