

# LKGE: Large-scale Knowledge Graph Embedding

## Abstract

There have been some effective embedding methods proposed for knowledge graphs in recent years. However, these existing methods are trained and evaluated in toy datasets. In reality, knowledge graphs often contain millions of entities or even more. This paper studies the problem of embedding very large realistic knowledge graphs into continuous low-dimensional vector spaces, which is useful in many tasks such as question answering and web searching. In this paper, we propose a novel large-scale knowledge graph embedding method called the “LKGE” based on the previous works, which overcome some difficulties exposed during large-scale knowledge graph embedding processing: (1) We use a parallel training method to combine neural networks and knowledge graph embedding methods so that text features can be fastly encoded to support knowledge graph embedding. The previous works usually adopt the serial training method, which is quite slow even on toy datasets. (2) Knowledge graphs in the real-world commonly have highly skewed power-law degree distributions, which challenge the efficient training for every entities and relations meanwhile. We propose an edge-sampling algorithm to improve the model optimization so that entities and relations can participate into training in different proportion. (3) We also divide the entire knowledge graph into some subgraphs and propose a graph-parallel curriculum learning framework based on the graph division. This lead to a high-speed and precise multiple threads optimization. Empirical experiments both on a benchmark toy dataset and a real knowledge graph “Wikidata” prove the effectiveness of the LKGE. And the experiments also prove our model is very efficient, which is able to learn large-scale knowledge graph embeddings in a few hours on a single machine without GPUs and distributed computing. The source code of the LKGE would be released in the public then.

## 1 Introduction

The *IJCAI-17 Proceedings* will be printed from electronic manuscripts submitted by the authors. These must be PDF (*Portable Document Format*) files formatted for 8-1/2" × 11" paper.

### 1.1 Length of Papers

Each accepted full paper is allocated six pages in the conference proceedings, excluded references. References can take up to one page. Up to two additional pages may be purchased at a price to be announced per page for any accepted paper. However, all *submissions* must be a maximum of six pages, plus at most one for references, in length.

### 1.2 Word Processing Software

As detailed below, IJCAI has prepared and made available a set of  $\text{\LaTeX}$  macros and a Microsoft Word template for use in formatting your paper. If you are using some other word processing software (such as WordPerfect, etc.), please follow the format instructions given below and ensure that your final paper looks as much like this sample as possible.

## 2 Style and Format

$\text{\LaTeX}$  and Word style files that implement these instructions can be retrieved electronically. (See Appendix A for instructions on how to obtain these files.)

### 2.1 Layout

Print manuscripts two columns to a page, in the manner in which these instructions are printed. The exact dimensions for pages are:

- left and right margins: .75"
- column width: 3.375"
- gap between columns: .25"
- top margin—first page: 1.375"
- top margin—other pages: .75"
- bottom margin: 1.25"
- column height—first page: 6.625"
- column height—other pages: 9"

All measurements assume an 8-1/2" × 11" page size. For A4-size paper, use the given top and left margins, column width, height, and gap, and modify the bottom and right margins as necessary.

## 2.2 Format of Electronic Manuscript

For the production of the electronic manuscript, you must use Adobe's *Portable Document Format* (PDF). A PDF file can be generated, for instance, on Unix systems using `ps2pdf` or on Windows systems using Adobe's Distiller. There is also a website with free software and conversion services: <http://www.ps2pdf.com/>. For reasons of uniformity, use of Adobe's *Times Roman* font is strongly suggested. In L<sup>A</sup>T<sub>E</sub>X2e, this is accomplished by putting

```
\usepackage{times}
```

in the preamble.<sup>1</sup>

Additionally, it is of utmost importance to specify the American **letter** format (corresponding to 8-1/2" × 11") when formatting the paper. When working with `dvips`, for instance, one should specify `-t letter`.

## 2.3 Title and Author Information

Center the title on the entire width of the page in a 14-point bold font. Below it, center the author name(s) in a 12-point bold font, and then center the address(es) in a 12-point regular font. Credit to a sponsoring agency can appear on the first page as a footnote.

### Blind Review

In order to make blind reviewing possible, authors must omit their names and affiliations when submitting the paper for review. In place of names and affiliations, provide a list of content areas. When referring to one's own work, use the third person rather than the first person. For example, say, "Previously, Gottlob [?] has shown that...", rather than, "In our previous work [?], we have shown that..." Try to avoid including any information in the body of the paper or references that would identify the authors or their institutions. Such information can be added to the final camera-ready version for publication.

## 2.4 Abstract

Place the abstract at the beginning of the first column 3" from the top of the page, unless that does not leave enough room for the title and author information. Use a slightly smaller width than in the body of the paper. Head the abstract with "Abstract" centered above the body of the abstract in a 12-point bold font. The body of the abstract should be in the same font as the body of the paper.

The abstract should be a concise, one-paragraph summary describing the general thesis and conclusion of your paper. A reader should be able to learn the purpose of the paper and the reason for its importance from the abstract. The abstract should be no more than 200 words long.

---

<sup>1</sup>You may want also to use the package `latexsym`, which defines all symbols known from the old L<sup>A</sup>T<sub>E</sub>X version.

## 2.5 Text

The main body of the text immediately follows the abstract. Use 10-point type in a clear, readable font with 1-point leading (10 on 11).

Indent when starting a new paragraph, except after major headings.

## 2.6 Headings and Sections

When necessary, headings should be used to separate major sections of your paper. (These instructions use many headings to demonstrate their appearance; your paper should have fewer headings.)

### Section Headings

Print section headings in 12-point bold type in the style shown in these instructions. Leave a blank space of approximately 10 points above and 4 points below section headings. Number sections with arabic numerals.

### Subsection Headings

Print subsection headings in 11-point bold type. Leave a blank space of approximately 8 points above and 3 points below subsection headings. Number subsections with the section number and the subsection number (in arabic numerals) separated by a period.

### Subsubsection Headings

Print subsubsection headings in 10-point bold type. Leave a blank space of approximately 6 points above subsubsection headings. Do not number subsubsections.

### Special Sections

You may include an unnumbered acknowledgments section, including acknowledgments of help from colleagues, financial support, and permission to publish.

Any appendices directly follow the text and look like sections, except that they are numbered with capital letters instead of arabic numerals.

The references section is headed "References," printed in the same style as a section heading but without a number. A sample list of references is given at the end of these instructions. Use a consistent format for references, such as that provided by BibT<sub>E</sub>X. The reference list should not include unpublished work.

## 2.7 Citations

Citations within the text should include the author's last name and the year of publication, for example [?]. Append lower-case letters to the year in cases of ambiguity. Treat multiple authors as in the following examples: [?] or [?] (for more than two authors) and [?] (for two authors). If the author portion of a citation is obvious, omit it, e.g., Nebel [?]. Collapse multiple citations as follows: [?; ?].

## 2.8 Footnotes

Place footnotes at the bottom of the page in a 9-point font. Refer to them with superscript numbers.<sup>2</sup> Separate them from the text by a short line.<sup>3</sup> Avoid footnotes as much as possible; they interrupt the flow of the text.

---

<sup>2</sup>This is how your footnotes should appear.

<sup>3</sup>Note the line separating these footnotes from the text.

### 3 Illustrations

Place all illustrations (figures, drawings, tables, and photographs) throughout the paper at the places where they are first discussed, rather than at the end of the paper. If placed at the bottom or top of a page, illustrations may run across both columns.

Illustrations must be rendered electronically or scanned and placed directly in your document. All illustrations should be in black and white, as color illustrations may cause problems. Line weights should be 1/2-point or thicker. Avoid screens and superimposing type on patterns as these effects may not reproduce well.

Number illustrations sequentially. Use references of the following form: Figure 1, Table 2, etc. Place illustration numbers and captions under illustrations. Leave a margin of 1/4-inch around the area covered by the illustration and caption. Use 9-point type for captions, labels, and other text in illustrations.

### Acknowledgments

The preparation of these instructions and the  $\text{\LaTeX}$  and Bib $\text{\TeX}$  files that implement them was supported by Schlumberger Palo Alto Research, AT&T Bell Laboratories, and Morgan Kaufmann Publishers. Preparation of the Microsoft Word file was supported by IJCAI. An early version of this document was created by Shirley Jowell and Peter F. Patel-Schneider. It was subsequently modified by Jennifer Balentine and Thomas Dean, Bernhard Nebel, and Daniel Pagenstecher. These instructions are the same as the ones for IJCAI-05, prepared by Kurt Steinkraus, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Lab.

### A $\text{\LaTeX}$ and Word Style Files

The  $\text{\LaTeX}$  and Word style files are available on the IJCAI-17 website, <http://www.ijcai-17.org/>. These style files implement the formatting instructions in this document.

The  $\text{\LaTeX}$  files are `ijcai17.sty` and `ijcai17.tex`, and the Bib $\text{\TeX}$  files are `named.bst` and `ijcai17.bib`. The  $\text{\LaTeX}$  style file is for version 2e of  $\text{\LaTeX}$ , and the Bib $\text{\TeX}$  style file is for version 0.99c of Bib $\text{\TeX}$  (*not* version 0.98i). The `ijcai17.sty` file is the same as the `ijcai07.sty` file used for IJCAI-07.

The Microsoft Word style file consists of a single file, `ijcai17.doc`. This template is the same as the one used for IJCAI-07.

These Microsoft Word and  $\text{\LaTeX}$  files contain the source of the present document and may serve as a formatting sample.

Further information on using these styles for the preparation of papers for IJCAI-17 can be obtained by contacting [pcchair@ijcai-17.org](mailto:pcchair@ijcai-17.org).