Project Form

Exploring analogical proportions between and within Knowledge Graphs (KG)

Supervisor:

- Team Orpailleur of the LORIA
- Main supervisor: Miguel Couceiro, miguel.couceiro@loria.fr
- Other supervisors: Esteban Marquer, esteban.marquer@loria.fr
  Pierre Monnin, pierre.monnin@orange.com

Description:

1. **Global Description** *Analogical proportions* are relations $A : B :: C : D$ between 4 elements A, B, C, and D, stating that the relation between A and B is equivalent to the one between C and D by capturing both similarities and dissimilarities between the elements. For instance, A, B, C, and D can be of a different but related nature: in $2 : 4 :: ab : abab$, 4 is the double of 2 in the domain of numbers while $abab$ is the "double" of ab in the domain of character strings. There are two main tasks associated with analogical proportions: analogy making and analogy solving. *Analogy making* (detection) consists of deciding whether four elements A, B, C, and D are in analogical proportion, and can be formulated as a classification problem. *Analogy solving* (resolution) focuses on the problem of finding a missing element x of an analogical equation $A : B : C : x$, and can be thought of as a regression problem.

   These two problems provide a logical framework to support various machine learning subtasks such as learning, transfer, and explainability, and that find noteworthy applications in artificial intelligence, recommandation and natural language processing (Mitchell 2021; Behrens 2017).

   Analogy making and solving have been addressed and tackled in the realm word semantics and morphology through both symbolic (Pierre-Alexandre Murena et al. 2020; Prade and Richard 2014; Miclet, Bayoudh, and Delhay 2008) and Deep Learning based approaches (Lim, Prade, and Richard 2019). A series of recent works (Lim, Prade, and Richard 2019; Alsaidi et al. 2021a; Alsaidi et al. 2021b; E. Marquer et al. 2022; Chan et al. 2022; Esteban Marquer, P.-A. Murena, and Couceiro 2022) explores Deep Learning (DL) to manipulate analogies that obtain competitive performance on word semantics and outperforms symbolic approaches on morphology.

   The current project aims to extend the results obtained on word morphology and semantics to more complex data, namely, Knowledge Graphs (KGs) (Hogan et al. 2021). KGs are useful graph based tools for encoding knowledge and that support several reasoning tasks. There are different approaches to encode KGs, including so called KG embeddings (KGEs) that are used for various knowledge discovery and engineering tasks, e.g., relation prediction, KG completion and reconciliation (Ji et al. 2022).

   We will first explore KG analogy making to detect interesting relations between KGs that can facilitate KG alignments and support relation prediction and KG completion (Monnin and Couceiro 2022). It will be necessary to gather KG data and organize it in a way that can be manipulated with analogy, which may lead to the creation of one or multiple datasets. It will be possible to reuse existing datasets published alongside existing approaches that have studied analogies and KGs (Liu et al. 2019; Portisch, Heist, and Paulheim 2022). Additional datasets could come from established challenges such as OAEI[a] or SemTab[b]. Such approaches will also constitute a baseline for comparison. Depending on the results of the approach, we will extend the work to KG analogy solving to directly tackle the above-mentioned tasks.

2. **Bibliography** (UE 705, semester 7) The bibliographic phase will be focused on four aspects:

    - studying the provided framework of analogy making and solving;
    - exploring KGE approaches to integrate the proposed framework;
    - studying knowledge discovery and engineering tasks and selecting some that can be solved using analogies;
    - studying competing approaches and setting baselines for knowledge discovery and engineering tasks;
    - selecting existing dataset sources for analogies between and within KGs.

3. **Implementation** (UE 805, semester 8) The project will have several implementation bits:

    - the baselines;
    - the necessary code to manipulate the data;
    - the dataset in itself;
    - the model in itself;
    - the model training and evaluation, including several empirical studies;
    - depending on the results, a demo of the system.

[a] http://oaei.ontologymatching.org/
[b] https://www.cs.ox.ac.uk/isg/challenges/sem-tab/

Information: Various material will be provided in the form of bibliographic sources and implemented codes to be further explored and tested.

Deliverables and Schedule:

- 1st & 2nd months: bibliographic work and problem(s) formalization.
- 3rd month: first hands-on and bibliographic report writing
- 4th & 5th months: dataset(s) preprocessing and code(s) implementation/adaptation
- 6th & 7th months: empirical study
- 8th month: thesis and paper writing.

References

Alsaidi, Safa et al. (October 2021a). “A Neural Approach for Detecting Morphological Analogies”. In: *The 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Porto/Online, Portugal. URL: https://hal.inria.fr/hal-03313556.

Alsaidi, Safa et al. (September 2021b). “On the Transferability of Neural Models of Morphological Analogies”. In: *AIMLAI, ECML PKDD 2021: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. Bilbao/Virtual, Spain. URL: https://hal.inria.fr/hal-03313591.

Behrens, Heike (2017). “The role of analogy in language processing and acquisition”. In: *The changing English language: Psycholinguistic perspectives*. Ed. by Marianne Hundt, Sandra Mollin, and Simone E.Editors Pfenninger, pp. 215–239.

Chan, Kevin et al. (2022). “Solving Morphological Analogies Through Generation”. In: *IARML@IJCAI-ECAI 2022*. Vol. 3174, pp. 29–39.

Hogan, Aidan et al. (2021). *Knowledge Graphs*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers. DOI: 10.2200/S01125ED1V01Y202109DSK022.

Ji, Shaoxiong et al. (2022). “A Survey on Knowledge Graphs: Representation, Acquisition, and Applications”. In: *IEEE Transactions on Neural Networks and Learning Systems* 33.2, pp. 494–514. DOI: 10.1109/TNNLS.2021.3070843.

Lim, Suryani, Henri Prade, and Gilles Richard (2019). “Solving Word Analogies: A Machine Learning Perspective”. In: *15th ECSQARU*. Ed. by Gabriele Kern-Isberner and Zoran Ognjanovic. Vol. 11726, pp. 238–250.

Liu, Chan et al. (2019). “A Survey of Recommendation Algorithms Based on Knowledge Graph Embedding”. In: *2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI)*, pp. 168–171. DOI: 10.1109/CSEI47661.2019.8938875.

Marquer, E. et al. (2022). “A Deep Learning Approach to Solving Morphological Analogies”. In: *30th ICCBR*, pp. 159–174.

Marquer, Esteban, P.-A. Murena, and Miguel Couceiro (2022). “Transferring Learned Models of Morphological Analogies”. In: *ATA@ICCBR 2022*.

Miclet, Laurent, Sabri Bayoudh, and Arnaud Delhay (2008). “Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning”. In: *Journal Artificial Intelligence Research* 32, pp. 793–824. DOI: 10.1613/jair.2519.

Mitchell, M. (2021). “Abstraction and Analogy-Making in Artificial Intelligence”. In: *Annals of the New York Academy of Sciences*.

Monnin, Pierre and Miguel Couceiro (2022). “Interactions Between Knowledge Graph-Related Tasks and Analogical Reasoning: A Discussion”. In: *ATA@ICCBR 2022*.

Murena, Pierre-Alexandre et al. (2020). “Solving Analogies on Words based on Minimal Complexity Transformation”. In: *29th IJCAI*, pp. 1848–1854.

Portisch, Jan, Nicolas Heist, and Heiko Paulheim (2022). “Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction - two sides of the same coin?” In: *Semantic Web* 13.3, pp. 399–422. DOI: 10.3233/SW-212892.

Prade, Henri and Gilles Richard (2014). “A Short Introduction to Computational Trends in Analogical Reasoning”. In: *Computational Approaches to Analogical Reasoning: Current Trends*. Ed. by Henri Prade and Gilles Richard. Vol. 548. Studies in Computational Intelligence. Springer, pp. 1–22.