



HAL
open science

Which Pairs to Choose? Exploring Analogical Competency for Knowledge Graph Pruning

Ndeye-Emilie Mbengue, Miguel Couceiro, Yves Lepage, Pierre Monnin

► To cite this version:

Ndeye-Emilie Mbengue, Miguel Couceiro, Yves Lepage, Pierre Monnin. Which Pairs to Choose? Exploring Analogical Competency for Knowledge Graph Pruning. IARML 2025 - Workshop on the Interactions between Analogical Reasoning and Machine Learning at IJCAI 2025, Aug 2025, Montreal, Canada. <hal-05162211>

HAL Id: hal-05162211

<https://inria.hal.science/hal-05162211v1>

Submitted on 15 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Which Pairs to Choose? Exploring Analogical Competency for Knowledge Graph Pruning

Ndeye-Emilie Mbengue^{1,*}, Miguel Couceiro², Yves Lepage³ and Pierre Monnin⁴

¹Université de Lorraine, CNRS, LORIA, Nancy, France

²INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

³Waseda University, IPS, Kitakyushu, Japan

⁴Université Côte d'Azur, Inria, CNRS, I3S, Sophia Antipolis, France

Abstract

Analogical reasoning is based on analogical proportions of the form $A : B :: C : D$ that can be read “ A is to B as C is to D ”. This type of reasoning has demonstrated its potential in enhancing machine learning algorithms, both in explainability and extrapolation. However, a crucial element lies in the choice of left-pairs (A, B) used to extrapolate on new situations (C, D) . In this paper, we propose to guide such a selection process by introducing different notions of competency that can be applied in the training or inference phases. We evaluate their impact on a binary classification task, namely knowledge graph pruning, and empirically demonstrate that competency-based selection not only improves the performance of the analogical reasoning model but also helps in reducing the number of necessary (A, B) pairs for accurate extrapolations. Our results open new perspectives to explore the impact of competency-based selection on knowledge graph-related tasks and models.

Keywords

Analogical Reasoning, Competency, Compression, Knowledge Graph, Pruning

1. Introduction

Analogical proportions are statements of the form “ A is to B as C is to D ”, e.g., “Paris is to France as Berlin is to Germany”, and are denoted by quadruples $A : B :: C : D$. They constitute a natural way for humans to reason about unfamiliar situations by drawing from known ones [1]. This form of reasoning thus consists in two steps: (i) detecting an analogy, and (ii) extrapolating the outcome of a new situation based on a past outcome. It is worth noting that analogical reasoning can also be leveraged in machine learning to improve generalization beyond the limits of train data objects [2]. In this view, its effectiveness has already been demonstrated across various domains, such as Natural Language Processing (NLP) [3, 4], or tasks involving

IARML@IJCAI'2025: Workshop on the Interactions between Analogical Reasoning and Machine Learning, at IJCAI'2025, August, 2025, Montreal, Canada

*Corresponding author.

✉ ndeye-emilie.mbengue@loria.fr (N. Mbengue); miguel.couceiro@inesc-id.pt (M. Couceiro);

yves.lepage@waseda.jp (Y. Lepage); pierre.monnin@inria.fr (P. Monnin)

🌐 <https://scholar.tecnico.ulisboa.pt/authors/ist430030> (M. Couceiro); <http://lepage-lab.ips.waseda.ac.jp/> (Y. Lepage);

<https://pmonnin.github.io/> (P. Monnin)

🆔 0009-0002-1289-6417 (N. Mbengue); 0000-0003-2316-7623 (M. Couceiro); 0000-0002-3059-4271 (Y. Lepage);

0000-0002-2017-8426 (P. Monnin)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

knowledge graphs [5, 6]. As with any machine learning approach, the quality of objects within the train dataset is decisive, but for analogy-based methods, this requirement is even more pronounced. In particular, the selection of pairs (A, B) plays a key role as they drive the process of extrapolating relevant and known knowledge to new situations [7]. Such a selection impacts not only model performance but also resource efficiency, as using competent pairs may lead to needing fewer pairs to infer from [8]. This motivates us to explore the following research question:

RQ How to select pairs (A, B) to enhance the performance of analogy-based models in classification and regression tasks?

These pairs would thus be the best suited to capture transferable knowledge to a new, unsolved task. In this paper we mainly address the former. Inspired by the case-based reasoning selection process [8, 9], we introduce and compare several notions of *pair competency* that we experiment on a concrete task, namely knowledge graph pruning [5, 10]. We compare their impact both in terms of task performance and number of needed (A, B) pairs, *i.e.*, (A, B) pair-base compression.

The remainder of this paper is organized as follows. Section 2 describes related work on competency definition and related applications, particularly for case-based and analogical reasoning. Section 3 presents the necessary preliminaries about knowledge graph pruning and how it can be framed as an analogical detection problem. In Section 4, we propose five different definitions of competency that are empirically evaluated in Section 5. Finally, we conclude the paper by outlining future work opened by our results in Section 6.

2. Related Work

The notion of *competency* has already been addressed in the literature, particularly within the framework of case-based reasoning [9]. Case-based reasoning relies on a collection of problems (*i.e.*, situations) and their corresponding solutions (*i.e.*, outcomes) in order to extrapolate the solution of new and unsolved problems. Analogical proportions “ A is to B as C is to D ” can be viewed in case-based reasoning such that A and C represent problems, and B and D represent solutions [9]. Given a new unsolved problem, the case-based reasoning process starts by retrieving already solved and similar problems from the case base. Then, the solutions associated with these retrieved problems are adapted to extrapolate the solution to the new unsolved problem. At the end of the process, an evaluation of the solution is performed and the unsolved task, along with its solution, is added to the case base.

Several studies have highlighted the importance of the initial retrieval step in case-based reasoning, and particularly of the similarity matching process that impacts both the performance of the reasoning system and the size of the case-base [11]. In this view, Marquer *et al.* [8] presents the CoAT indicator, an energy-based [12] model designed to measure a transfer similarity between the situation space and the outcome space. Their results show a direct correlation between their indicator and the case-based reasoning performance on a binary classification task, as well as its capacity for case-base compression. Their idea of measuring the transfer similarity of a couple formed by an object, *e.g.*, a vector, and its outcome on a task, *e.g.*, a class

prediction, to guide the selection of adequate objects for inference has inspired this study and the studied notion of *competency*. Similarly to them, we will tackle insights in terms of performance and (A, B) pair-base compression, for the specific task of KG pruning. It is worth noting that our study is not the first to introduce the notion of competency for analogical reasoning. For instance, Bounhas *et al.* [13] introduces a definition of competency for an analogy-based classification task on datasets with boolean-valued and nominal attributes. In their proposal, competency is framed as a dual concept, on one side on the train dataset objects, and the other side on a particular test object. Within the train set, a competent pair (A, B) is a pair of objects that minimizes the prediction error of the inference on this training set. With respect to a test object, a competent pair is defined as a pair of objects that reflects the same level of (dis)similarity as the pair of objects being tested. This dual view of competency will be adopted and extended in our study in order to address the task of knowledge graph pruning.

3. Analogical Detection for Knowledge Graph Pruning

3.1. Knowledge Graph Pruning

In this paper, we study the competency of (A, B) pairs used in analogical reasoning for the task of *knowledge graph pruning* [5, 10]. A *knowledge graph* (KG) can be defined as a “graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities” [14]. A KG can be formalized as $G = \{(h, r, t) \mid h, t \in V, r \in R\}$ where V is the set of entities (or nodes), R is the set of relation (or edge) types, and each triple (h, r, t) indicates that the head entity h is connected to the tail entity t with a directed edge labeled by the relation r . For instance, the triple (Colosseum, in, Rome) in Figure 1 represents the location of the Colosseum within the city of Rome.

In a KG, the pruning task (illustrated in Figure 1) reduces to a binary classification task to decide whether to *keep* or to *prune* the neighbors of seed entities of interest given as input. This classification is iteratively performed on neighbors of kept neighbors until no more neighbors are to process. The interest of the KG pruning task lies in extracting sub-graphs of interest from a huge KG. These sub-graphs can then be used in another application (*e.g.*, building a new KG focused on the topics of interest of a company, recommender systems) while avoiding application-irrelevant knowledge from the original huge KG, thus increasing the quality of the downstream application and reducing potential scalability issues.

3.2. From Analogy Classification to Knowledge Graph Pruning

Jarnac *et al.* [5] proposed to use analogy classification to perform the task of KG pruning. In particular, they use candidate analogies of the form $A : B :: C : D$ where A and C are two seed entities, B is a neighbor reached during the expansion from A , and D is a neighbor reached during the expansion from C . Using analogical inference, if this candidate is a valid analogy and the decision (keep / prune) is known for the pair (A, B) , then it can be extrapolated for the pair (C, D) . To illustrate, if the analogy Hadoop : Big Data :: SharePoint : Content Management System is valid and Big Data is kept, then Content Management Sys-

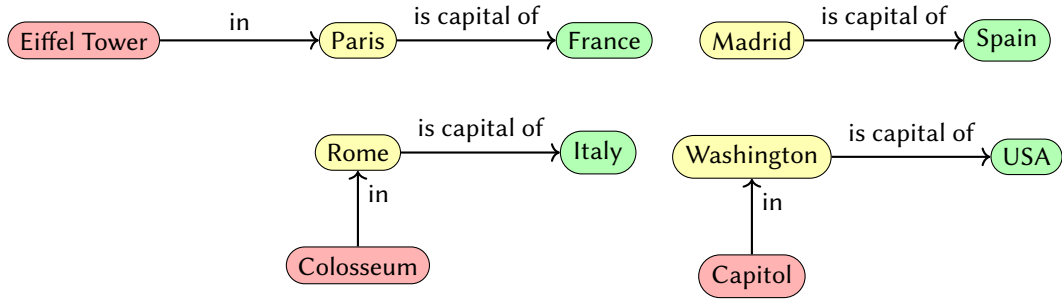


Figure 1: Illustration of the knowledge graph pruning task, where neighbors of seed entities of interest \circ are either kept \circ or pruned \circ .

tem should also be kept. Such an analogy-based framework was implemented with the two-layer convolution model from Lim *et al.* [15] represented in Figure 2 and applied on the vectorial representation of KG entities, *i.e.*, their KG embeddings. This framework empirically proved to significantly enhance pruning performance with a drastically lower number of parameters for the underlying deep learning model compared to other models from the state of the art [5].

Note that the notions of informativeness and interestingness are highly contextual. In the considered setting, the pruning process thus relies on (i) the set of seed nodes of interest and (ii) pairs (e_s^1, e_r^1) whose keep / prune decisions are known to implicitly define a specific domain of interest. In other words, a neighbor is kept or pruned depending on its adequacy with respect to these seed nodes and known pairs. This observation motivated the present study, whose goal is to demonstrate the critical impact of the selection of (A, B) pairs in the analogical inference process to classify new neighbors as kept or pruned. We propose to compare a random-selection approach with approaches based on different notions of *competency* of (A, B) pairs that we introduce in Section 4. Experiments presented in Section 5 are designed to highlight the ability of analogical reasoning to generalize across different domains of interest, and to showcase the importance of the (A, B) pair-base in this process.

4. Measuring (A, B) Pairs Competency

Following the work of Bounhas *et al.* [13], competency is approached from two perspectives: on one side from the perspective of a particular test object (C, D) , and the other side from the perspective of all train objects. Recall that, differently from [13] that considers boolean and nominal attributes, in our case, we apply analogical inference on vector objects, *i.e.*, the KG embeddings that represent the involved entities A, B, C , and D .

4.1. From the Perspective of a Particular Test Object

From the perspective of a test pair of objects (C, D) at inference, we propose to assess and retrieve competent train pairs (A, B) to generate candidate analogies from their similarity with the pair (C, D) . One way of estimating the similarity between pairs (A, B) and (C, D) is to use the distance between their vector representations inside a latent space [16]. Since we only have

at our disposal vector representations of entities A , B , C , and D and not of pairs (A, B) and (C, D) , we propose to explore several methodologies to estimate the distance between these pairs, which are based on different combinations of embeddings of individual entities that make up the pairs:

M1 computes the average of the distances between the corresponding elements of the two pairs, *i.e.*, seed-seed and neighbor-neighbor: $d_1 = \frac{1}{2}(\|C - A\| + \|D - B\|)$

M2 computes the distance between the first objects of each pair, *i.e.*, the seed entities: $d_2 = \|C - A\|$

M3 computes the distance between the second objects of each pair, *i.e.*, the neighbors: $d_3 = \|D - B\|$

M4 computes the distance between the barycenters (midpoints) of each pair: $d_4 = \left\| \frac{A+B}{2} - \frac{C+D}{2} \right\|$

These methodologies enable the generation of candidate analogies for the test pair (C, D) by selecting so-called *competent* (A, B) pairs, *i.e.*, pairs that minimize the proposed distances.

4.2. From the Perspective of Train Objects

Here, we build upon the description proposed in [17] as “a set of pairs of objects from the train dataset that minimizes the prediction error during analogical inference”. In our case, we extend this perspective by building a first base $\tilde{\mathcal{P}}$ of pairs (A, B) whose classification label is known and that can come from any dataset. That is, such pairs can come from the train dataset or another dataset unused in training or testing.

We select competent pairs (A, B) from this first pair-base $\tilde{\mathcal{P}}$ to build the *competent pair-base* \mathcal{P} . These pairs (A, B) are selected because they minimize the prediction error as follows. For each pair $(A, B) \in \mathcal{P}$, we build a set of candidate analogies $\mathcal{C}(A, B) = \{A : B :: C : D \mid (C, D) \in \tilde{\mathcal{P}} \text{ and } (A, B) \neq (C, D)\}$. Recall that we know the classification label of each pair in $\tilde{\mathcal{P}}$, and thus we know for each candidate analogy whether it is valid. We compute the prediction error of (A, B) using the considered analogy model and the Log loss from Equation (1):

$$\text{LogLoss}(A, B) = -\frac{1}{|\mathcal{C}(A, B)|} \sum_{A:B::C:D \in \mathcal{C}(A, B)} [y_{A:B::C:D} \log(\hat{y}_{A:B::C:D}) + (1 - y_{A:B::C:D}) \log(1 - \hat{y}_{A:B::C:D})], \quad (1)$$

where $y_{A:B::C:D}$ is the true label of the candidate analogy $A : B :: C : D$, equal to 1 or 0 in our binary classification task, and $\hat{y}_{A:B::C:D}$ is its probability to pertain to class 1 as computed by the analogy model. We then rank (A, B) pairs by increasing order of their $\text{LogLoss}(A, B)$ and select the top- k to form the competent pair-base \mathcal{P} . These pairs are thus computed once, independently of test pairs, and used at inference for any test pair (C, D) . Recall that, on the contrary, M1-M4 selects specific competent pairs for each test pair (C, D) . We denote this methodology by **M5**. Note that M5 comes with a computational cost of $O(n^2)$ due to the computation of the different error scores.

5. Experiments

In this section we present an empirical study to illustrate the pertinence of the proposed competency measures on three real-world datasets.

5.1. Knowledge Graph and Knowledge Graph Embedding Model

The experiments were conducted on the Wikidata Knowledge Graph¹, which contains approximately 117 million entities and 12,515 relation types. All entities are represented through a numerical vector representation (*i.e.*, an embedding), generated using the TransE model [18], a translational embedding approach that models KG triples (h, r, t) as translations r between h and t in the Euclidean space, *i.e.*, $h + r = t$. The embedding training process was carried out using the PyTorch-BigGraph framework, which is designed to efficiently handle large scale graph data like Wikidata. All the materials (*i.e.*, code and datasets) necessary to carry out this study are available on Github².

Due to the nature of the embedding model [18], that represents relations between entities as translations in the Euclidean space, the Euclidean distance (2) is selected to compute d_1, d_2, d_3, d_4 .

$$d_{\text{Euclidean}}(\mathbf{s}_1, \mathbf{s}_2) = \sqrt{\sum_{i=1}^n (s_{1i} - s_{2i})^2} \quad (2)$$

where $\mathbf{s}_1 = (s_{11}, \dots, s_{1n})$ and $\mathbf{s}_2 = (s_{21}, \dots, s_{2n})$ are the embeddings of two entities.

5.2. Datasets

Three labeled datasets were used for the knowledge graph entity pruning task. Each dataset comprises pairs of entities (seed, neighbor) extracted from Wikidata, annotated with decisions *prune* or *keep*. The first two datasets (DS1 and DS2)³, originally introduced by [5], focus on entities from the domains of Computer Science and General Knowledge, respectively. We introduce here a third dataset (DS3)⁴ that targets entities related to the domain of Works of Art. Key statistics describing the size and label distribution across the three datasets are presented in Table 1 and justify the need of using a weighted measure of performance in order to take into account the imbalance inside the datasets.

To conduct the different experiments, each dataset was divided into three splits: two splits (train and validation) for the training of the analogical model with the early stopping mechanism based on the validation loss, and one (test) for the evaluation phase of the different selection methodologies. Table 1 reports the size of the data splits used for each phase and each dataset.

¹<https://www.wikidata.org>

²https://github.com/lielie-dot2/competency_knowledge_graph_pruning

³<https://doi.org/10.5281/zenodo.8091583>

⁴<https://doi.org/10.5281/zenodo.15600971>

Table 1

Datasets characteristics. For each dataset, “Size” represent the total number of decisions, while “# Keep” and “# Prune” respectively represent the number of keep and prune decisions.

| Dataset | Size | # Keep | # Prune | Train size | Validation size | Test size |
|---------|------|--------|---------|------------|-----------------|-----------|
| DS1 | 5233 | 1769 | 3464 | 3764 | 941 | 524 |
| DS2 | 982 | 594 | 388 | 705 | 177 | 100 |
| DS3 | 2111 | 933 | 1178 | 1519 | 380 | 212 |

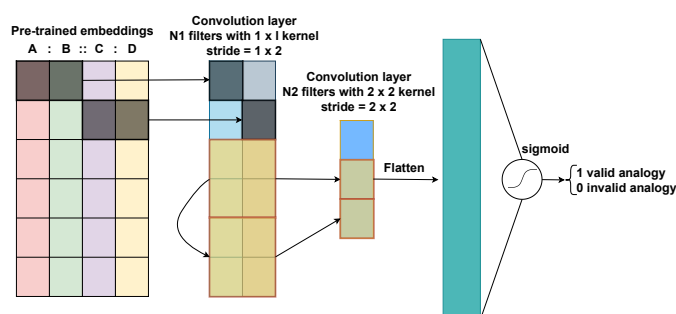


Figure 2: The analogy-based classifier model with two convolution layers and one fully connected layer from [15]. Image from [5].

5.3. Analogical Model

Following Jarnac *et al.* [5] and as aforementioned, we rely on the model proposed by Lim *et al.* [15] represented in Figure 2 to perform analogical inference. This model is a convolutional neural network (CNN)-based architecture comprising two convolutional layers. Key hyperparameters including the number of filters per layer, learning rate, and dropout rate were tuned using Bayesian optimization within predefined search ranges (Table 2). An early stopping mechanism with a patience of 5 on the validation loss was implemented to prevent overfitting.

The model is trained according to the first configuration described in [5], *i.e.*, “*valid analogies are of the form $k :: k$ and invalid analogies are of the form $k :: p$* ”. *Valid analogies* are defined as analogies formed with pairs (A, B) and (C, D) that both belong to the class *keep* in a labeled dataset, whereas *invalid analogies* involve pairs where (A, B) belongs to the class *keep* and (C, D) to the class *prune*. During training, the random selection of examples from the train dataset described in [5] is preserved, since the current study focuses on comparing selection methodologies during analogical inference. However, it is worth noting that competency-based selection strategies could also be integrated during training to potentially enhance the model’s performance. The methodology used to extrapolate classes on unseen pairs follows a majority strategy. For each test pair (C, D) , N inference candidate analogies $A : B :: C : D$ are built, based on labeled pairs (A, B) known to belong to the class *keep*. For each of these analogies, if the probability predicted by the model is higher than a given threshold (0.6) then a vote is accounted for a *keep* extrapolation. At the end of the process, if there is a majority of votes for *keep*, then the test pair is classified as *keep*.

Table 2

Hyperparameters of the analogical model and their search ranges.

| Hyperparameter | Search Range / Values |
|---|----------------------------------|
| Learning rate | $[10^{-5}, 10^{-1}]$ (log scale) |
| Batch size | {16, 32, 64} |
| Number of filters of the first convolutional layer | {8, 16, 32} |
| Number of filters of the second convolutional layer | {8, 16} |
| Dropout rate | [0.0, 0.5] |
| Number of epochs | 30 |
| First convolutional layer kernel size | (1,2) |
| Second convolutional layer kernel size | (2,2) |
| First convolutional layer strides | (1,2) |
| Second convolutional layer strides | (2,2) |

5.4. Main Results

We tested our competency-based selection strategies on three distinct configurations:

- C1** The model is trained on the train set of a particular dataset, tested on the test set of this dataset, and (A, B) pairs are retrieved from the train set of the same dataset at inference.
- C2** The model is trained on the train set of a particular dataset, tested on the test set of this dataset, and (A, B) pairs are retrieved from the train set of the three datasets at inference.
- C3** The model is trained on the train set of a particular dataset, tested on the test set of another dataset, and (A, B) pairs are retrieved from the train set of the three datasets at inference.

These three configurations aim at comparing the different selection methodologies by training and testing the model with entities that either belong or not to the same domain and evaluate their contributions. As a baseline, we adopt the random-based selection methodology.

In the following, only the best weighted F-measures obtained and the associated number of (A, B) pairs for each dataset are presented. In particular, we tested with a number of inference candidate analogies from 0 to 20 and retained the number of pairs (A, B) leading to the best weighted F-measure. It is worth noting that for the random selection, each experiment was repeated 150 times, and the reported F-measure corresponds to the mean weighted F-measure of those different repetitions along with its standard deviation.

Results obtained with configuration C1. Table 3 presents the best weighted F-measures obtained for the experimental configuration **C1**, *i.e.*, considering (A, B) pairs of the same domain.

Table 3 shows that on all datasets, the random selection achieves the lowest F-measure, highlighting the interest of studying the competency of pairs (A, B) to use during the analogical inference process. However, we note that competency-based methodologies achieve heterogeneous results depending on the dataset used. For instance, for DS1, the highest F-measure (0.89) is achieved using M3. This improvement represents a 10-point gain compared to the

baseline, *i.e.*, the random methodology. It also constitutes a 9-point gain compared to the lowest performing methodology M5. For DS2, the methodology M3 also stands out, contributing to a 5-point gain compared to the random methodology. For DS3, competency-based methodologies and the random methodology provide quite similar weighted F-measures, which might suggest a difficulty for those competency approaches to capture effectively the structural properties of this dataset.

Table 3

Best weighted F-measure obtained on the three considered datasets with the different selection methodologies and experimental configuration **C1**. The random selection process averaged on 150 run is given as a baseline. Bold numbers represent the best result obtained for each dataset.

| Datasets | Random | M1 | M2 | M3 | M4 | M5 |
|----------|------------------|------|-------------|-------------|-------------|-------------|
| DS1 | 0.79 ± 0.002 | 0.80 | 0.86 | 0.89 | 0.80 | 0.80 |
| DS2 | 0.81 ± 0.009 | 0.85 | 0.83 | 0.86 | 0.86 | 0.85 |
| DS3 | 0.83 ± 0.007 | 0.83 | 0.84 | 0.83 | 0.83 | 0.84 |

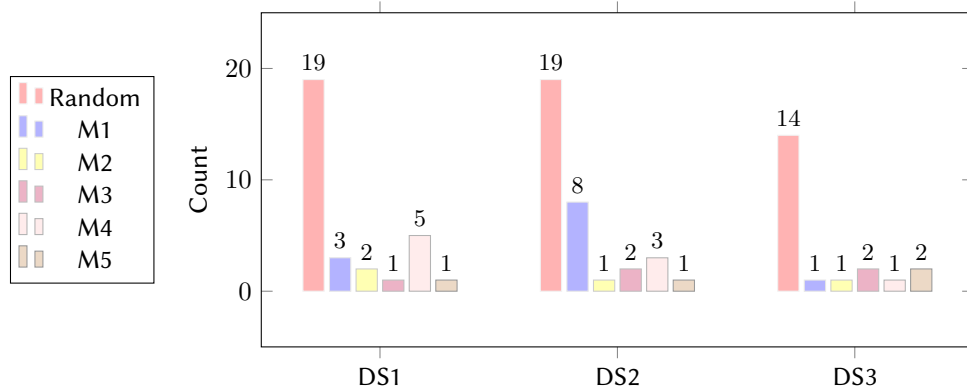


Figure 3: Number of pairs (A, B) used from each dataset using the different selection methodologies to obtain the best weighted F-measure reported in Table 3 on the experimental configuration **C1**.

In terms of (A, B) pair-base compression, Figure 3 shows that all competency-based selection methodologies reduce the number of pairs needed to achieve the best weighted F-measure. For example, for DS1, methodology M3 achieves a 10-point performance gain compared to the random methodology while reducing the required number of analogical pairs from 19 to just 1. In DS2, methodology M3 achieves a 5-point performance improvement while reducing the number of pairs from 19 to 2. Even in DS3, where the performance gain is minimal in terms of weighted F-measure, the number of pairs needed to reach the best weighted F-measure is drastically reduced. Recall, even if methodology M5 also competitively reduces the number of (A, B) pairs needed to reach the best F-measure, it comes with a computational cost induced by the need of pre-computing the different error scores (*i.e.*, $O(n^2)$), which may hinder scalability in real-world, large-scale KGs. We leave this study as future work.

Results obtained with configuration C2. Table 4 presents the best-performing selection methodologies and their associated F-measure for the experimental configuration C2, *i.e.*, training and testing on one dataset, and selecting at inference competent (A, B) pairs not necessarily from the same domain but retrieved from all datasets.

Table 4

Results obtained on the experimental configuration C2, *i.e.*, selecting competent (A, B) pairs of different domains, for a model trained and tested on one dataset. For each dataset to select competent (A, B) pairs from, the best performing selection methodology and its associated weighted F-measure are reported.

| Train & test | (A, B) pairs dataset | F-measure | Competency method | Random method |
|--------------|------------------------|-------------|-------------------|------------------|
| DS1 | DS1 | 0.89 | M3 | 0.79 ± 0.002 |
| | DS2 | 0.81 | M2 | 0.79 ± 0.006 |
| | DS3 | 0.79 | M2 | 0.79 ± 0.003 |
| DS2 | DS1 | 0.86 | M2 | 0.84 ± 0.004 |
| | DS2 | 0.86 | M3 | 0.83 ± 0.009 |
| | DS3 | 0.86 | M4 | 0.84 ± 0.009 |
| DS3 | DS1 | 0.85 | M2 | 0.84 ± 0.002 |
| | DS2 | 0.85 | M2 | 0.84 ± 0.004 |
| | DS3 | 0.84 | M2 | 0.83 ± 0.007 |

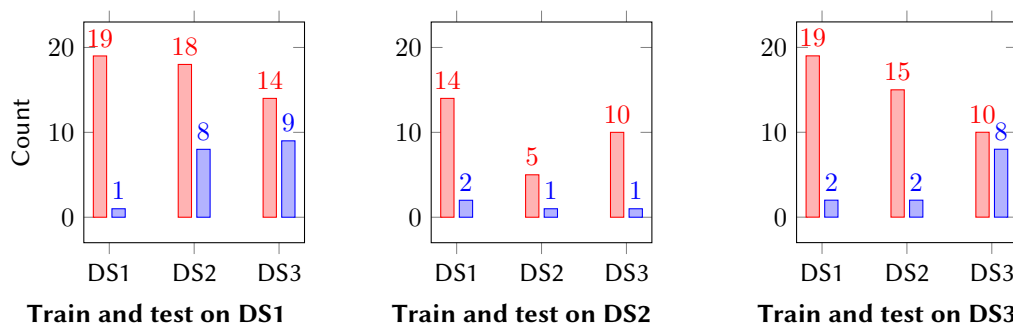


Figure 4: Number of (A, B) pairs used to obtain the best F-measures for the random methodology ■ and the best competency-based methodology ■ on experimental configuration C2, *i.e.*, training and testing on one dataset and selecting competent (A, B) pairs at inference from all datasets.

For DS1, using the competency-based methodology M2 with (A, B) pairs from DS1 yields the best results, as shown in Table 4. For pair-base compression (Figure 4), the number of analogies needed to achieve the best F-measure is also lower when retrieving (A, B) pairs from DS1. However, contrary to DS1, competent (A, B) pairs from the train dataset are not always the ones leading to the best results. Indeed, for DS2 and DS3, selecting (A, B) pairs from DS1 is as competitive with selecting from the train dataset. This observation highlights a potential notion of *competent dataset*, that may not only be associated to similarity between domains but potentially to other analogical characteristics. Thus, this result could emphasize the ability of analogical reasoning in extrapolating from one domain to another with minimal reliance on

domain-specific data or similarity. Note that the random methodology provides stable results that are always lower than the ones obtained with the competency-based approach. Similarly, using this dataset as a source of (A, B) pairs leads to results that are always below those obtained using DS1 or DS2. Again, this may be explained by the nature of DS3 that is more fine-grained in its decisions than other datasets.

Results obtained with configuration C3. For the experimental configuration C3, results in Table 5 are obtained by training a model on dataset DS1 and testing on another dataset while selecting competent (A, B) pairs at inference from the different datasets available. Additional experimental results with other training datasets are available in Appendices.

Table 5

Results obtained on the experimental configuration C3, *i.e.*, selecting competent (A, B) pairs of different domains, for a model trained on one dataset and tested on another dataset. For each dataset from which competent (A, B) pairs are selected, the best performing selection methodology and its associated weighted F-measure are reported.

| Train | Test | (A, B) pairs dataset | F-measure | Competency method | Random method |
|-------|------|------------------------|-------------|-------------------|------------------|
| DS1 | DS2 | DS1 | 0.67 | M2 | 0.53 ± 0.001 |
| | | DS2 | 0.66 | M2 | 0.53 ± 0.006 |
| | | DS3 | 0.55 | M4 | 0.52 ± 0.008 |
| DS1 | DS3 | DS1 | 0.71 | M2 | 0.60 ± 0.004 |
| | | DS2 | 0.61 | M2 | 0.60 ± 0.006 |
| | | DS3 | 0.60 | M2 | 0.60 ± 0.004 |

As seen for Configuration C2, selecting competent (A, B) pairs from the same dataset used for test does not necessarily lead to the best results: in Table 5, the highest weighted F-measure is achieved when competent pairs are selected from DS1. Even on DS3, where competency-based methodologies struggle to outperform random selection, selecting competent pairs from DS1 allows to obtain an 11-point gain. Interestingly, Figure 5 shows that for a model trained on one dataset and tested on another, selecting (A, B) pairs from one of these datasets only leads to a small reduction in the number of pairs needed, even if still under the number required for random selection. However, selecting (A, B) pairs from another dataset (DS3) leads to a drastic reduction, but a smaller F-measure. Note that additional results in Appendix also show a general ability for our competency-based methodologies to provide a compression gain for the pair-base.

6. Conclusion and Perspectives

This study has introduced several competency-based methodologies for the selection of (A, B) pairs to enhance performance in analogy-based machine learning tasks. By comparing each competency-based methodology with a naive, random selection approach on the task of knowledge graph pruning, we have empirically demonstrated the benefits of using some methodologies, both in terms of task performance and pair-base compression (*i.e.*, number of (A, B) pairs

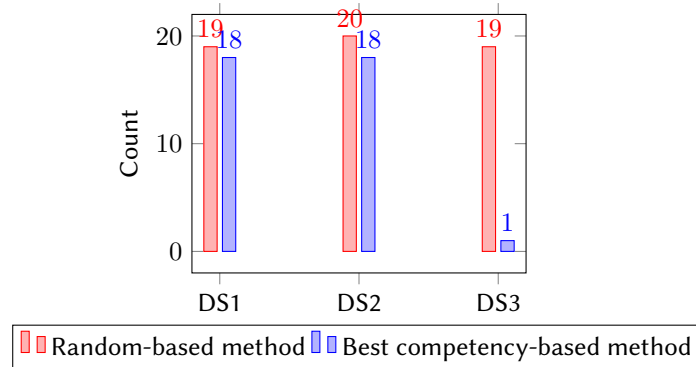


Figure 5: Number of (A, B) pairs needed from each dataset to obtain the best F-measure for a model trained on DS1 and tested on DS2 with (A, B) pairs coming from different domains (experimental configuration C3).

needed). Furthermore, the study reveals that the best analogical (A, B) pairs, in terms of task performance, are not necessarily those retrieved from the same domain dataset. To illustrate, using pairs from our dataset DS1 consistently yields the best results across all experimental configurations, also suggesting a better analogical potential for this dataset. This observation also emphasizes the ability of analogical reasoning to generalize across several domains. This cross-domain generalization and our results indicate that our approach could especially be useful to reuse an analogical model trained on a dataset for inference on a new dataset. Even though the proposed competency-based selection methodologies did not always enable reaching the best weighted F-measure compared to a fully retrained model, a high performance gain is still obtained, which could limit the need and cost of retraining an entire model on domain specific data. The compression benefits of our competency-based methodologies also contributes to this cost-effective aspect.

These findings suggest many directions for future work, and especially in extending competency-based approaches to other tasks. One direction could be the application of these selection methodologies to Natural Language Processing (NLP) where information is also embedded in vector spaces. Target Sense Verification (TSV) that has already been framed into an analogy detection task by [19] could be a relevant task. Finally, the robustness of the competency-based methodologies proposed and the distance metrics employed remains to be challenged, especially when applied with different embedding models (*i.e.*, DistMult [20], ConvE [21]), as all experiments in this study were conducted using TransE [18].

Acknowledgments

This work was supported by the French National Research Agency (ANR)⁵ project “Analogies: from theory to tools and applications” (AT2TA), ANR-22-CE23-0023.

⁵Agence Nationale de la Recherche.

References

- [1] S. D. Afantenos, H. Prade, G. Richard, L. C. Bernardes, Analogical proportions and creativity: A preliminary study, in: ICCC 2024, Association for Computational Creativity, 2024, pp. 289–298.
- [2] J. Thibaut, Y. Glady, R. M. French, Understanding the *What* and *When* of analogical reasoning across analogy formats: An eye-tracking and machine learning approach, *Cogn. Sci.* 46 (2022).
- [3] K. Combs, T. Bihl, S. Ganapathy, D. Staples, Analogical reasoning: An algorithm comparison for natural language processing, in: HICSS 2022, ScholarSpace, 2022, pp. 1–10.
- [4] B. Elayeb, M. Bounhas, M. F. Ettih, Morphological disambiguation of texts based on analogical proportions, in: ICAART 2025, SCITEPRESS, 2025, pp. 1207–1214.
- [5] L. Jarnac, M. Couceiro, P. Monnin, Relevant entity selection: Knowledge graph bootstrapping via zero-shot analogical pruning, in: CIKM 2023, ACM, 2023, pp. 934–944.
- [6] P. Monnin, M. Couceiro, Interactions between knowledge graph-related tasks and analogical reasoning: A discussion, in: ATA@ICCB2022, volume 3389 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 57–67.
- [7] L. Wang, H. Wang, Y. Lepage, AnaScore: understanding semantic parallelism in proportional analogies, in: L. Chiruzzo, A. Ritter, L. Wang (Eds.), NAACL 2025, Albuquerque, New Mexico, 2025, pp. 1175–1188.
- [8] E. Marquer, F. Badra, M. Lesot, M. Couceiro, D. Leake, Less is better: An energy-based approach to case base competence, in: ATA@ICCB2023, volume 3438 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 27–42.
- [9] B. Fuchs, J. Lieber, L. Miclet, A. Mille, A. Napoli, H. Prade, G. Richard, Case-based reasoning, analogy, and interpolation, in: *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*, Springer, 2020, pp. 307–339.
- [10] F. Lin, X. Zhu, Z. Zhao, D. Huang, Y. Yu, X. Li, T. Xu, E. Chen, Knowledge graph pruning for recommendation, *CoRR abs/2405.11531* (2024). doi:10.48550/ARXIV.2405.11531. arXiv:2405.11531.
- [11] J. L. Kolodner, An introduction to case-based reasoning, *Artificial Intelligence Review* 6 (1992) 3–34. doi:10.1007/BF00155578.
- [12] D. Carbone, Hitchhiker’s guide on energy-based models: a comprehensive review on the relation with other generative models, sampling and statistical physics, *CoRR abs/2406.13661* (2024). arXiv:2406.13661.
- [13] M. Bounhas, H. Prade, G. Richard, Analogy-based classifiers for nominal or numerical data, *International Journal of Approximate Reasoning* 91 (2017) 36–55.
- [14] A. Hogan, et al., *Knowledge Graphs, Synthesis Lectures on Data, Semantics, and Knowledge*, Morgan & Claypool Publishers, 2021.
- [15] S. Lim, H. Prade, G. Richard, Solving word analogies: A machine learning perspective, in: ECSQARU 2019, volume 11726 of *LNCS*, Springer, 2019, pp. 238–250.
- [16] N. Hubert, H. Paulheim, A. Brun, D. Monticolo, Do similar entities have similar embeddings?, *CoRR abs/2312.10370* (2023). arXiv:2312.10370.
- [17] M. Bounhas, H. Prade, Analogy-based classifiers: An improved algorithm exploiting competent data pairs, *International Journal of Approximate Reasoning* 158 (2023) 108923.

- [18] A. Bordes, N. Usunier, A. García-Durán, J. Weston, O. Yakhnenko, Translating embeddings for modeling multi-relational data, in: NIPS 2013, 2013, pp. 2787–2795.
- [19] G. Zervakis, E. Vincent, M. Couceiro, M. Schoenauer, E. Marquer, An analogy based approach for solving target sense verification, in: NLP4IR 2022, ACM, 2022, pp. 144–151.
- [20] B. Yang, W. Yih, X. He, J. Gao, L. Deng, Embedding entities and relations for learning and inference in knowledge bases, in: ICLR 2015, 2015.
- [21] M. Shi, J. Zhao, A knowledge graph link prediction model with combined 1d and 2d convolutional embeddings, in: DSAA 2022, IEEE, 2022, pp. 1–2.

A. Appendices

Table 6

Results obtained on the experimental configuration **C3**, *i.e.*, selecting competent (A, B) pairs of different domains, for a model trained on one dataset (DS2 or DS3) and tested on another dataset. For each dataset from which competent (A, B) pairs are selected, the best performing selection methodology and its associated weighted F-measure are reported.

| Train | Test | (A, B) pairs dataset | F-measure | Competency method | Random method |
|-------|------|------------------------|-------------|-------------------|------------------|
| DS2 | DS1 | DS1 | 0.53 | M2 | 0.51 ± 0.003 |
| | | DS2 | 0.53 | M2 | 0.49 ± 0.007 |
| | | DS3 | 0.53 | M4 | 0.50 ± 0.006 |
| DS2 | DS3 | DS1 | 0.53 | M2 | 0.49 ± 0.004 |
| | | DS2 | 0.53 | M2 | 0.49 ± 0.009 |
| | | DS3 | 0.52 | M4 | 0.50 ± 0.009 |
| DS3 | DS1 | DS1 | 0.59 | M2 | 0.51 ± 0.004 |
| | | DS2 | 0.51 | M3 | 0.50 ± 0.007 |
| | | DS3 | 0.55 | M4 | 0.51 ± 0.003 |
| DS3 | DS2 | DS1 | 0.53 | M2 | 0.50 ± 0.003 |
| | | DS2 | 0.51 | M2 | 0.50 ± 0.005 |
| | | DS3 | 0.51 | M2 | 0.51 ± 0.003 |

Table 7

Results obtained on the experimental configuration **C3**, *i.e.*, selecting competent (A, B) pairs of different domains, for a model trained on one dataset (DS2 or DS3) and tested on another dataset. For each dataset from which competent (A, B) pairs are selected, the best performing selection methodology and its associated number of (A, B) pairs are reported.

| Train | Test | (A, B) pairs dataset | $\#(A, B)$ pairs | Competency method | Random method |
|-------|------|------------------------|------------------|-------------------|---------------|
| DS2 | DS1 | DS1 | 2 | M2 | 20 |
| | | DS2 | 2 | M2 | 11 |
| | | DS3 | 2 | M4 | 17 |
| DS2 | DS3 | DS1 | 3 | M2 | 19 |
| | | DS2 | 1 | M2 | 16 |
| | | DS3 | 6 | M4 | 18 |
| DS3 | DS1 | DS1 | 2 | M2 | 19 |
| | | DS2 | 2 | M3 | 15 |
| | | DS3 | 8 | M4 | 10 |
| DS3 | DS2 | DS1 | 6 | M2 | 8 |
| | | DS2 | 4 | M2 | 4 |
| | | DS3 | 5 | M2 | 8 |