A Web-based Tool for Detecting Argument Validity and Novelty

Demonstration Track

Sandrine Chausson
University of Edinburgh
Edinburgh, United Kingdom
s.chausson@ed.ac.uk

Ameer Saadat-Yazdi
University of Edinburgh
Edinburgh, United Kingdom
ameer.saadat@ed.ac.uk

Xue Li
University of Edinburgh
Edinburgh, United Kingdom
xue.shirley.li@ed.ac.uk

Jeff Z. Pan
University of Edinburgh/Huawei
Edinburgh Centre, CSI, Huawei
Edinburgh, United Kingdom
j.z.pan@ed.ac.uk

Vaishak Belle
University of Edinburgh
Edinburgh, United Kingdom
belle.vaishak@ed.ac.uk

Nadin Kökciyan
University of Edinburgh
Edinburgh, United Kingdom
nadin.kokciyan@ed.ac.uk

Björn Ross
University of Edinburgh
Edinburgh, United Kingdom
b.ross@ed.ac.uk

ABSTRACT

Individuals engage in arguments on an everyday basis as they seek to obtain information about current affairs and engage with social media. While fact-checkers are available to help dispel misinformation, it is almost impossible for users to verify every single claim they encounter. This means that oftentimes, it is left to the user to decide whether a claim is well supported. To address this, we have developed a Web interface that allows users to input an argument, and our developed framework automatically detects its validity (soundness of logical deduction) and novelty (whether the argument is non-circular). Our Web-based tool could be used by social media users who wish to evaluate the information they consume. As part of one of the modules developed at the University of Edinburgh, our tool will be deployed as a teaching tool for the students who study argumentation.

KEYWORDS

Natural Language Inference; Computational Argumentation; Neural Networks; Knowledge Graphs

ACM Reference Format:
Sandrine Chausson, Ameer Saadat-Yazdi, Xue Li, Jeff Z. Pan, Vaishak Belle, Nadin Kökciyan, and Björn Ross. 2023. A Web-based Tool for Detecting Argument Validity and Novelty: Demonstration Track. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

When writing persuasive texts, there is strong evidence to suggest that logical sound and consistent arguments tend to be more effective [4, 7]. However, someone who is untrained in logic and argumentation will often struggle to determine whether the arguments they are presented with are logically valid. Having a tool to help users determine the quality of the arguments that they are making could help guide the development and assessment of arguments in critical texts. Another challenge that people often face is the rise of misinformation [8]. One step towards relieving users of the burden of verifying every claim a user encounters online would be to allow them to automatically assess whether the deduction in an argument is sound. Alongside tools for fact-checking premises, this would enable an end-to-end claim verification pipeline.

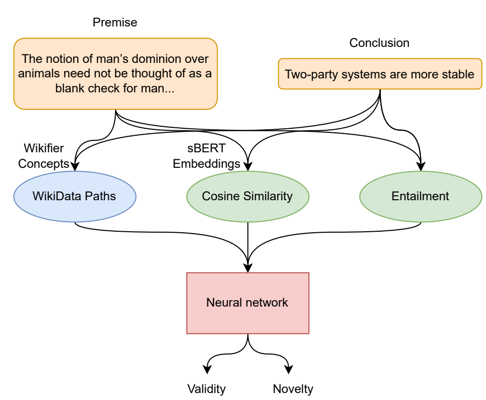

Figure 1: KEViN uses WikiData knowledge, and pre-trained transformers to predict similarity and entailment prior to feeding the data into the neural network.

In our previous work [10], we approached the problem of determining the validity and novelty of a conclusion given a set of premises. For this, we extracted paths from WikiData [11] that link the premise to the conclusion, and generate numerical features from these paths. We combined these features with those obtained from pre-trained language models for computing semantic similarity [9]

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

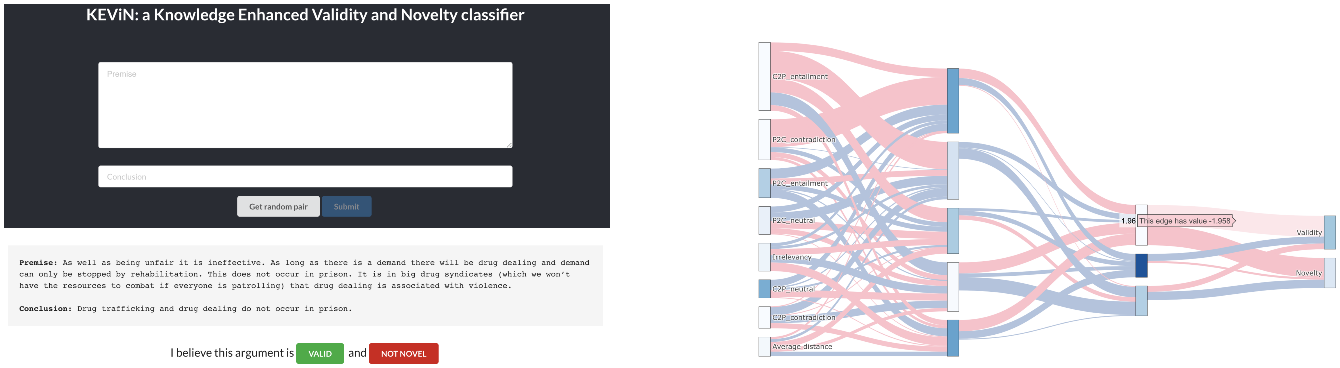

Figure 2: Users can input a premise or conclusion or select a random example from the dataset. The user input is then passed to the API and prediction results are returned and displayed. (Left) Visualisation of classifier activations are also shown. (Right)

and textual entailment [12] and ran a grid-search algorithm to identify the most relevant features to the task. We found that textual entailment serves as a strong indicator of argument validity, while a combination of textual entailment and knowledge graph distance improves the model's ability to detect novelty. The model trained on the set of most relevant features was named KEViN (Knowledge Enhanced Val[i]dity and Novelty classifier), depicted in Figure 1.

Based on these findings, we sought to translate our classifier into a usable tool that can be accessed via a Web interface to allow anyone who may not be familiar with the theory and code used to build the model to take make use of it (Figure 2). This interface allows users to input their own premises and conclusions and make inferences using KEViN. In addition to making inferences, our Web-based tool provides visualizations of the features obtained from knowledge graphs and pre-trained language models to help explain its behavior. We believe that this tool could be useful for students writing arguments in essays as well as for researchers in argument mining to explore how the various features interact given different inputs. While initially intended as a tool for introducing students in argumentation to some computational approaches, we believe that this Web-based tool could serve as a reasoning assistant for everyday users; particularly, as we iteratively improve upon the underlying classification model, KEViN. In other words, everyday users could use this tool to evaluate the information they consume online such as social media.

2 IMPLEMENTATION DETAILS

Our system consists of three components: (i) the API for performing inference and retrieving relevant data, (ii) the inference model based on KEViN, and (iii) the user interface. These were all written modular in such a way that components can be easily replaced to, for example, update the model, compute inferences for different input types, add/remove features and so on.

Web API. The Web API was built in Python using `flask` [3]. The API exposes a number of methods that can be accessed by the web portal via *GET* requests. The most relevant of these methods are `get_predictions`, which returns the classification predictions as well as the set of features used to make the classification. The system also has a number of cached instances pre-computed by running the model on the test set from the ArgMining 2022 shared task [5] via the `get_random` method.

Table 1: Performance of KEViN vs. a RoBERTa baseline on the test set for the combined task of validity and novelty prediction.

Model	Precision	Recall	F1
RoBERTa	0.21	0.26	0.21
KEViN	0.44	0.43	0.43

KEViN. The details on the implementation of our classifier model are introduced in Saadat-Yazdi et al. [10]. Given a premise and conclusion, KEViN (1), uses the Wikifier API [2] to obtain the two sets of WikiData entities found in the premise and the conclusion. A *SPARQL* query is then made to QLever in order to retrieve paths from WikiData [1] that connect premise entities to conclusion entities. For predicting textual entailment, a pre-trained BART model [12] was used without fine-tuning on our dataset. After pre-processing the dataset provided by Heinisch et al. [6] to obtain the aforementioned features, a small neural network was trained to predict both novelty and validity. The performance of the model in comparison to fine-tuned RoBERTa is given in Table 1.

User Interface (UI). The user interface was implemented in *React.js* (https://reactjs.org/). The UI allows the user to input a set of premises and a conclusion; or it also provides the user with the option to randomly select a pre-existing example from the dataset. This allows the user to understand the format and kind of text they are expected to input. Once an input is provided, the UI displays a prediction as well as the model's confidence in making this prediction. The user can also visualize some of the features such as the extracted paths, the entailment predictions, and the activations of the classification layers. Figure 2 shows the UI.

Tool Access. We provide our code with instructions on setting up and running the host server locally on GitLab[1]. A video demonstrating the functioning of the system can also be found here.

[1] https://git.ecdf.ed.ac.uk/s1876087/kevin-ui.git

ACKNOWLEDGMENTS

This work was supported by the ELIAI (Edinburgh Laboratory for Integrated Artificial Intelligence) EPSRC (grant no EP/W002876/1); the UKRI (grant EP/S022481/1) and the University of Edinburgh; the Edinburgh-Huawei Joint Lab and Huawei's grant CIENG4721/LSC.

REFERENCES

[1] Hannah Bast and Björn Buchhold. 2017. Qlever: A query engine for efficient sparql+ text search. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 647–656.

[2] Janez Brank, Gregor Leban, and Marko Grobelnik. 2017. Annotating documents with relevant wikipedia concepts. *Proceedings of SiKDD* 472 (2017).

[3] Miguel Grinberg. 2018. *Flask web development: developing web applications with python.* " O'Reilly Media, Inc.".

[4] Kyra Hamilton and Blair T. Johnson. 2020. *Attitudes and Persuasive Communication Interventions.* Cambridge University Press, 445–460. https://doi.org/10.1017/9781108677318.031

[5] Philipp Heinisch, Anette Frank, Juri Opitz, Moritz Plenz, and Philipp Cimiano. 2022. Overview of the 2022 Validity and Novelty Prediction Shared Task. In *Proceedings of the 9th Workshop on Argument Mining*. 84–94.

[6] Philipp Heinisch, Anette Frank, Juri Opitz, Moritz Plenz, and Philipp Cimiano. 2022. Overview of the 2022 Validity and Novelty Prediction Shared Task. In *Proceedings of the 9th Workshop on Argument Mining*. International Conference on Computational Linguistics, Online and in Gyeongju, Republic of Korea, 84–94. https://aclanthology.org/2022.argmining-1.7

[7] Blair T. Johnson and Alice H. Eagly. 1989. Effects of involvement on persuasion: a meta-analysis. *Psychological Bulletin* 106 (1989), 290–314.

[8] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096. https://doi.org/10.1126/science.aao2998 arXiv:https://www.science.org/doi/pdf/10.1126/science.aao2998

[9] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. http://arxiv.org/abs/1908.10084

[10] Ameer Saadat-Yazdi, Xue Li, Sandrine Chausson, Vaishak Belle, Björn Ross, Jeff Z. Pan, and Nadin Kökciyan. 2022. KEViN: A Knowledge Enhanced Validity and Novelty Classifier for Arguments. In *Proceedings of the 9th Workshop on Argument Mining*. International Conference on Computational Linguistics, Online and in Gyeongju, Republic of Korea, 104–110. https://aclanthology.org/2022.argmining-1.9

[11] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[12] Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. *CoRR* abs/1909.00161 (2019). arXiv:1909.00161 http://arxiv.org/abs/1909.00161