Unsupervised Semantic Frame Induction using Triclustering: Supplementary Materials

Dmitry Ustalov[†], **Alexander Panchenko**[‡], **Andrei Kutuzov**[⋆], **Chris Biemann**[‡], and **Simone Paolo Ponzetto**[†]

[†]University of Mannheim, Germany
{dmitry,simone}@informatik.uni-mannheim.de
[⋆]University of Oslo, Norway
andreku@ifi.uio.no
[‡]University of Hamburg, Germany
{panchenko,biemann}@informatik.uni-hamburg.de

Abstract

We use dependency triples automatically extracted from a Web-scale corpus to perform unsupervised semantic frame induction. We cast the frame induction problem as a *triclustering* problem that is a generalization of clustering for *triadic* data. Our replicable benchmarks demonstrate that the proposed graph-based approach, *Triframes*, shows state-of-the art results on this task on a FrameNet-derived dataset and performing on par with competitive methods on a verb class clustering task. **This document contains supplementary materials to the main paper**.

1 Triple Vector Representation

Figure 1 illustrates our approach for triple vector representation. In our representation, given a syntactic subject-verb-object (SVO) triple (people, make, money), we concatenate the word embeddings corresponding to these words into a single vector representing the whole triple. This explains the core assumption underlying in the Triframes approach: *triples representing similar roles appear in similar contexts*.

2 Implementation Details

We use a parallel implementation of the Watset[1] algorithm in Java for graph clustering, the Gensim[2] library for handling word embeddings, and the Faiss[3] library for indexing of word embeddings and retrieval of nearest neighbors. The source code and the data presented in this paper are available online under a permissive license.[4]

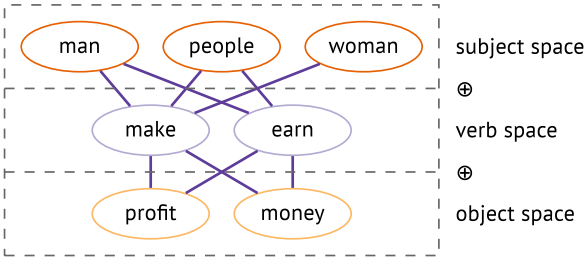

Figure 1: Concatenation of the vectors corresponding to the SVO triple elements expresses structural similarity of the triples.

Method	# of clusters
Triframes Watset	37,535
HOSG	10,000
NOAC	46,984
Triadic Spectral	500
Triadic k-means	500
Triframes CW	1862
LDA-Frames	109
Singletons	648,432
Whole	1

Table 1: Number of induced frames.

3 Cluster Sizes

Table 1 shows the amount of clusters produced by clustering algorithms during the frame induction experiment. Note that the Singletons baseline produced a distinct cluster for each triple and yet received low scores on each scale.

[1] https://github.com/dustalov/watset-java
[2] https://radimrehurek.com/gensim/
[3] https://github.com/facebookresearch/faiss
[4] https://github.com/uhh-lt/triframes

4 Examples of Induced Frames

Figures 2, 3 and 4 demonstrate examples of "good" frames, i.e. those which are semantically plausible according to our human judgment during a post-hoc manual analysis of clustering results. Figures 5, 6 and 7 show examples of "bad" frames according to the same criteria. All the frames are produced by the Triframes WATSET method ranked best as according to the Frame F_1 in the frame induction experiment. In particular, the number of nearest neighbors is $n = 30$, and the WATSET[CW_{top}, CW_{top}] fuzzy clustering algorithm has been used. These frames are available in the file `triw2v-watset-n30-top-top-triples.txt` available in the "Downloads" section of our GitHub repository (cf. Section 2).

Frame # 848	
Subjects:	Company, firm, company
Verbs:	buy, supply, discharge, purchase, expect
Objects:	book, supply, house, land, share, company, grain, which, item, product, ticket, work, this, equipment, House, it, film, water, something, she, what, service, plant, time

Figure 2: An example of a "good" frame.

Frame # 849	
Subjects:	student, scientist, we, pupil, member, company, man, nobody, you, they, US, group, it, people, Man, user, he
Verbs:	do, test, perform, execute, conduct
Objects:	experiment, test

Figure 3: An example of a "good" frame.

Frame # 3207	
Subjects:	people, we, they, you
Verbs:	feel, seek, look, search
Objects:	housing, inspiration, gold, witness, partner, accommodation, Partner

Figure 4: An example of a "good" frame.

Frame # 1	
Subjects:	you, she, he, return, they, we, themselves, road, help, who
Verbs:	govern, discourage, resemble, encumber, urge, pummel, *...911 more verbs ...*, demolish, swarm, anticipate, spew, derail, emit, snap
Objects:	you, pass, she, he, it, product, change, solution, total, any, wall, they, something, people, classic, this, interest, itself, flat, place, part, controversy

Figure 5: An example of a "bad" frame.

Frame # 852	
Subjects:	Word, glue, pill, speed, drug, pot, they, those, mine, item, resource, this, its, it, something, most, horse, material, chemical, plant, information, word
Verbs:	use, attach, apply, follow
Objects:	we, they, you, it, report, he

Figure 6: An example of a "bad" frame.

Frame # 37535	
Subjects:	he
Verbs:	phone, book
Objects:	you

Figure 7: An example of a "bad" frame.

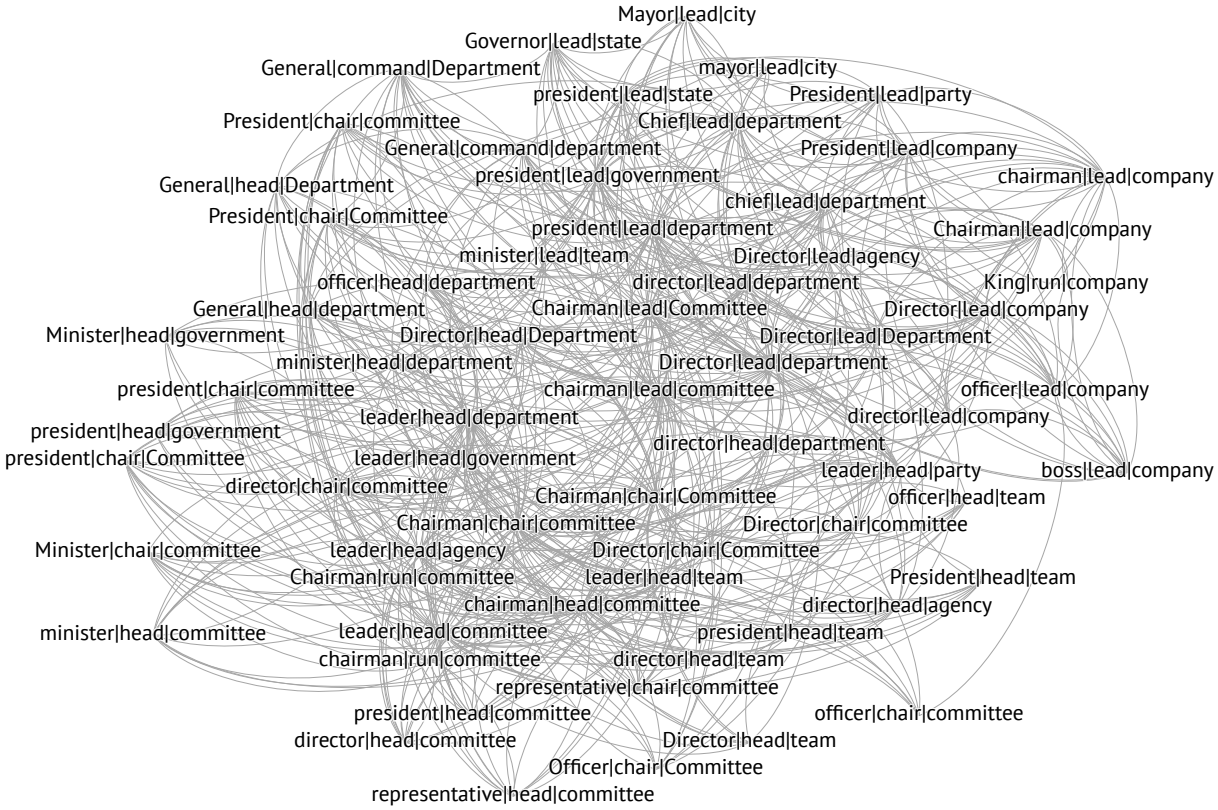

Figure 8: Visualization of an SVO triple graph, where edges represent distributional relatedness of the triples estimated using word embeddings.

5 Visualization of Triple Graph

Figure 8 presents a densly connected part of the triple graph related to the concept of "leadership". A similar cluster of triples can represent a semantic frame induced automatically from text using our approach.