

Interpretable explanations for sequence models

Mateusz Kobyłka

Faculty of Mathematics and Information Science
Warsaw University of Technology



WHY R? 2019 CONFERENCE
Warsaw, 29th September 2019



What is XAI?

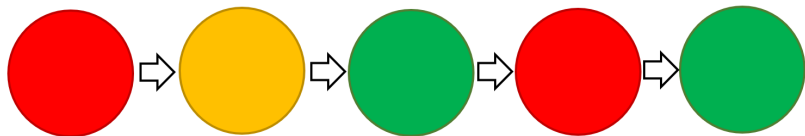
Explainable AI (XAI) refers to methods and techniques in the application of artificial intelligence technology (AI) such that the results of the solution can be understood by human experts.

Popular techniques:

- LIME,
- Anchors,
- SHAP,
- Ceteris Paribus,
- ...

Sequential data

Generally speaking, **sequential data** is any type of data where order matters.

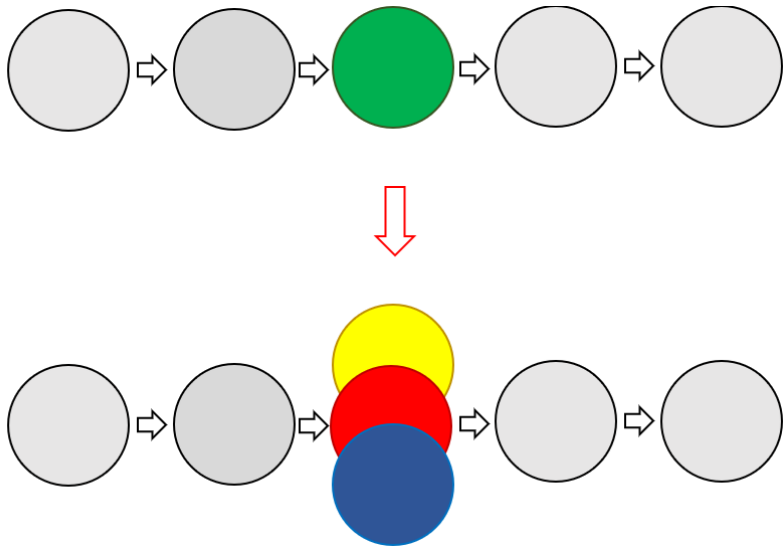


What is RME?

Recurrent Memory Explainer (RME) is a tool for creating interpretable explanations, dedicated to models operating on sequential data.

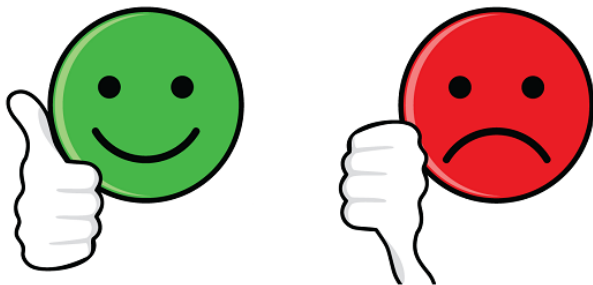


Intuition



Machine learning model for sentiment prediction

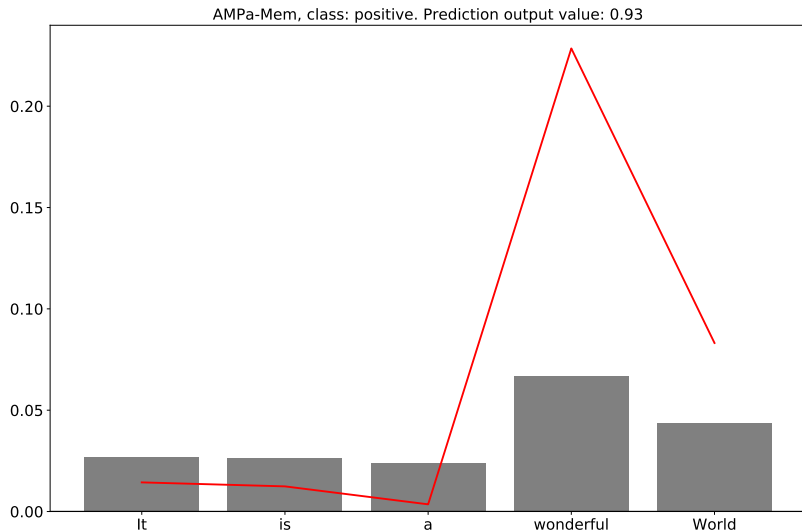
Movie review data: collections of movie-review documents labeled with respect to their overall sentiment polarity.



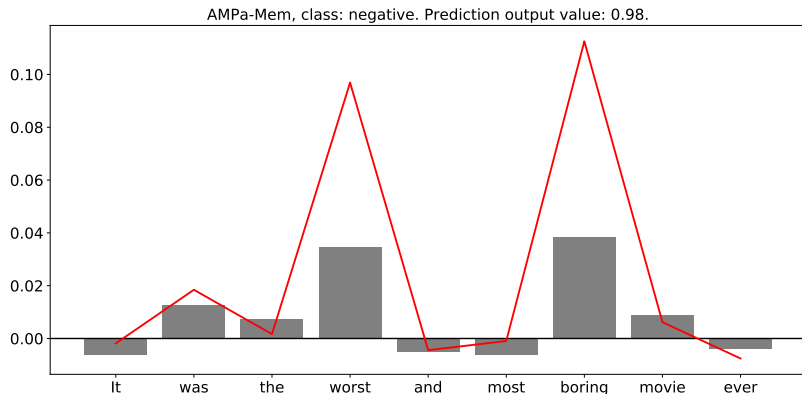
Source:

<http://www.cs.cornell.edu/people/pabo/movie-review-data/>

RME - demo



RME - demo



Find RME and me at:



<https://www.facebook.com/mateusz.kobylka.1>



<https://github.com/ModelOriented/RME>



kobylkam95@gmail.com



MI2DataLab research lab is looking for you!

Are you are intreseted in XAI, AutoML, AutoEDA other innovations in the next generation of ML?

Come to us

<http://bit.do/MI2isHiring>