



# MMKG: Multi-modal Knowledge Graphs

Ye Liu<sup>1</sup>(✉), Hui Li<sup>2</sup>, Alberto Garcia-Duran<sup>3</sup>, Mathias Niepert<sup>4</sup>,  
Daniel Onoro-Rubio<sup>4</sup>, and David S. Rosenblum<sup>1</sup>

<sup>1</sup> National University of Singapore, Singapore, Singapore  
{liuye,david}@comp.nus.edu.sg

<sup>2</sup> Xiamen University, Xiamen, China  
huili.xmu@gmail.com

<sup>3</sup> EPFL, Lausanne, Switzerland  
alberto.duran@epfl.ch

<sup>4</sup> NEC Labs Europe, Heidelberg, Germany  
{mathias.niepert,daniel.onoro}@neclab.eu

**Abstract.** We present MMKG, a collection of three knowledge graphs that contain both numerical features and (links to) images for all entities as well as entity alignments between pairs of KGs. Therefore, multi-relational link prediction and entity matching communities can benefit from this resource. We believe this data set has the potential to facilitate the development of novel multi-modal learning approaches for knowledge graphs. We validate the utility of MMKG in the **sameAs** link prediction task with an extensive set of experiments. These experiments show that the task at hand benefits from learning of multiple feature types.

**Github Repository:** <https://github.com/nle-ml/mmkb>

**Permanent URL:** <https://zenodo.org/record/1245698>

## 1 Introduction

A large volume of human knowledge can be represented with a multi-relational graph. Binary relationships encode *facts* that can be represented in the form of RDF [11] type triples (**head**, **predicate**, **tail**), where **head** and **tail** are entities and **predicate** is the relation type. The combination of all triples form a multi-relational graph, where nodes represent entities and directed edges represent relationships. The resulting multi-relational graph is often referred to as a Knowledge Graph.

Knowledge Graphs (KGs) provide ways to efficiently organize, manage and retrieve this type of information, being increasingly used as external source of knowledge for a number of problems. While ranging from general purpose (DBPEDIA [2] or FREEBASE [3]) to domain-specific (IMDB or UNIPROTKB), KGs are often highly incomplete and, therefore, research has focused heavily on

---

Y. Liu, H. Li and A. Garcia-Duran—Contributed equally. Work done while at NEC Labs Europe.

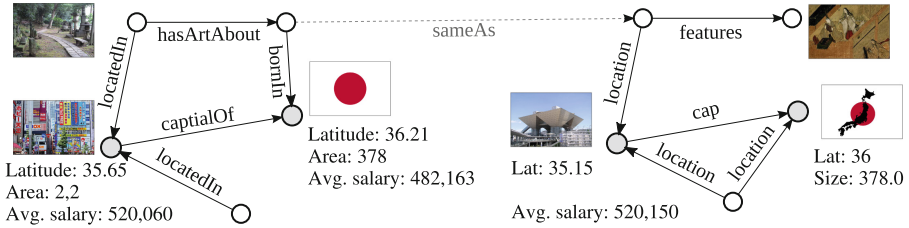
the problem of knowledge graph completion [15]. Link prediction [15], relationship extraction [18] and ontology matching [20] are some of the different ways to tackle the incompleteness problem.

Novel data sets for benchmarking knowledge graph completion approaches, therefore, are important contributions to the community. This is especially true since one method performing well on one data set might perform poorly on others [23]. With this paper we introduce MMKG (Multi-Modal Knowledge Graphs), a collection of three knowledge graphs for link prediction and entity matching research. Contrary to existing data sets, these knowledge graphs contain both numerical features and images for all entities as well as entity alignments between pairs of KGs. There is a fundamental difference between MMKG (Fig. 1) and other visual-relational resources (e.g. [12, 24]). While MMKG is intended to perform relational reasoning across different entities and images, previous resources are intended to perform visual reasoning within the same image.

We use FREEBASE15K [4] as the blue print for the multi-modal knowledge graphs we constructed. FREEBASE15K is the major benchmark data set in the recent link prediction literature. In a first step, we aligned most FB15k entities to entities from DBPEDIA and YAGO through the `sameAs` links contained in DBPEDIA and YAGO dumps. Since the degree of a node relates to the probability of an entity to appear in a subsampled version of a KG, we use this measure to populate our versions of DBPEDIA and YAGO with more entities. For each knowledge graph, we include entities that are highly connected to the aligned entities so that the number of entities in each KG is similar to that of FB15K. Lastly, we have populated the three knowledge graphs with numeric literals and images for (almost) all of their entities. We name the two new data sets DBPEDIA15K and YAGO15K. Although all three data sets contain a similar number of entities, this does not prevent potential users of MMKG from filtering out entities to benchmark approaches in scenarios where KGs largely differ with respect to the number of entities that they contain.

The contributions of the present paper are the following:

- The creation of two knowledge graphs DBPEDIA15K and YAGO15K, that are the DBPEDIA and YAGO [21] counterparts, respectively, of FREEBASE15K. Furthermore, all three KGs are enriched with numeric literals and image information, as well as `sameAs` predicates linking entities from pairs of knowledge graphs. `sameAs` predicates, numerical literals and (links to) images for entities so as the relational graph structure are released in separate files.
- We validate our hypothesis that knowledge graph completion related problems can benefit from multi-modal data:
  - We elaborate on a previous learning framework [9] and extend it by also incorporating image information. We perform completion in queries such as (`head?`, `sameAs`, `tail`) and (`head`, `sameAs`, `tail?`), where `head` and `tail` are entities, each one from a different KG. This task can be deemed something in-between link prediction and entity matching.
  - We analyze the performance of the different modalities in isolation for different percentages of known aligned entities between KGs, as well as for different combinations of feature types.



**Fig. 1.** Illustration of MMKG.

## 2 Relevance

There are a number of problems related to knowledge graph completion. *Named-entity linking* (NEL) [6] is the task of linking a named-entity mention from a text to an entity in a knowledge graph. Usually a NEL algorithm is followed by a second procedure, namely *relationship extraction* [18], which aims at linking relation mentions from text to a canonical relation type in a knowledge graph. Hence, relation extraction methods are often used in conjunction with NEL algorithms to perform KG completion from natural language content.

Link prediction and entity matching are two other popular tasks for knowledge graph completion. MMKG has been mainly created targeting these two tasks.

**Link Prediction.** It aims at answering completion queries of the form (head?, predicate, tail) or (head, predicate, tail?), where the answer is supposed to be always within the KG.

**Entity Matching.** Given two KGs, the goal is to find pairs of records, one from each KG, that refer to the same entity. For instance,  $DBpedia:NYC \equiv FB:NewYork$ .

### 2.1 Relevance for Multi-relational Link Prediction Research

The core of most of multi-relational link prediction approaches is a scoring function. The scoring function is a (differentiable) function whose parameters are learned such that it assigns high scores to true triples and low scores to triples assumed to be false. The majority of recent work fall into one of the following two categories:

1. Relational approaches [10, 14] wherein features are given as logical formulas which are evaluated in the KG to determine the feature's value. For instance, the formula  $\exists x (A, \text{bornIn}, x) \wedge (x, \text{capitalOf}, B)$  corresponds to a binary feature which is 1 if there exists a path of that type from entity A to entity B, and 0 otherwise.
2. Latent approaches [15] learn fixed-size vector representations (embeddings) for all entities and relationships in the KG.

While previous work has almost exclusively focused on the relational structure of the graph, recent approaches have considered other feature types like numerical literals [9, 17]. In addition, recent work on visual-relational knowledge graphs [16] has introduced novel visual query types such as “*How are these two unseen images related to each other?*” and has proposed novel machine learning methods to answer these queries. Different to the link prediction problem addressed in this work, the methods evaluated in [16] solely rely on visual data.

MMKG provides three data sets for evaluating multi-relational link prediction approaches where, in addition to the multi-relational links between entities, all entities have been associated with numerical and visual data. An interesting property of MMKG is that the three knowledge graphs are very heterogeneous (w.r.t. the number of relation types, their sparsity, and so on) as we show in Sect. 3. It is known that the performance of multi-relational link prediction methods depends on the characteristics of the specific knowledge graphs [23]. Therefore, MMKG is an important benchmark data set for measuring the robustness of the approaches.

## 2.2 Relevance for Entity Matching Research

There are numerous approaches to find `sameAs` links between entities of two different knowledge graphs. Though there are works [8] that solely incorporate the relational graph structure, there is an extensive literature on methods that perform the matching by combining relational structural information with literals of entities, where literals are used to compute prior confidence scores [13, 20].

A large number of approaches of the entity matching literature have been evaluated as part of the Ontology Alignment Evaluation Initiative (OAEI) [1] using data sets such as YAGO, FREEBASE, and IMDB [13, 20]. Contrary to the proposed multi-modal knowledge graph data sets, however, the OAEI does not focus on tasks with visual and numerical data. The main advantages of MMKG over existing benchmark data sets for entity matching are: (1) MMKG’s entities are associated with visual and numerical data, and (2) the availability of ground truth entity alignments for a high percentage of the KG entities. The former encourages research in entity matching methods that incorporate visual and numerical data. The latter allows one to measure the robustness in performance of entity matching approaches with respect to the number of given alignments between two KGs. The benchmark KGs can also be used to evaluate different active learning strategies. Traditional active learning approaches ask a user for a small set of alignments that minimize the uncertainty and, therefore, maximize the quality of the final alignments.

## 3 MMKG: Dataset Generation

We chose FREEBASE-15K (FB15K), a data set that has been widely used in the knowledge graph completion literature, as a starting point to create the multi-modal knowledge graphs. Facts of this KG are in N-Triples format, a line-based plain text format for encoding an RDF graph. For example, the triple

**Table 1.** Files from which we extract the different subcomponents of MMKG.

DB15k	
sameAs	<a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/freebase_links_en.ttl.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/freebase_links_en.ttl.bz2</a>
Relational Graph	<a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/mappingbased_objects_en.ttl.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/mappingbased_objects_en.ttl.bz2</a> <a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/instance_types_en.ttl.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/instance_types_en.ttl.bz2</a>
Numeric Literals	<a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/geo_coordinates_en.tql.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/geo_coordinates_en.tql.bz2</a> <a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/mappingbased_literals_en.tql.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/mappingbased_literals_en.tql.bz2</a> <a href="http://downloads.dbpedia.org/2016-10/core-i18n/en/persondata_en.tql.bz2">http://downloads.dbpedia.org/2016-10/core-i18n/en/persondata_en.tql.bz2</a>
YAGO15k	
sameAs	<a href="http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoDBpediaInstances.ttl.7z">http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoDBpediaInstances.ttl.7z</a>
Relational Graph	<a href="http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoFacts.ttl.7z">http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoFacts.ttl.7z</a>
Numeric Literals	<a href="http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoDateFacts.ttl.7z">http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoDateFacts.ttl.7z</a> <a href="http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoGeonamesOnlyData.ttl.7z">http://resources.mpi-inf.mpg.de/yago-naga/yago3.1/yagoGeonamesOnlyData.ttl.7z</a>

**Table 2.** Statistics of the MMKG knowledge graphs.

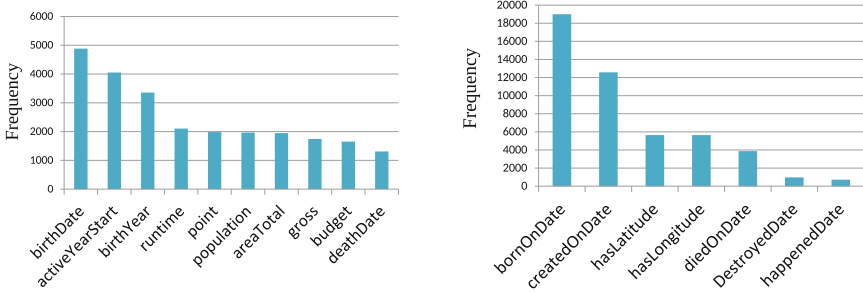
			Number of triples			
KG	#Entities	#Relationships	Relational Graph	Numeric Literals	Images	sameAs
FB15k	14,951	1,345	592,213	29,395	13,444	-
DB15k	14,777	279	99,028	46,121	12,841	12,846
Yago15k	15,283	32	122,886	48,405	11,194	11,199

</ns/g.112ygbz6> </ns/type.object.type> </ns/film.film>.

indicates that the entity with identifier </ns/g.112ygbz6> is connected to the entity with identifier </ns/film.film> via the relationship </ns/type.object.type>.

We create versions of DBPEDIA and YAGO, called DBPEDIA-15K (DB15K) and YAGO15K, by aligning entities in FB15K with entities in these other knowledge graphs. More concretely, for DB15K we performed the following steps.

1. **sameAs.** We extract alignments between entities of FB15K and DBPEDIA in order to create DB15K. These alignments link one entity from FB15K to one from DBPEDIA via a **sameAs** relation.
2. **Relational Graph.** A high percentage of entities from FB15K can be aligned with entities in DBPEDIA. However, to make the two knowledge graphs have roughly the same number of entities and to also have entities that cannot be aligned across the knowledge graphs, we include additional entities in DB15K. We chose entities with the highest connectivity to the already aligned entities to complete DB15K. We then collect all the triples where both **head** and **tail** entities belong to the set of entities of DB15K. This collection of triples forms the relational graph structure of DB15K.
3. **Numeric Literals.** We collect all triples that associate entities in DB15K with numerical literals. For example, the relations **/location/geocode/latitude** links entities to their latitude. We refer to these relation types as numerical relations. Figure 2 shows the most common numerical



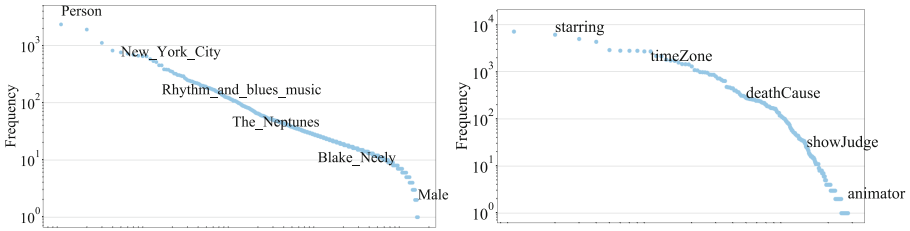
**Fig. 2.** Most common numerical relationships in DB15K (left) and YAGO15K (right).

relationships in the knowledge graphs. In previous work [9] we have extracted numeric literals for FB15K only.

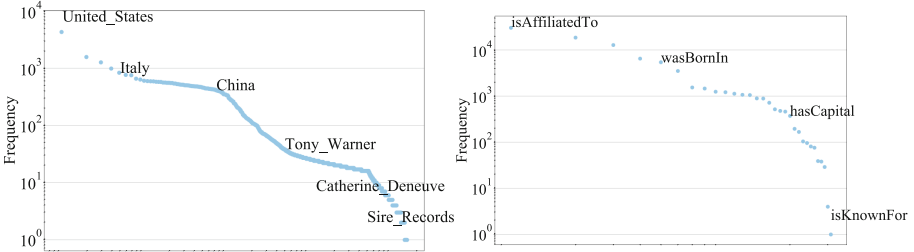
- Images.** We obtain images related to each of the entities of FB15K. To do so we implemented a web crawler that is able to parse query results for the image search engines Google Images, Bing Images, and Yahoo Image Search. To minimize the amount of noise due to polysemous entity labels (for example, there are two FREEBASE entities with the text label “Paris”) we extracted, for each entity in FB15K, all Wikipedia URIs from the 1.9 billion triple FREEBASE RDF dump<sup>1</sup>. For instance, for Paris, we obtained URIs such as `Paris(ile-de-France,France)` and `Paris(City_of_New_Orleans,Louisiana)`. These URIs were processed and used as search queries for disambiguation purposes. We crawled web images also following other type of search queries, and not only the Wikipedia URIs. For example, we used (i) the entity name, and (ii) the entity name followed by the entity’s notable type as query strings, among others. After visual inspection of polysemous entities (as they are the most problematic entities), we observed that using Wikipedia URIs as query strings was the strategy that alleviated most the polysemy problem. We used the crawler to download a number of images per entity. For each entity we stored the 20 top ranked images retrieved by each browser. We filtered out images with a side smaller than 224 pixels, and images with a side 2.5 bigger than the other. We also removed corrupted, low quality, and duplicate images (pairs of images with a pixel-wise distance below a certain threshold). After all these steps, we kept 55.8 images per entity on average. We also scaled the images to have a maximum height or width of 500 pixels while maintaining their aspect ratio. Finally, for each entity we distribute a distinct image to FB15K and DB15K.

We repeat the same sequence of steps for the creation of YAGO15K with one difference. `sameAs` predicates from the YAGO dump align entities from that knowledge graph to DBPEDIA entities. We used them along with the previously extracted alignments between DB15K and FB15K to eventually create the alignment between YAGO and FB15K entities. Table 1 depicts the hyperlinks from where we extracted the different component for the generation of DB15K and YAGO15K.

<sup>1</sup> <https://developers.google.com/freebase/>.



**Fig. 3.** Entity (left) and relation type (right) frequencies in YAGO15K.



**Fig. 4.** Entity (left) and relation type (right) frequencies in DB15K.

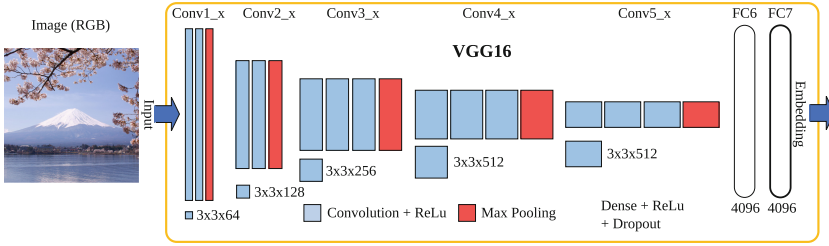
Statistics of FB15K, DB15K and YAGO15K are depicted in Table 2. The frequency of entities and relationships in YAGO15K and DB15K are depicted in Figs. 3 and 4, respectively. Entities and relationships are sorted according to their frequency. They show in logarithmic scale the number of times that each entity and relationship occurs in YAGO15K and DB15K. Relationships like `starring` or `timeZone` occur quite frequently in YAGO15K, while others like `animator` are rare. Contrary to FB15K, the entity `Male` is unusual in YAGO15K, which illustrates, to a limited extent, the heterogeneity of the KGs.

### 3.1 Availability and Sustainability

MMKG can be found in the Github repository <https://github.com/nle-ml>. We will actively use Github issues to track feature requests and bug reports. The documentation of the framework has been published on the repository’s Wiki as well. To guarantee the future availability of the resource, it has also been published on Zenodo. MMKG is released under the BSD-3-Clause License.

The repository contains a number of files, all of them formatted following the N-Triples guidelines (<https://www.w3.org/TR/n-triples/>). These files contain information regarding the relational graph structure, numeric literals and visual information. Numerical information is formatted as RDF literals, entities and relationships point to their corresponding RDF URIs<sup>2</sup>. We also provide separate files that link both DB15K and YAGO15K entities to FB15K ones via `sameAs` predicates, also formatted as N-Triples.

<sup>2</sup> Unfortunately, Freebase has deprecated RDF URIs.



**Fig. 5.** Low-dimensional embeddings learned for images through VGG16.

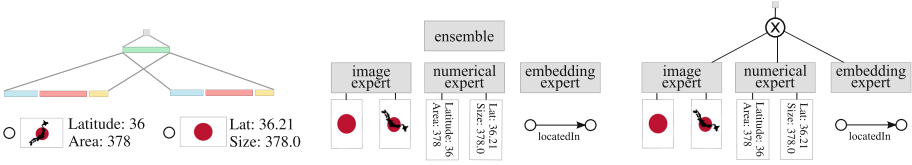
To avoid copyright infringement and guarantee the access to the visual information (i.e. URLs to images are not permanent), we learn embeddings for the images through the VGG16 model introduced in [19]. The VGG16 model used for this work was previously trained on the ILSVRC 2012 data set derived from IMAGENET [5]. The architecture of this network is illustrated in Fig. 5. We remove the softmax layer of the trained VGG16 and obtain the 4096-dimensional embeddings for all images of MMKG. We provide these embeddings in hdf5 [22] format. The Github repository contains documentation on how to access these embeddings. Alternatively, one can use the crawler (also available in the Github repository) to download the images from the different search engines.

## 4 Technical Quality of MMKG

We provide empirical evidence that knowledge graph completion related tasks can benefit from the multi-modal data of MMKG. Our hypothesis is that different data modalities contain complementary information beneficial for both multi-relational link prediction and entity matching. For instance, in the entity matching problem if two images are visually similar they are likely to be associated with the same entity and if two entities in two different KGs have similar numerical feature values, they are more likely to be identical. Similarly, we hypothesize that multi-relational link prediction can benefit from the different data modalities. For example, learning that the mean difference of birth years is 0.4 for the Freebase relation `/people/marriage/spouse`, can provide helpful evidence for the linking task.

In recent years, numerous methods for merging feature types have been proposed. The most common strategy is the concatenation of either the input features or some intermediate learned representation. We compare these strategies to the recently proposed learning framework [9], which we have found to be superior to the concatenation and an ensemble type of approach.





**Fig. 6.** Illustration of the methods we evaluated to combine various data modalities.

#### 4.1 Task: SAMEAS Link Prediction

We validate the hypothesis that different modalities are complementary for the `sameAs` link prediction task. Different to the standard link prediction problem, here the goal is to answer queries such as  $(\text{head}?, \text{sameAs}, \text{tail})$  or  $(\text{head}, \text{sameAs}, \text{tail}?)$  where `head` and `tail` are entities from different KGs. We do not make the one-to-one alignment assumption, that is, the assumption that one entity in one KG is identical to exactly (at most) one in the other. A second difference is that in the evaluation of the SAMEAS prediction task, and in general in the link prediction literature, *only* one argument of a triple is assumed to be missing at a time. That partial knowledge of the ground truth is not given in the entity matching literature.

#### 4.2 Model: Products of Experts

We elaborate on previous work [9] and extend it by incorporating visual information. Such learning framework can be stated as a Product of Experts (PoE).

In general, a PoE’s probability distribution is

$$p(\mathbf{d} \mid \theta_1, \dots, \theta_n) = \frac{\prod_i f_i(\mathbf{d} \mid \theta_i)}{\sum_{\mathbf{c}} \prod_i f_i(\mathbf{c} \mid \theta_i)},$$

where  $\mathbf{d}$  is a data vector in a discrete space,  $\theta_i$  are the parameters of individual model  $f_i$ ,  $f_i(\mathbf{d} \mid \phi_i)$  is the value of  $d$  under model  $f_i$ , and the  $\mathbf{c}$ ’s index all possible vectors in the data space. The PoE model is now trained to assign high probability to observed data vectors.

In the KG context, the data vector  $\mathbf{d}$  is always a triple  $\mathbf{d} = (\mathbf{h}, \mathbf{r}, \mathbf{t})$  and the objective is to learn a PoE that assigns high probability to true triples and low probabilities to triples assumed to be false. For instance, the triple  $(\text{Paris}, \text{locatedIn}, \text{France})$  should be assigned a high probability and the triple  $(\text{Paris}, \text{locatedIn}, \text{Germany})$  a low probability. If  $(\mathbf{h}, \mathbf{r}, \mathbf{t})$  holds in the KG, the pair’s vector representations are used as positive training examples. Let  $\mathbf{d} = (\mathbf{h}, \mathbf{r}, \mathbf{t})$ . We can now define one individual expert  $f_{(\mathbf{r}, \mathbf{F})}(\mathbf{d} \mid \phi_{(\mathbf{r}, \mathbf{F})})$  for each (relation type  $\mathbf{r}$ , feature type  $\mathbf{F}$ ) pair

$$\begin{aligned} f_{(\mathbf{r}, \mathbf{L})}(\mathbf{d} \mid \theta_{(\mathbf{r}, \mathbf{L})}) &: \text{the embedding expert for relation type } \mathbf{r} \\ f_{(\mathbf{r}, \mathbf{R})}(\mathbf{d} \mid \theta_{(\mathbf{r}, \mathbf{R})}) &: \text{the relational expert for relation type } \mathbf{r} \\ f_{(\mathbf{r}, \mathbf{N})}(\mathbf{d} \mid \theta_{(\mathbf{r}, \mathbf{N})}) &: \text{the numerical expert for relation type } \mathbf{r} \\ f_{(\text{sameAs}, \mathbf{I})}(\mathbf{d} \mid \theta_{(\mathbf{r}, \mathbf{I})}) &: \text{the visual expert for relation type } \text{sameAs} \end{aligned}$$

The joint probability for a triple  $\mathbf{d} = (\mathbf{h}, \mathbf{r}, \mathbf{t})$  of the PoE model is now

$$p(\mathbf{d} \mid \theta_1, \dots, \theta_n) = \frac{\prod_{F \in \{\mathbf{R}, \mathbf{L}, \mathbf{N}, \mathbf{I}\}} f_{(r,F)}(\mathbf{d} \mid \theta_{(r,F)})}{\sum_{\mathbf{c}} \prod_{F \in \{\mathbf{R}, \mathbf{L}, \mathbf{N}, \mathbf{I}\}} f_{(r,F)}(\mathbf{c} \mid \theta_{(r,F)})},$$

where  $\mathbf{c}$  indexes all possible triples.

For information regarding the latent, relational and numerical experts, we refer the reader to [9]. Although entity names are not used to infer `sameAs` links in this work, one may also define an expert for such feature.

**Visual Experts.** The visual expert is only learned for the `sameAs` relation type. The scores for the image experts is computed by the cosine similarity between two 4096-dimensional feature vectors from the two images.

Let  $\mathbf{d} = (\mathbf{h}, \mathbf{r}, \mathbf{t})$  be a triple. The visual expert for relation type  $\mathbf{r}$  is defined as

$$\begin{aligned} f_{(r,I)}(\mathbf{d} \mid \theta_{(r,I)}) &= \exp(\mathbf{i}_{\mathbf{h}} \cdot \mathbf{i}_{\mathbf{t}}) \text{ and} \\ f_{(r',I)}(\mathbf{d} \mid \theta_{(r',I)}) &= 1 \text{ for all } \mathbf{r}' \neq \mathbf{r}, \end{aligned}$$

where  $\cdot$  is the dot product and  $\mathbf{i}_{\mathbf{h}}$  and  $\mathbf{i}_{\mathbf{t}}$  are embeddings of the images for the head and tail entities.

**Learning.** The logarithmic loss for the given training triples  $\mathbf{T}$  is defined as

$$\mathcal{L} = - \sum_{\mathbf{t} \in \mathbf{T}} \log p(\mathbf{t} \mid \theta_1, \dots, \theta_n).$$

To fit the PoE to the training triples, we follow the derivative of the log likelihood of each observed triple  $\mathbf{d} \in \mathbf{T}$  under the PoE

$$\frac{\partial \log p(\mathbf{d} \mid \theta_1, \dots, \theta_n)}{\partial \theta_m} = \frac{\partial \log f_i(\mathbf{d} \mid \theta_i)}{\partial \theta_m} - \frac{\partial \log \sum_{\mathbf{c}} \prod_i f_i(\mathbf{c} \mid \theta_i)}{\partial \theta_m}$$

We follow [9] and we generate for each triple  $\mathbf{d} = (\mathbf{h}, \mathbf{r}, \mathbf{t})$  a set  $\mathbf{E}$  consisting of  $N$  triples  $(\mathbf{h}, \mathbf{r}, \mathbf{t}')$  by sampling exactly  $N$  entities  $\mathbf{t}'$  uniformly at random from the set of all entities. In doing so, the right term is then approximated by

$$\frac{\partial \log \sum_{\mathbf{c} \in \mathbf{E}} \prod_i f_i(\mathbf{c} \mid \theta_i)}{\partial \theta_m}.$$

This is often referred to as negative sampling.

### 4.3 Additional Baseline Approaches

Apart from the product of experts, we also evaluate other approaches to combine various data modalities. All the evaluated approaches are illustrated in Fig. 6.

**Concatenation.** Given pairs of aligned entities, each pair is characterized by a single vector wherein all modality features of both entities are concatenated. For each pair of aligned entities we create a number of negative alignments, each of which is also characterized by a concatenation of all modality features of both entities. A logistic regression is trained taking these vectors as input, and their corresponding class label (+1 and -1 for positive and negative alignments, respectively). The output of the logistic regression indicates the posterior probability of two entities being the same. In Sect. 5 we refer to this approach as **CONCAT**.

**Ensemble.** The ensemble approach combines the various expert models into an ensemble classifier. Instead of training the experts jointly and end-to-end, here each of the expert models is first trained independently. At test time, the scores of the expert models are added and used to rank the entities. We refer to this approach as **ENSEMBLE**.

## 5 Experiments

We conducted experiments on two pairs of knowledge graphs of MMKG, namely, (FB15K *vs.* DB15K and YAGO15K *vs.* FB15K). We evaluate a number of different instances of the product of experts (PoE) model, as well as the other baseline methods, in the **sameAs** prediction task. Because of its similarity with link prediction, we use metrics commonly used for this task. The main objective of the experiments is to demonstrate that MMKG is suitable for the task at hand, and specifically that the related problems can benefit from learning of multiple feature types.

### 5.1 Evaluation

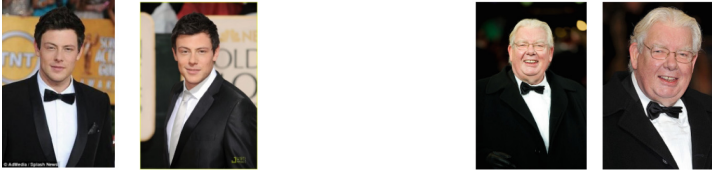
MMKG allows to experiment with different percentages of aligned entities between KGs. These alignments are given by the **sameAs** predicates that we previously found. We evaluate the impact of the different modalities in scenarios wherein the number of given alignments  $P$  [%] between two KGs is low, medium and high. We reckon that such scenarios would correspond to 20%, 50% and 80% out of all **sameAs** predicates, respectively. We use these alignments along with the two KGs as part of our observed triples  $\mathbf{T}$ , and split equally the remaining **sameAs** triples into validation and test.

We use AMIE+ [7] to mine relational features for the relational experts. We used the standard settings of AMIE+ with the exception that the minimum absolute support was set to 2 and the maximum number of entities involved in the rule to four. The latter is important to guarantee that AMIE+ retrieves rules like  $(x, \mathbf{r}_1, w), (w, \mathbf{SameAs}, z), (z, \mathbf{r}_2, y) \Rightarrow (x, \mathbf{SameAs}, y)$ , wherein  $\mathbf{r}_1$  is a relationship that belongs to the one KG, and  $\mathbf{r}_2$  to the other KG. One example of retrieved rule by AMIE+ is:

$$(x, \mathbf{father\_of}_{\text{DB15k}}, w), (w, \mathbf{SameAs}, z), (z, \mathbf{children\_of}_{\text{FB15k}}, y) \Rightarrow (x, \mathbf{SameAs}, y)$$

**Table 3.** sameAs queries for which numerical experts led to good performance. Left and right column correspond to FB15k and DB15k, respectively.

/m/015dcj		Marc_Christian	
date_of_birth	1925.11	birthDate	1925.11
date_of_death	1985.10	deathDate	1985.10
height_meters	1.93	height	1.9558
/m/07zhjj		How_i_met_your_mother	
number_of_seasons	9.0	numberOfSeasons	9.0
air_date_of_final_episode	2014.03	completionDate	2014.03
number_of_episodes	208.0	numberOfEpisodes	208.0



**Fig. 7.** sameAs queries for which visual experts led to good performance. Left and right images within each pair correspond to FB15k and DB15k, respectively.

In this case both  $\text{father\_of}_{\text{DB15k}}$  and  $\text{children\_of}_{\text{FB15k}}$  are (almost) functional relationships. A relationship  $r$  is said to be functional if an entity can only be mapped exactly to one single entity via  $r$ . The relational expert will learn that the body of this rule leads to a sameAs relationship between entities  $x$  and  $y$  and with a very high likelihood.

We used ADAM for parameter learning in a mini-batch setting with a learning rate of 0.001, the categorical cross-entropy as loss function and the number of epochs was set to 100. We validated every 5 epochs and stopped learning whenever the MRR (Mean Reciprocal Rank) values on the validation set decreased. The batch size was set to 512 and the number  $N$  of negative samples to 500 for all experiments.

We follow the same evaluation procedure as previous works of the link prediction literature. Therefore, we measure the ability to answer completion queries of the form  $(h, \text{SameAs}, t?)$  and  $(h?, \text{SameAs}, t)$ . For queries of the form  $(h, \text{SameAs}, t?)$ , wherein  $h$  is an entity of the first KG, we replaced the tail by each of the second KB’s entities in turn, sorted the triples based on the scores or probabilities, and computed the rank of the correct entity. We repeated the

**Table 4.** sameAs prediction on FB15K-DB15K for different percentages of  $P$ .

$P$ [%]	20%			50%			80%		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
PoE- $n$	12.8	10.1	18.6	23.0	16.8	34.1	28.2	21.8	37.4
PoE- $r$	12.9	10.7	16.5	26.3	22.9	31.7	35.9	33.6	38.6
PoE- $l$	12.2	7.9	20.3	42.8	34.9	58.2	63.1	55.6	76.6
PoE- $i$	1.6	0.8	2.7	2.3	1.3	3.8	3.3	1.7	5.9
PoE- $lni$	16.7	12.0	25.6	48.1	40.9	62.1	68.5	62.0	79.3
PoE- $rni$	<b>28.3</b>	<b>23.2</b>	<b>39.0</b>	44.2	38.0	55.7	55.8	50.2	64.1
PoE- $lri$	12.5	8.8	19.1	40.4	33.4	53.9	67.0	60.3	78.3
PoE- $lrn$	16.0	11.5	24.1	47.9	41.2	60.3	67.1	60.6	79.1
PoE- $lrni$	17.0	12.6	25.1	<b>53.3</b>	<b>46.4</b>	<b>65.8</b>	<b>72.1</b>	<b>66.6</b>	<b>82.0</b>

**Table 5.** sameAs prediction on FB15K-YAGO15K for different percentages of  $P$ .

$P$ [%]	20%			50%			80%		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
PoE- $n$	22.2	15.4	33.7	38.9	29.7	56.1	35.8	27.1	53.5
PoE- $r$	9.9	8.4	12.3	20.0	18.0	23.1	29.9	28.1	31.9
PoE- $l$	10.1	6.4	16.9	32.0	25.8	44.1	50.5	43.7	63.6
PoE- $i$	1.4	2.4	0.7	2.0	1.1	3.2	3.2	1.7	5.5
PoE- $lni$	15.4	10.9	24.1	39.8	32.8	52.6	59.0	52.5	70.5
PoE- $rni$	<b>33.4</b>	<b>25.0</b>	<b>49.5</b>	<b>49.8</b>	<b>41.1</b>	<b>66.9</b>	57.2	49.2	70.5
PoE- $lri$	11.3	7.7	18.1	34.0	28.1	44.7	55.5	49.3	66.7
PoE- $lrn$	13.9	10.2	20.9	37.3	31.6	47.4	57.7	51.3	68.9
PoE- $lrni$	15.4	11.3	22.9	41.4	34.7	53.6	<b>63.5</b>	<b>57.3</b>	<b>74.6</b>

same process for the queries of type  $(h?, \text{SameAs}, t)$ , wherein  $t$  in this case corresponds to an entity of the second KG and we iterate over the entities of the first KG to compute the scores. The mean of all computed ranks is the Mean Rank (lower is better) and the fraction of correct entities ranked in the top  $n$  is called hits@ $n$  (higher is better). We also compute the Mean Reciprocal Rank (higher is better) which is an evaluation metric that is less susceptible to outliers. Note that the filtered setting described in [4] does not make sense in this problem, since an entity can be linked to an entity via a SameAs relationship only once.

We report the performance of the PoE in its full scope in Tables 4 and 5. We also show feature ablation experiments, each of which corresponds to removing one modality from the full set. The performance of each modality in isolation is also depicted. We use the abbreviations PoE-*suffix* to refer to the different

**Table 6.** Performance comparison for  $P = 80\%$ .

		MRR	HITS@1	HITS@10
FB-DB	CONCAT	2.1	1.7	2.7
	ENSEMBLE	40.1	34.3	50.2
	PoE- <i>lrni</i>	<b>72.1</b>	<b>66.6</b>	<b>82.0</b>
FB-YAGO	CONCAT	0.18	0.18	0.04
	ENSEMBLE	47.6	42.3	57.5
	PoE- <i>lrni</i>	<b>63.5</b>	<b>57.3</b>	<b>74.6</b>

instances of PoE. *suffix* is a combination of the letters L (Latent), R (Relational), N (Numerical) and I (Image) to indicate the inclusion of each of the four feature types. Generalizations are complicated to make, given that performance of PoE’s instances differ across percentages of aligned entities and pairs of knowledge graphs. Nevertheless, there are two instances of our PoE approach, PoE-*lrni* and PoE-*rni*, that tend to outperform all others for low and high percentages of aligned entities, respectively. Results seem to indicate that the embedding expert response dominates over others, and hence its addition to PoE harms the performance when such expert is not the best-performing one. Table 3 and Fig. 7 provides examples of queries where numerical and visual information led to good performance, respectively. It is hard to find one specific reason that explains when adding numerical and visual information is beneficial for the task at hand. For example, there are entities with a more canonical visual representation than others. This relates to the difficulty of learning from visual data in the `sameAs` link prediction problem, as visual similarity largely varies across entities. Similarly, the availability of numerical attributes largely varies even for entities of the same type within a KG. However, Tables 4 and 5 provide empirical evidence of the benefit from including additional modalities.

Table 6 depicts results for the best-performing instance of PoE and baselines discussed in Sect. 4. The best performing instance of PoE significantly outperforms the approaches CONCAT and ENSEMBLE. This validates the choice of the PoE approach, which can incorporate data modalities to the link prediction problem in a principled manner.

## 6 Conclusion

We present MMKG, a collection of three knowledge graphs that contain multi-modal data, to benchmark link prediction and entity matching approaches. An extensive set of experiments validate the utility of the data set in the `sameAs` link prediction task.

## References

1. Achichi, M., et al.: Results of the ontology alignment evaluation initiative 2016. In: OM: Ontology Matching, pp. 73–129 (2016). No commercial editor
2. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-76298-0\\_52](https://doi.org/10.1007/978-3-540-76298-0_52)
3. Bollacker, K.D., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: SIGMOD Conference, pp. 1247–1250 (2008)
4. Bordes, A., Usunier, N., García-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS, pp. 2787–2795 (2013)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
6. Dredze, M., McNamee, P., Rao, D., Gerber, A., Finin, T.: Entity disambiguation for knowledge base population. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 277–285. Association for Computational Linguistics (2010)
7. Galárraga, L., Teflioudi, C., Hose, K., Suchanek, F.M.: Fast rule mining in ontological knowledge bases with AMIE+. *VLDB J.* **24**(6), 707–730 (2015)
8. Galárraga, L.A., Preda, N., Suchanek, F.M.: Mining rules to align knowledge bases. In: Proceedings of the 2013 Workshop on Automated Knowledge Base Construction, pp. 43–48. ACM (2013)
9. Garcia-Duran, A., Niepert, M.: KBLRN: end-to-end learning of knowledge base representations with latent, relational, and numerical features. In: Uncertainty in Artificial Intelligence Proceedings of the 34th Conference (2018)
10. Gardner, M., Mitchell, T.M.: Efficient and expressive knowledge base completion using subgraph feature extraction. In: EMNLP, pp. 1488–1498 (2015)
11. Klyne, G., Carroll, J.J., McBride, B.: Resource description framework (RDF): concepts and abstract syntax. W3C Recommendation, February 2004
12. Krishna, R., et al.: Visual genome: connecting language and vision using crowd-sourced dense image annotations (2016). <https://arxiv.org/abs/1602.07332>
13. Lacoste-Julien, S., Palla, K., Davies, A., Kasneci, G., Graepel, T., Ghahramani, Z.: SiGMa: simple greedy matching for aligning large knowledge bases. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 572–580. ACM (2013)
14. Lao, N., Mitchell, T., Cohen, W.W.: Random walk inference and learning in a large scale knowledge base. In: EMNLP, pp. 529–539 (2011)
15. Nickel, M., Murphy, K., Tresp, V., Gabrilovich, E.: A review of relational machine learning for knowledge graphs. *Proc. IEEE* **104**(1), 11–33 (2016)
16. Oñoro-Rubio, D., Niepert, M., García-Durán, A., González-Sánchez, R., López-Sastre, R.J.: Representation learning for visual-relational knowledge graphs. arXiv preprint [arXiv:1709.02314](https://arxiv.org/abs/1709.02314) (2017)
17. Pezeshkpour, P., Chen, L., Singh, S.: Embedding multimodal relational data for knowledge base completion. In: EMNLP (2018)
18. Riedel, S., Yao, L., McCallum, A., Marlin, B.M.: Relation extraction with matrix factorization and universal schemas. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 74–84 (2013)

19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
20. Suchanek, F.M., Abiteboul, S., Senellart, P.: PARIS: probabilistic alignment of relations, instances, and schema. *Proc. VLDB Endow.* **5**(3), 157–168 (2011)
21. Suchanek, F.M., Kasneci, G., Weikum, G.: YAGO: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*, pp. 697–706. ACM (2007)
22. The HDF Group: Hierarchical Data Format, version 5 (1997-NNNN). <http://www.hdfgroup.org/HDF5/>
23. Toutanova, K., Chen, D.: Observed versus latent features for knowledge base and text inference. In: *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pp. 57–66 (2015)
24. Wu, Q., Wang, P., Shen, C., Dick, A., van den Hengel, A.: Ask me anything: free-form visual question answering based on knowledge from external sources. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4622–4630 (2016)