|  |  |  |  |
| --- | --- | --- | --- |
| Title | Generating and Modifying Natural Language Explanations | | |
| Author(s) Name | Abdus Salam, Rolf Schwitter, Mehmet A. Orgun | | |
| Contact Email(s) | abdus.salam@aiub.edu | | |
| Published Journal Name | The 19th Annual Workshop of the Australasian Language Technology Association (ALTA 2021) | | |
| Type of Publication | Conference Article | | |
| Volume |  | Issue |  |
| Publisher | Association for Computational Linguistics (ACL) | | |
| Publication Date | December 2021 | | |
| ISSN |  | | |
| DOI |  | | |
| URL | https://aclanthology.org/2021.alta-1.0 | | |
| Other Related Info. | Pages 149–157 | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| HESIP is a hybrid explanation system for image predictions that combines sub-symbolic and symbolic machine learning techniques to explain the predictions of image classification tasks. The sub-symbolic component makes a prediction for an image and the symbolic component learns probabilistic symbolic rules in order to explain that prediction. In HESIP, the explanations are generated in controlled natural language from the learned probabilistic rules using a bi-directional logic grammar. In this paper, we present an explanation modification method where a human-in-the-loop can modify an incorrect explanation generated by the HESIP system and afterwards, the modified explanation is used by HESIP to learn a better explanation. | |