CALAMR: Component ALignment for Abstract Meaning Representation

Paul Landes, Barbara Di Eugenio
Department of Computer Science
University of Illinois Chicago
{plande2, bdieugen}@uic.edu

Overview

Previous Work

Method

Scoring

Results

Contributions

- Novel approach to aligning abstract meaning representation (AMR) graphs.
- New summarization based scoring methods for similarity of AMR subgraphs composed of one or more sentences.
- The entire reusable source code to reproduce our results.

Previous Work

Method

Scoring

Results

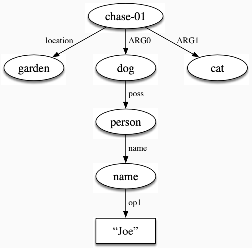
Abstract Meaning Representation (AMR)

AMR captures "who is doing what to whom" as semantic representation language that describes the abstract meaning of a sentence[1].

Example: *"Joe's dog was chasing a cat in the garden."*

```
# ::snt Joe's dog was chasing a cat in...
(c / chase-01~e.4
   :ARG0 (d / dog~e.2
            :poss (p / person
                     :name (n / name
                              :op1 "Joe")))
   :ARG1 (c2 / cat~e.6)
   :location~e.7 (g / garden~e.9))
```

Penman Notation



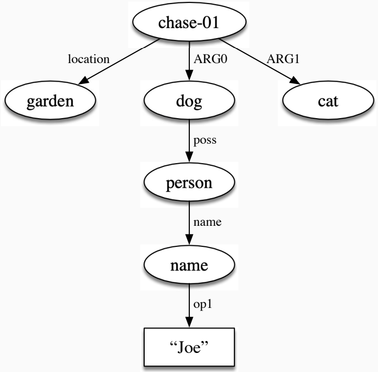
Directed Acyclic Graph

[1] Banarescu et al., "Abstract Meaning Representation for Sembanking" 2013 [1]

Propbank

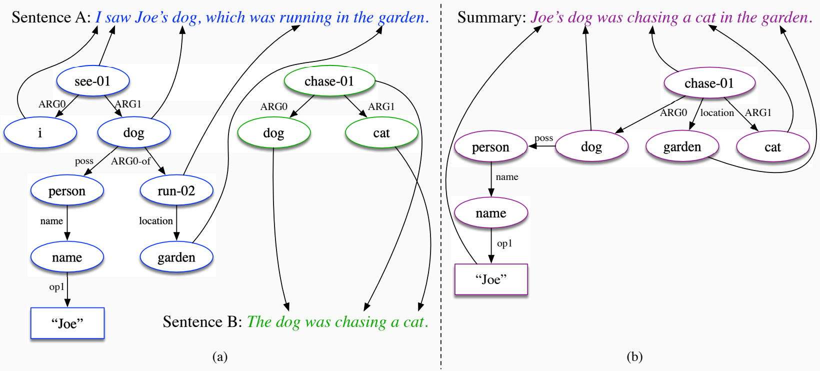
PropBank[2]: verb database of word senses.

- Roleset: `chase-01`
- Role 0 (`ARG0`)
  - Description: follower
  - Function: prototypical agent
  - Concept instance: dog
- Role 1 (`ARG1`)
  - Description: thing followed
  - Function: prototypical patient
  - Concept instance: cat



[2] Kingsbury et al., "From TreeBank to PropBank" 2002 [5]

AMR Token to Graph Alignment

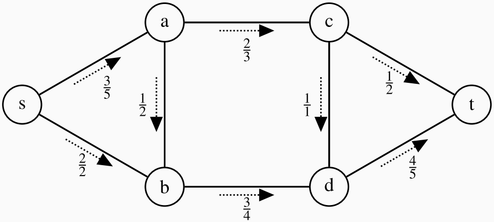


Example taken from Liu et al. 2015[3].

[3] Liu et al., "Toward Abstractive Summarization Using Semantic Representations" 2015 [7]

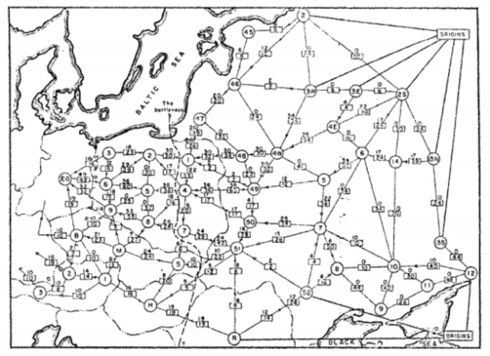
What is a flow network?

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	a flow network is a graph
$f : \mathcal{E} \rightarrow \mathbb{R}$	a s-t flow that assigns a flow
$\forall e, \in \mathcal{E}, 0 \leq f(e) \leq c_e$	capacity constraint
$\forall v, \in \mathcal{V} \sum_{\text{e into v}} f(e) = \sum_{\text{e leaving v}} f(e)$	conservation of flow constraint



Flow Network Origins

1. 1930: A. N. Tolstoĭ on Soviet railroad planning[4]
2. 1962: Ford-Fulkerson max-cut min-flow algorithm[5]



[4] Schrijver, "On the History of the Transportation and Maximum Flow Problems" 2002 [8]
[5] Ford et al., "Flows in Networks" 1962 [3]

Previous Work

Method

Scoring

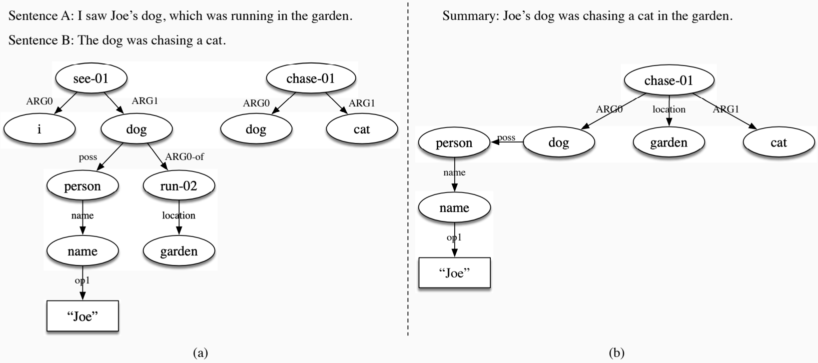
Results

AMR Graph Alignment

"Graph component alignment": the process of connecting two separate components together as a connected bipartite graph.

Create AMRs

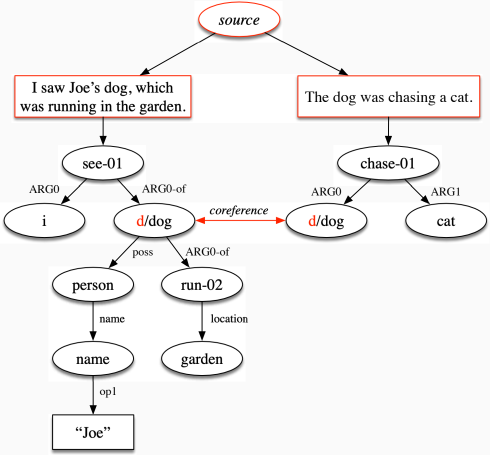
Human annotated AMR graphs from the “proxy report” used[6].



[6] Knight et al., *Abstract Meaning Representation (AMR) Annotation Release 3.0* 2021 [6]

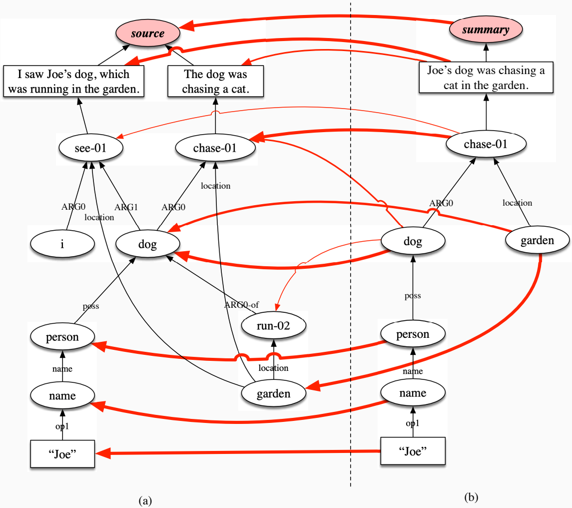
Create Graph Components

Create source and summary components.



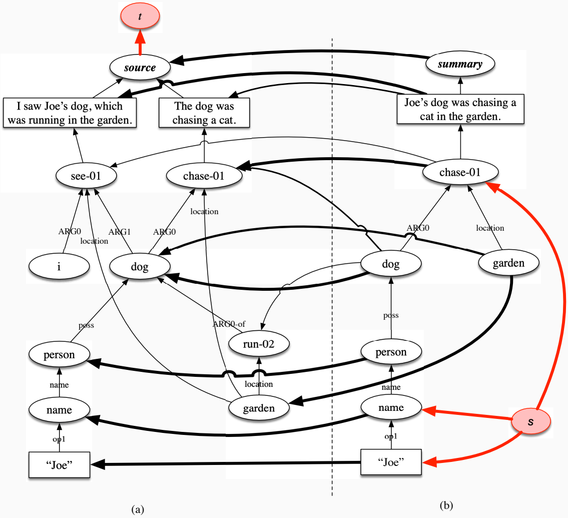
Add Alignment Edges

Connect the source and summary components as a bipartite graph.



Finalize the Flow Network

Add source and sink flow network nodes.



Compute Flow

Compute alignment edge capacities by:

1. Attach graph embeddings from text-to-graph tokens and PropBank on concept verb nodes.
2. Compute and set alignment edge capacities.
3. Run the max flow algorithm.
4. Normalize flow-per-node.
5. Clamp and remove low flow alignment edges.
6. Go to step 3 until convergence: flow remains static.

AMR Graph Alignment: Global vs. Local

Where do alignment edge capacities come from?

- Local:
  - PropBank entries
  - Text-to-graph alignment tokens
  - Sentence dampening
- Global:
  - Max flow algorithm
  - Subtree capacity constriction

Previous Work

Method

Scoring

Results

Apply the flow definition to the score

Our flow network once again: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

$$\forall v, \in \mathcal{V} \sum_{\text{e into v}} f(e) = \sum_{\text{e leaving v}} f(e) \tag{1}$$

$$\forall e, \in \mathcal{E}, 0 \leq f(e) \leq c_e, \; v(f) \triangleq f^{\text{out}}(s) \tag{2}$$

Using Equation 2, the value of flow exiting the source node to the sink is $\mathbf{C}_{f\mathbf{c}} \triangleq f^{\text{out}}(s_{\text{source}})$.

The value of flow exiting the summary node to the sink is $\mathbf{C}_{f\mathbf{y}} \triangleq f^{\text{out}}(s_{\text{summary}})$.

Score method definitions

Aggregate flow:

$$\mathbf{C}_f = 2\frac{\mathbf{C}_{f\mathbf{c}} \cdot \mathbf{C}_{f\mathbf{y}}}{\mathbf{C}_{f\mathbf{c}} + \mathbf{C}_{f\mathbf{y}}}$$

Aggregate alignment portion:

$$\widetilde{\mathbf{C}} = 2\frac{\widetilde{\mathbf{C}}_{\mathbf{c}} \cdot \widetilde{\mathbf{C}}_{\mathbf{y}}}{\widetilde{\mathbf{C}}_{\mathbf{c}} + \widetilde{\mathbf{C}}_{\mathbf{y}}}$$

Previous Work

Method

Scoring

Results

Document Summarization Scores

Scoring matched vs. mismatch corpus.

Corpus	$\widetilde{\mathbf{C}}_\mathbf{y}$	$\widetilde{\mathbf{C}}_\mathbf{c}$	$\mathbf{C}_{f\mathbf{y}}$	$\mathbf{C}_{f\mathbf{c}}$
Proxy report	**86.6%**	**43.2%**	7.21	0.67
Mismatch	35.1%	14.6%	2.61	0.2

AMR Sentence Pearson correlations (ρ) between aggregate alignment portion $(\widetilde{\mathbf{C}})$ALAMR and previous scoring methods (S)MATCH and (W)LK.

Corpus	**Parser**	$\rho\,\widetilde{\mathbf{C}},\mathbf{S}$	$\rho\,\widetilde{\mathbf{C}},\mathbf{W}$
Biomedical	Gsii	41.2	31.8
Biomedical	Jamr	66.2	65.2
Biomedical	Spring	50.1	41.3
Little prince	Gsii	38.8	35.7
Little prince	Jamr	**67.7**	**69.2**
Little prince	Spring	41.3	47.1
Proxy report	Gsii	22.9	30.8
Proxy report	Jamr	53.2	56.2
Proxy report	Spring	37.3	48.2

Aligned Node Text Comparison

Unigram (bag of words) aligned source to summary overlap of text-to-graph tokens.

Method	Precision	Recall	F1
Liu et al. [7]	51.9%	39.0%	44.3%
Dohare et al. [2]	52.4%	55.7%	51.3%
Fu et al. [4]	-	-	49.1%
CALAMR	69.0%	68.6%	**68.8%**

References

[1] Laura Banarescu et al. "Abstract Meaning Representation for Sembanking". In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 178–186 (cit. on p. 5).

[2] Shibhansh Dohare, Harish Karnick, and Vivek Gupta. *Text Summarization Using Abstract Meaning Representation*. 2017. URL: `http://arxiv.org/abs/1706.01678`. preprint (cit. on p. 24).

[3] Lester Randolph Ford and Delbert Ray Fulkerson. "Flows in Networks". In: *Flows in Networks*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1962, p. 212. ISBN: 978-0-691-65184-2 (cit. on p. 9).

[4] Qiankun Fu et al. "End-to-End AMR Coreference Resolution". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. ACL-IJCNLP 2021. Online: Association for Computational Linguistics, Aug. 2021, pp. 4204–4214. DOI: `10.18653/v1/2021.acl-long.324` (cit. on p. 24).

References (cont.)

[5] Paul Kingsbury and Martha Palmer. "From TreeBank to PropBank". In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. LREC 2002. Las Palmas, Canary Islands - Spain: European Language Resources Association (ELRA), May 2002 (cit. on p. 6).

[6] Kevin Knight et al. *Abstract Meaning Representation (AMR) Annotation Release 3.0*. Abacus Data Network, Sept. 3, 2021. DOI: 10.35111/44cy-bp51 (cit. on p. 12).

[7] Fei Liu et al. "Toward Abstractive Summarization Using Semantic Representations". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Denver, Colorado: Association for Computational Linguistics, 2015, pp. 1077–1086. DOI: 10.3115/v1/N15-1114 (cit. on pp. 7, 24).

[8] Alexander Schrijver. "On the History of the Transportation and Maximum Flow Problems". In: *Mathematical Programming* 91.3 (Feb. 1, 2002), pp. 437–445. ISSN: 1436-4646. DOI: 10.1007/s101070100259 (cit. on p. 9).