Proceedings of TIAD-2019 Shared Task – Translation Inference Across Dictionaries

Language, Data and Knowledge conference (LDK)
Leipzig, Germany, 20 May 2019

The second shared task for Translation Inference Across Dictionaries (TIAD 2019) is aimed at exploring methods and techniques for automatically generating new bilingual (and multilingual) dictionaries from existing ones. Although considerable work has been done to this end in the past, it was usually conducted on different types of datasets and evaluated in different ways, applying various algorithms that are often not comparable. Thus, the main aim of TIAD is to support a coherent experiment framework that enables reliable validation of results and solid comparison of the processes used. This initiative also aims to enhance further research on the topic of inferring translations across languages, and continues the first TIAD edition, that took place in the first Language Data and Knowledge (LDK-2017) conference on June 18, 2017 in Galway, Ireland (*https://tiad2017.wordpress.com/*).

The results of this second TIAD shared task were presented in a workshop that took place in Leipzig, Germany, on May 20, 2019 co-located with the 2nd Language, Data and Knowledge conference (LDK-2019). These proceedings assemble the works that were presented in this event. See *https://tiad2019.unizar.es/* for more information.

Jorge Gracia
Besim Kabashi
Ilan Kernerman
TIAD 2019 organisers

The shared task

The objective of the TIAD shared task was to explore and compare methods and techniques that infer translations indirectly between language pairs, based on other bilingual resources. Such techniques would help in auto-generating new bilingual and multilingual dictionaries based on existing ones.

In this edition, the evaluation data selection was different from that in the first edition, and the overall process was largely simplified. In particular, the participating systems were asked to generate new translations automatically among three languages, English, French, Portuguese, based on known translations contained in the Apertium RDF graph (http://linguistic.linkeddata.es/apertium/). As these languages (EN, FR, PT) are not directly connected in this graph, no translations can be obtained directly among them there. Based on the available RDF data, the participants had to apply their methodologies to derive translations, mediated by any other language in the graph, between the pairs EN/FR, FR/PT and PT/EN.

Participants could also make use of other freely available sources of background knowledge (e.g. lexical linked open data and parallel corpora) to improve performance, as long as no direct translation among the studied language pairs were available.

The evaluation of the results was carried out by the organisers against manually compiled pairs of K Dictionaries, extracted from its Global Series (https://www.lexicala.com/), particularly the following pairs: BR-EN, EN-BR, FR-EN, EN-FR, FR-PT, PT-FR. The translation pairs extracted from these dictionaries served as a golden standard and remained blind to the participants. Notice that the Brazilian Portuguese variant was used for the translations to/from English, which might introduce a bias; however its influence should be equivalent to every participant system thus still allowing for a valid comparison.

Results

Eleven systems participated in the shared task, coming from four teams. Differently to the first edition, all of them were able to complete the evaluation. The participant teams submitted a system description paper including: a description of their system or systems, the way data was processed, the applied algorithms, the obtained results, as well as the conclusions and ideas for future improvements. The system papers were peer reviewed by members of the reviewing committee to confirm that all these aspects were well covered.

An additional paper was submitted as a regular paper describing a system that did not participate in the shared task but that described a technique that was relevant for the topic of the workshop. This paper was also peer reviewed on the basis of its scientific quality.

Finally, the proceedings include a sixth paper, which consists of a summary in which the TIAD organisers give more details about the task, the evaluation data, and the system results.

The list of works published in the proceedings is as follows:

Summary paper:
● Jorge Gracia, Besim Kabashi, Ilan Kernerman, Marta Lanau-Coronas, and Dorielle Lonke. "Results of the Translation Inference Across Dictionaries 2019 Shared Task"

Regular paper:
● Mihael Arcan, Daniel Torregrosa, Sina Ahmadi, and John P. McCrae. "Inferring translation candidates for multilingual dictionary generation with multi-way neural machine translation"

System description papers:
● Daniel Torregrosa, Mihael Arcan, Sina Ahmadi, and John P. McCrae. "TIAD 2019 shared task: Leveraging knowledge graphs with neural machine translation for automatic multilingual dictionary generation"
● Marcos Garcia, Marcos García Salido, and Miguel A. Alonso. "Exploring cross-lingual word embeddings for the inference of bilingual dictionaries"
● Kathrin Donandt and Christian Chiarcos. "Translation inference through multi-lingual word embedding similarity"
● John P. McCrae "TIAD Shared Task 2019: orthonormal explicit topic analysis for translation inference across dictionaries"

Organisation

Organisers

- Jorge Gracia, University of Zaragoza, Spain
- Besim Kabashi, Friedrich-Alexander University of Erlangen-Nuremberg and Ludwig-Maximilian University of Munich, Germany
- Ilan Kernerman, K Dictionaries, Tel Aviv, Israel

Review committee

Julia Bosque-Gil, Universidad Politécnica de Madrid, Spain
Thierry Fontenelle, Translation Centre for the Bodies of the EU, Luxembourg
Jorge Gracia, University of Zaragoza, Spain
Besim Kabashi. Friedrich-Alexander University of Erlangen-Nuremberg & Ludwig-Maximilian University of Munich, Germany
Ilan Kernerman, K Dictionaries, Israel
Nikola Ljubešić, University of Zagreb, Croatia
Shervin Malmasi, Harvard University, USA
Elena Montiel-Ponsoda, Universidad Politécnica de Madrid, Spain
John McCrae, National University of Ireland, Galway
Georg Rehm, German Research Center for Artificial Intelligence (DFKI), Berlin
Arvi Tavast, Institute of the Estonian Language, Tallinn
Liling Tan, Saarland University, Germany & Nanyang Technological University, Singapore
Marcos Zampieri, University of Köln, Germany

Workshop schedule

9:00 - 9:30	Jorge Gracia, Besim Kabashi, Ilan Kernerman, Marta Lanau, and Dorielle Lonke: *2nd TIAD shared task description and overview of the results*
9:30 - 9:45	Mihael Arcan, Daniel Torregrosa, Sina Ahmadi, and John P. McCrae: *Inferring translation candidates for multilingual dictionary generation with multi-way neural machine translation*
9:45 - 10:00	Daniel Torregrosa, Mihael Arcan, Sina Ahmadi, and John P. McCrae: *TIAD 2019 Shared Task: leveraging knowledge graphs with neural machine translation for automatic multilingual dictionary generation*
10:00 -10:30	Coffee break
10:30 -10:55	Marcos Garcia, Marcos García Salido, and Miguel A. Alonso: *Exploring cross-lingual word embeddings for the inference of bilingual dictionaries*
10:55 -11:20	Kathrin Donandt and Christian Chiarcos: *Translation inference through multi-lingual word embedding similarity*
11:20 -11:45	John P. McCrae: *TIAD Shared Task 2019: orthonormal explicit topic analysis for translation inference across dictionaries*
11:45 - 12:00	Jorge Gracia and Besim Kabashi: *Closing words on TIAD 2019 shared task*

Acknowledgments