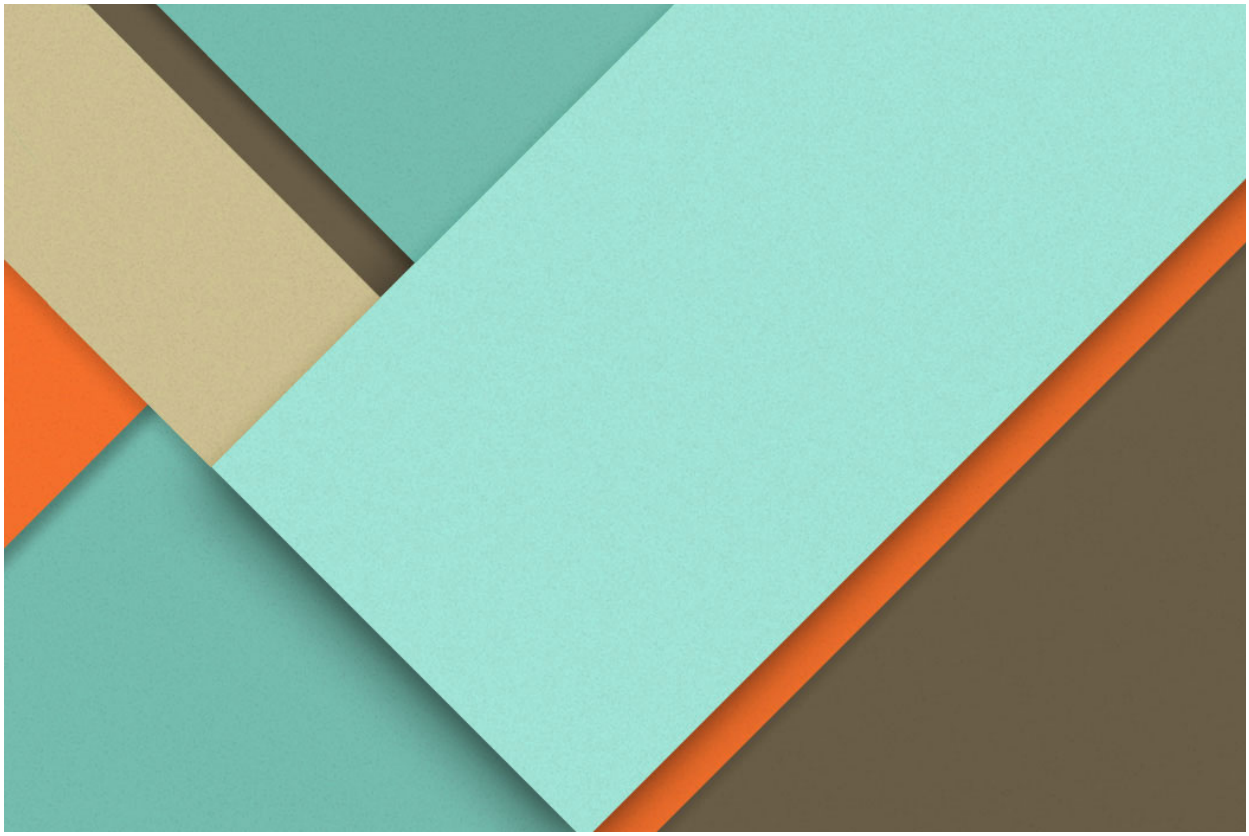

Holistic NLP Evaluator for Information Retrieval Systems

10.04.2019

Atri Basu

Overview

During the course of reviewing various NLP solutions for a particular use case, one thing I found I wish I had was a tool that would allow me to do a holistic apples to apples comparison of all the different Information Retrieval Systems. These types of problems, where you put some text into your model and get some other text out of it, are known as sequence to sequence or string transduction problems. Unfortunately for folks who are not familiar with NLP, there's no straightforward answer about what metrics should be used to evaluate the model. Even for someone familiar with NLP, no single metric provides a holistic appraisal as every metric has major drawbacks, especially when applied to tasks that it was never intended to evaluate.

Bottomline: NLP can't be measured as a whole, but rather specific to what one is trying to achieve.

Goals

1. Create an application that can provide a holistic and interpretable evaluation of an NLP model for the specific use case of information retrieval.
2. The application will be retrieval method agnostic, i.e. it will be able to evaluate the NLP model irrespective of whether it's a similarity-based model or a probabilistic model

Specifications

Information Retrieval is an empirically defined problem. Therefor, this evaluator will use user feedback gleaned in two different ways to evaluate the NLP model:

1. Directly from users, using a three star feedback mechanism defined as follows:
    a. Three stars - exact match
    b. Two stars - relevant match
    c. One star - irrelevant match
2. From linkage information between the documents created by the users to indicate relevance, i.e. whenever one document is referenced by another document it represents at least a two star relevance in the above scale.

The application will take the user feedback as input, generate an evaluation query set based on this feedback and when the results of an NLP model are provided as input to the application it will generate an analysis of the NLP model

What is the function of the tool?

The tool will provide an easily interpretable way for users to get a holistic evaluation of their NLP model based created for the use case of information retrieval. They will be guided through a step by step approach to providing certain output from their model for the evaluator to run on. Once the evaluator finishes it's run it will present not just the metrics but a visualisation of the metrics that make it easy to understand.

Who will benefit from such a tool?

Decision makers who may not be very well versed on how to evaluate NLP models but need to make a call on which solution to fund.

Does this kind of tools already exist? If similar tools exist, how is your tool different from them? Would people care about the difference?

There are libraries that make it easy to do an evaluation using individual metrics but no consolidated tool that can guide users on what metrics to collect, what those metrics mean or an understandable visualisation of these metrics.

What existing resources can you use?

Existing python libraries can be used to generate visualisations and do a lot of the calculations.

What techniques/algorithms will you use to develop the tool? (It's fine if you just mention some vague idea.)

We will use techniques like BLEU, TREC, Entropy Loss and Normalised Discounted Gain to develop the tool and various standard visualisation methods

How will you demonstrate the usefulness of your tool.

We will demonstrate the usefulness of the tool by using the tool to evaluate a few NLP models for one use case, like information retrieval.

Milestones

A very rough timeline to show when you expect to finish what. (The timeline doesn't have to be accurate.)

I. Team Selection - Oct 7th

Team Member : Atri Basu

II. Project Proposal - Oct 14th

III. Build Application -

A. Identify metrics to use and visualisations for each metric
B. Build application
C. Complete end to end testing
D. Identify demo data set

IV. Software code submission with documentation - Dec 16th 2019

V. Software usage tutorial presentation - Dec 16th