

Formula Concept Discovery using Wikidata

Background

Current content-based recommender and plagiarism detection systems perform poorly for STEM literature because they ignore mathematical formulae. We propose using the semantic database Wikidata (sister project of Wikipedia) to match formula concepts to named items.

Goal

Effective seeding of formula concepts from Wikipedia articles or arXiv documents to Wikidata items (Mathematical Language Processing task).

Tasks

- Propose three methods to discover semantic formula concepts in Wikipedia articles or arXiv documents and seed them as items to Wikidata or add defining formulae (including the semantics of the identifiers) to existing items.
- Evaluate your methods in comparison to each other. How many formula concepts were correctly seeded?



WIKIPEDIA
Die freie Enzyklopädie



Philipp Scharpf
philipp.scharpf@uni-konstanz.de



Moritz Schubotz
schubotz@uni-wuppertal.de

Neural Networks for Formula Recognition

Background

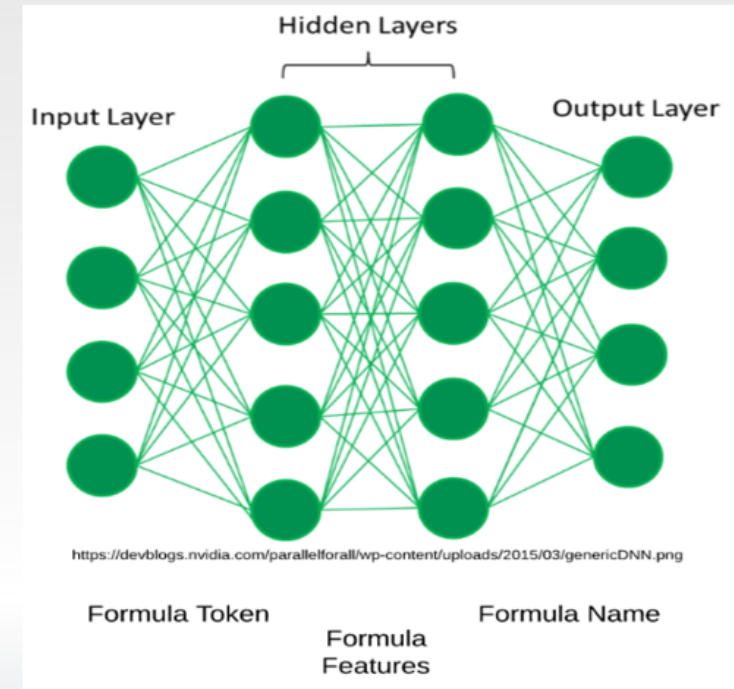
While advanced methods using Neural Networks exist for the recognition of handwritten text (OCR), images and human faces, no framework is currently able to identify mathematical formula concepts (e.g., " $E=mc^2$ " as "energy-mass-equivalence"). To solve this challenging problem, a given formula needs to be decomposed into its constituents, which can serve, e.g., as an input layer for of a Neural Network that matches them to formula names.

Goal

Design and train a suitable Neural Network to enable and optimize automatic formula recognition.

Tasks

- Review state-of-the-art Neural Network architectures that can be adapted to formulae recognition.
- Determine suitable formula features to train and evaluate the network.



Philipp Scharpf
philipp.scharpf@uni-
konstanz.de



Moritz Schubotz
schubotz@uni-wuppertal.de

