|  |  |  |  |
| --- | --- | --- | --- |
| Title | Human-understandable and Machine-processable Explanations for Sub-symbolic Predictions | | |
| Author(s) Name | Abdus Salam, Rolf Schwitter, Mehmet A. Orgun | | |
| Contact Email(s) | abdus.salam@aiub.edu | | |
| Published Journal Name | The 7th International Workshop on Controlled Natural Language, CNL 2020/21 | | |
| Type of Publication | Conference Article | | |
| Volume |  | Issue |  |
| Publisher | Association for Computational Linguistics (ACL) | | |
| Publication Date | September 2021 | | |
| ISSN |  | | |
| DOI |  | | |
| URL | https://aclanthology.org/2021.cnl-1.12/ | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| HESIP is a hybrid machine learning system in which a sub-symbolic machine learning component makes a prediction for an image classification and afterwards a symbolic machine learning component learns probabilistic rules that are used to explain that prediction. In this paper, we present an extension to HESIP that generates human-understandable and machine-processable explanations in a controlled natural language for the learned probabilistic rules. In order to achieve this, the literals of the probabilistic rules are first reordered, and then aggregated and disambiguated according to linguistic principles so that the rules can be verbalised with a bi-directional grammar. A human-in-the-loop can modify incorrect explanations and the same bi-directional grammar can be used to process these explanations to improve the decision process of the machine. | |