|  |  |
| --- | --- |
|  |  |
| Document Ranking with AllenNLP |  |
|  |  |
|  | 15 Dec 2020Artificial Intelligence |
|  | Devbrat Anuragi17078 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | What is AllenNLP AllenNLP is an open-source library for building deep learning models for natural language processing, developed by [the Allen Institute for Artificial Intelligence](https://allenai.org/). It is built on top of PyTorch and is designed to support researchers, engineers, students, etc., who wish to build high quality deep NLP models with ease. It provides high-level abstractions and APIs for common components and models in modern NLP. It also provides an extensible framework that makes it easy to run and manage NLP experiments.  In a nutshell, AllenNLP is   * A library with well-thought-out abstractions encapsulating the common data and model operations that are done in NLP research * A command line tool for training PyTorch models * A collection of pre-trained models that you can use to make predictions * A collection of readable reference implementations of common / recent NLP models * an experiment framework for doing replicable science * a way to demo your research * open source and community driven | |  |
|  | Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book.  Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. | Image |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | Lorem Ipsum is simply dummy text of the printing and typesetting industry. |  |  |
|  | An Overview on ranking There are several different kinds of document ranking tasks, such as ad-hoc retrieval, clickthrough rate prediction and more. Its simplest form, however, is the task of assigning a score to a query-document pair, which we call pointwise ranking. This is similar to textual entailment, but instead of our labels being entailment, contradiction, and neutral, our labels instead represent if the document is relevant to the query. This can be presented as either binary classification, or regression with continuous relevance scores between 0 and 1.  Because datasets for document ranking tend to be rather large, such as [MS-MARCO](https://microsoft.github.io/msmarco/) and [ClueWeb](https://lemurproject.org/clueweb09/" \t "_blank), we’ll use the [MIMICS](https://github.com/microsoft/MIMICS) query clarification dataset as a lightweight alternative. Each line contains a query, a clarifying question, and a list of options presented to the user. This is [used in Bing](https://twitter.com/albondarenko2/status/1225802655504781312/photo/1) to help refine search results.  {  "query": "headaches",  "question": "What do you want to know about this medical condition?",  "options": [ "symptom", "treatment", "causes", "diagnosis", "diet" ],  "scores": [0.05, 0.67, 0.42, 0.2, 0.0]  }  Each option will come with its own continuous score; our goal is to predict this score as closely as possible, given the query, the question, and the option itself. | |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Lorem Ipsum is simply dummy text of the printing and typesetting industry. |
|  |