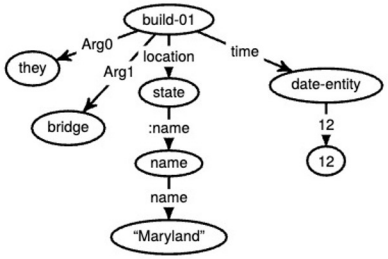
Anchor and Broadcast

An Efficient Concept Alignment Approach for Evaluation of Semantic Graphs

Abstract Meaning Representation



(1) (b / build-01
:ARG0 (t / they)
:ARG1 (b2 / bridge)
:location (s / state
:name (n / name :op1 "Maryland"))
:time (d / date-entity :month 12))

Figure 1: An example AMR representation for the sentence "They built a bridge in Maryland in December."

- Rooted directed acyclic graph
- Concrete concepts: word tokens
- Abstract concepts: inferred from contexts

Problem

(2) He likes apples. She hates oranges.

```
(l / like                    (h / hate
   :ARG0 (j / he)               :ARG0 (m / she)
   :ARG1 (a / apple))           :ARG1 (o / orange))
```

(3) He likes her.

```
(l / like                    (l / like
   :ARG0 (h / he)               :ARG0 (s / she)
   :ARG1 (s / she))             :ARG1 (h / he))
(h / he
   :ARG0 (l / likes)
   :ARG1 (s / she))
```

- Measure IAA
- Evaluate parser
- Previous metrics:
  - Intuitive?
  - Interpretable?
- Smatch:
  - 0.5; 0.67, 0.5, 0.5.
- Sembleu:
  - 0; 0.5, 0.5, 0.5.

Improvement

She is reading my book in the house.

```
(r1 / read-01                (r1 / read-03
 :location (h2 / house)       :location (h2 / home)
 :ARG0 (h / he)               :ARG0 (s / she)
 :ARG1 (b / book)             :ARG1 (p / paper)
         :poss (i / i))               :part (i / i))
```

- Efficient node alignment
  - Interpretability
- Fine-grained metric
  - Intuitiveness

Content

- How to align nodes?
  - Intrinsic node similarity
  - Anchor & Broadcast: integrating intrinsic node similarity with neighborhood information
- How to evaluate the similarity of two graphs?
  - Concept F-1
  - Labeled relation F-1
- Evaluation results compared to other metrics

Intrinsic Node Similarity

(4) (f / fry-03
:quant 5
:polarity -
:mode imperative)

(s / stir-fry-01
:quant 7
:polarity -)

$$S_{ij} = \frac{S_{ij}^{(l)}(1+\gamma(S_{ij}^{(s)}-1))+S_{ij}^{(a)}}{1+\mathbb{1}(|R^{(i,j)}|)}$$

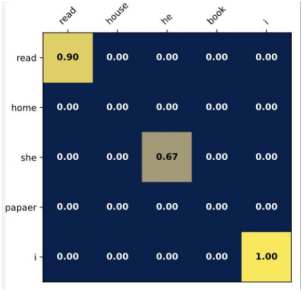
- String similarity
  - fry / stir-fry: 37.5%
- Attribute similarity
  - 2 shared properties, with 1 having the same value: 50%
- Sense-ID similarity
  - 03 ≠ 01, 0%
- Overall Similarity
  - (0.375*0.9+0.5)/2=0.41875 ≈ 0.42

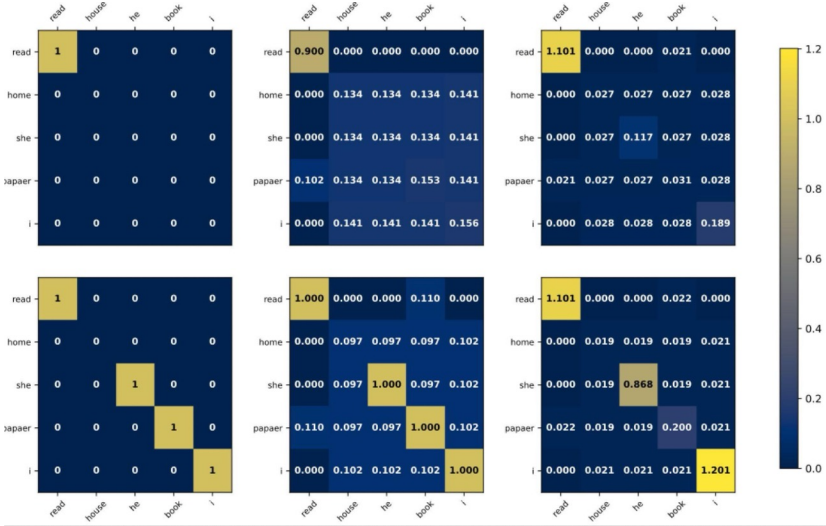
Anchors & Broadcast

She is reading my book in the house.

(r1 / read-01
 :location (h2 / house)
 :ARG0 (h / he)
 :ARG1 (b / book)
 :poss (i / i))

(r1 / read-03
 :location (h2 / home)
 :ARG0 (s / she)
 :ARG1 (p / paper)
 :part (i / i))





- Initial anchors:
  - unique
  - concrete concept
  - same word tokens
- Inferred anchors:
  - mutually best match
  - combination of neighborhood information and intrinsic similarity

Metrics: Concept F-1

She is reading my book in the house.

```
(r1 / read-01                (r1 / read-03
 :location (h2 / house)       :location (h2 / home)
 :ARG0 (h / he)               :ARG0 (s / she)
 :ARG1 (b / book)             :ARG1 (p / paper)
         :poss (i / i))               :part (i / i))
```

- F-1 of intrinsic node similarity of all aligned pairs.
- Whether two graphs are about the same content.

Metrics:
Labeled Relation F-1

She is reading my book in the house.

```
(r1 / read-01                (r1 / read-03
 :location (h2 / house)       :location (h2 / home)
 :ARG0 (h / he)               :ARG0 (s / she)
 :ARG1 (b / book)             :ARG1 (p / paper)
       :poss (i / i))               :part (i / i))
```

$$s_c = (S_{v_1 w_1} + S_{v_2 w_2})/2$$

$$s_p = s_c s_{ol}$$

- Triple similarities are calculated within pairs with the same head and tail
- Label similarity is binary.

<read-01, location, house>	<read-03, location, home>	0.45
<read-01, ARG0, he>	<read-03, ARG0, she>	0.785
<read-01, ARG1, book>	<read-03, ARG1, paper>	0.45
<book, poss, i>	<paper, part, i>	0

Experiment: Comparison to Other Metrics

(%)	Transition	BiBL	AMRBART	ATP-AMR	SPRING
Smatch (4 random starts)	81.15	83.83	86.17	85.80	84.55
Mock Smatch	79.59	82.63	85.01	84.57	83.23
Pearson Correlation	97.34	98.36	98.21	97.93	98.07
Sembleu	64.66	71.00	72.43	X	X
Labeled Relation Micro F1	70.14	76.61	76.16	75.99	74.76
Labeled Relation Macro F1	70.00	75.28	79.02	78.50	76.39
Concept F1	90.29	90.13	92.67	92.83	91.54
Unlabeled Relation F1	74.83	78.51	82.27	81.73	79.57
Weighted Relation F1	70.10	73.85	77.59	76.91	74.69
Runtime of Smatch	36.56s	33.97s	33.41s	37.32s	31.95s
Runtime of Sembleu	0.08s	0.08s	0.08s	X	X
Runtime of A & B	14.33s	15.79s	15.63s	15.68s	15.63s

Table 1: Comparison of different methods on 1898 AMRs in AMR 3.0 dataset. All non-micro F1's are macro F1, including Smatch, Mock Smatch and Sembleu. Sembleu is incapable of parsing graphs with early re-entrancies.

- Alignment algorithm is faster and retaining Smatch property;
- Provide more fine-grained metrics enabling analysis of more facets.

Conclusion

- New metrics that include concept F1, labeled relation F-1, etc. which are closer to human's intuition of semantic graph similarity.
- Efficient Anchor & Broadcast algorithm that has a polynomial runtime for node alignment.
- This represents a significant improvement over most widely used MR metric.