Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Alexander Panchenko

FROM UNSUPERVISED INDUCTION OF LINGUISTIC STRUCTURES FROM TEXT TOWARDS APPLICATIONS IN DEEP LEARNING

Induction of semantic frames

FrameNet: frame “Kidnapping”

Definition:

The words in this frame describe situations in which a Perpetrator carries off and holds the Victim against his or her will by force.

Two men KIDNAPPED a Millwall soccer club employee, police said last night.

Not even the ABDUCTION of his children by Captain Hook and his scurvy sidekick, Smee, can shake Peter's scepticism.

FEs:

Core:

Perpetrator [Perp] **Semantic Type:** Sentient	The Perpetrator is the person (or other agent) who carries off and holds the Victim against his or her will.
Victim [Vict] **Semantic Type:** Sentient	The Victim is the person who is carried off and held against his/her will.

Non-Core:

Co-participant [Co-p]	An additional abductee taken along with the Victim. She was ABDUCTED with her brother.

Frame induction as a triclustering

- **ACL'2018** [Ustalov et al., 2018]

Example of a LU tricluster corresponding to the "Kidnapping" frame from FrameNet.

FrameNet	Role	Lexical Units (LU)
Perpetrator	Subject	kidnapper, alien, militant
FEE	Verb	snatch, kidnap, abduct
Victim	Object	son, people, soldier, child

Triframes frame induction

Require: an embedding model $v \in V \rightarrow \vec{v} \in \mathbb{R}^d$,
Require: a set of SVO triples $T \subseteq V^3$,
Require: the number of nearest neighbors $k \in \mathbb{N}$,
Require: a graph clustering algorithm **CLUSTER**.

Triframes frame induction

Require: an embedding model $v \in V \rightarrow \vec{v} \in \mathbb{R}^d$,
Require: a set of SVO triples $T \subseteq V^3$,
Require: the number of nearest neighbors $k \in \mathbb{N}$,
Require: a graph clustering algorithm **CLUSTER**.
Ensure: a set of triframes F.

Triframes frame induction

Require: an embedding model $v \in V \rightarrow \vec{v} \in \mathbb{R}^d$,
Require: a set of SVO triples $T \subseteq V^3$,
Require: the number of nearest neighbors $k \in \mathbb{N}$,
Require: a graph clustering algorithm **CLUSTER**.
Ensure: a set of triframes F.

$S \leftarrow \{t \rightarrow \vec{t} \in \mathbb{R}^{3d} : t \in T\}$
$E \leftarrow \{(t, t') \in T^2 : t' \in \mathrm{NN}_k^S(\vec{t}), t \neq t'\}$
$F \leftarrow \emptyset$
for all $C \in$ **CLUSTER**(T, E) **do**
  $f_s \leftarrow \{s \in V : (s, v, o) \in C\}$
  $f_v \leftarrow \{v \in V : (s, v, o) \in C\}$
  $f_o \leftarrow \{o \in V : (s, v, o) \in C\}$
  $F \leftarrow F \cup \{(f_s, f_v, f_o)\}$
return F

Evaluation datasets

Dataset	# instances	# unique	# clusters
FrameNet Triples	99,744	94,170	383
Poly. Verb Classes	246	110	62

Evaluation settings

Dataset	# instances	# unique	# clusters
FrameNet Triples	99,744	94,170	383
Poly. Verb Classes	246	110	62

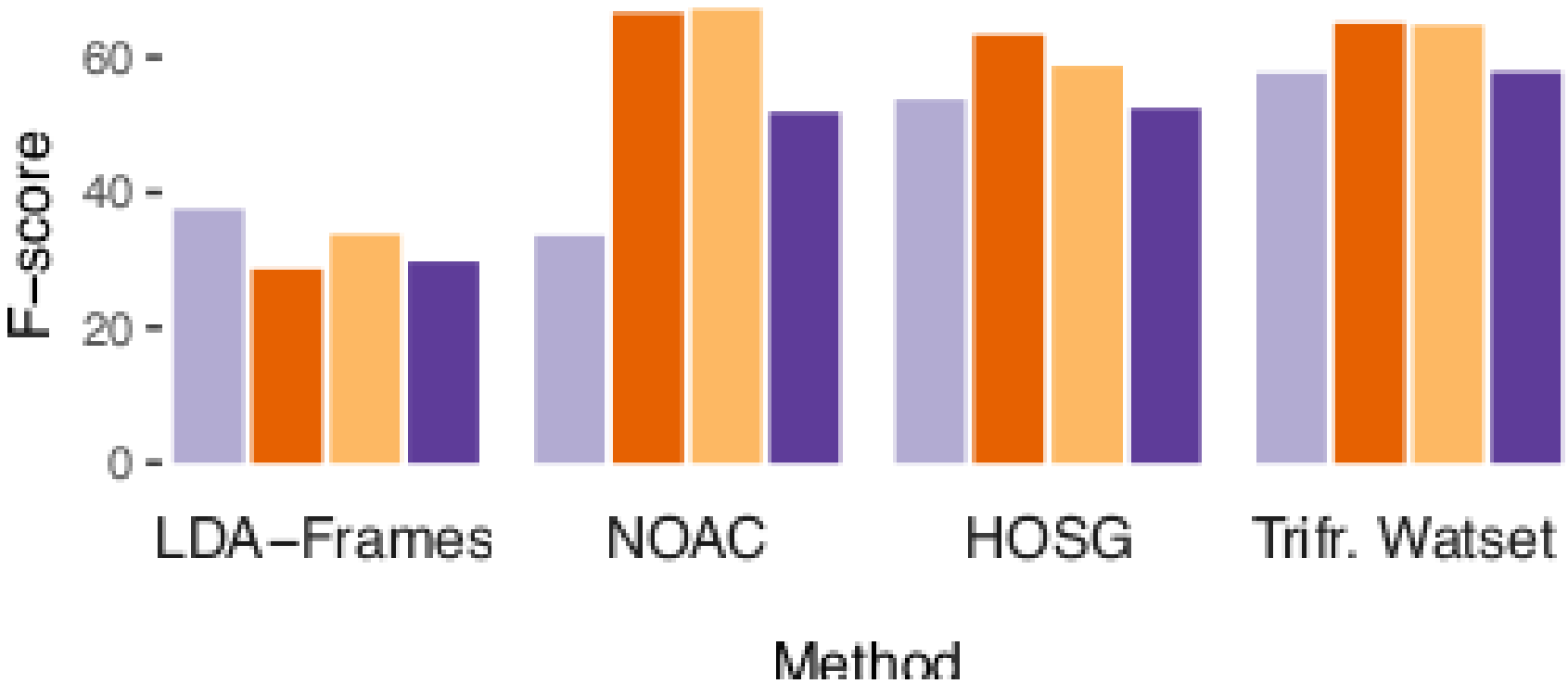
Evaluation settings

Dataset	# instances	# unique	# clusters
FrameNet Triples	99,744	94,170	383
Poly. Verb Classes	246	110	62

Quality Measures:

- nmPU: normalized modified purity,
- niPU: normalized inverse purity.

Results: comparison to state-of-art



F_1-scores for verbs, subjects, objects, frames


Ustalov, D., Panchenko, A., Kutuzov, A., Biemann, C., & Ponzetto, S. P. (2018).
Unsupervised semantic frame induction using triclustering.
arXiv preprint arXiv:1805.04715.