



Align-smatch: A Novel Evaluation Method for Chinese Abstract Meaning Representation Parsing based on Alignment of Concept and Relation

Liming Xiao¹, Bin Li¹, Zhixing Xu¹, Kairui Huo¹, Minxuan Feng¹, Junsheng Zhou², Weiguang Qu²

1. School of Chinese Language and Literature, Nanjing Normal University, Nanjing, Jiangsu, China

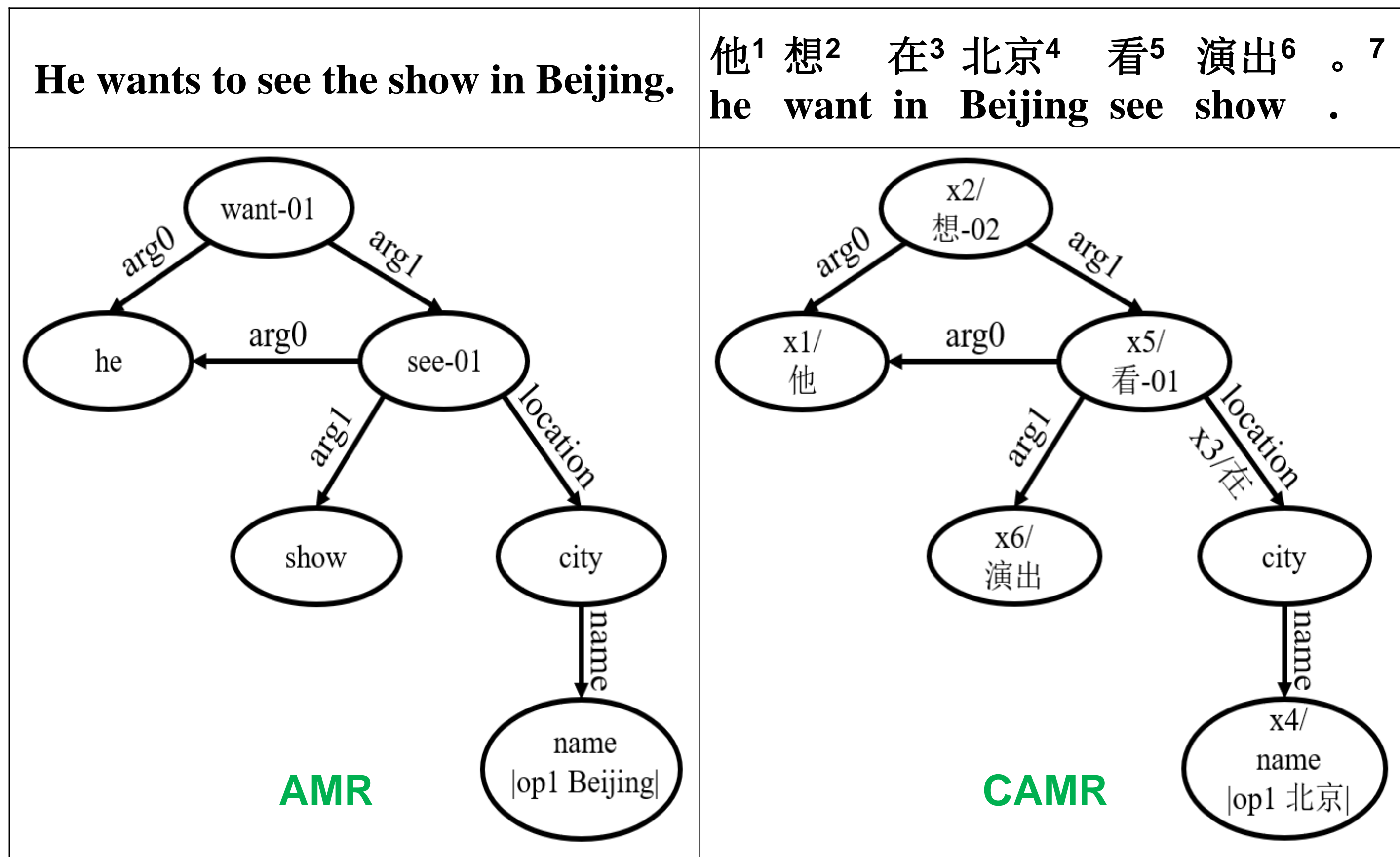
2. School of Computer, Electronics and Information, Nanjing Normal University, Nanjing, Jiangsu, China

✉: lmxiao1@foxmail.com, libin.njnu@gmail.com, xzx0828@live.com, kairui.huo.nj@gmail.com, fennel_2006@163.com, {zhoujs, wgqu}@njnu.edu.cn

1. From AMR to Chinese AMR

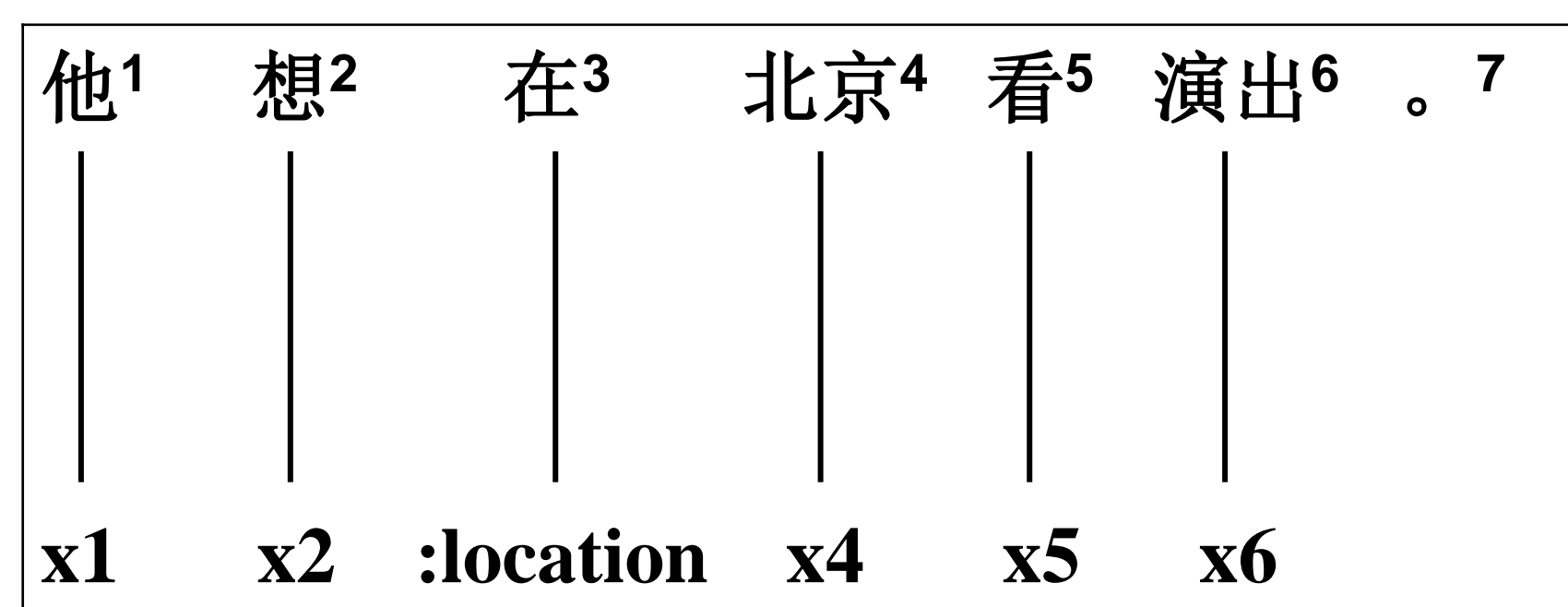
Abstract Meaning Representation (AMR) [1]:

- Sentence's semantic structure
- Rooted, directed, acyclic graph with nodes (concepts) and labels on arcs (relations)
- Lack of explicit alignments required for AMR parsing



Chinese AMR (CAMR) [2]:

- Concept Alignment: number after "x" in node's ID is the index of the aligned word in sentence
- Relation Alignment: the function word on arcs, whose label has the same function as the function word



2. Smatch

Smatch^[3] is a popular evaluation method for AMR parsing.

Process:

- Reset node's ID and transform two AMRs into two triple sets, each of which consists of node triple, arc triple and node property triple

$Num(Triples\ of\ Gold)=9$

$Num(Triples\ of\ Parsed)=4$

- Use hill-climbing method to find the optimal matching number of two triple sets

$Num(Match\ Triples)=2$

- Calculate Precision(P), Recall(R) and F1 score

Shortcoming:

- Concepts are not considered when comparing two arcs and two top nodes

- Alignments cannot be evaluated

	在 ¹ 北京 ² 看 ³ 演出 ⁴ in Beijing see show	
	Gold CAMR	Parsed CAMR
	Smatch Triples	
Node	'instance'(a0, 看-01) 'instance'(a1, 演出) 'instance'(a2, city) 'instance'(a3, name)	'instance'(b0, 看-04) 'instance'(b1, 北京)
Arc	arg1(a0, a1) location(a0, a2) name(a2, a3)	location(b0, b1)
Prop.	top(a0, 'TOP') op1(a3, 北京)	top(b0, 'TOP')
	Match Triples	
	location(a0, a2)-location(b0, b1) top(a0, 'TOP')-top(b0, 'TOP')	
	Smatch scores	
	P=2/9, R=2/4, F1=4/13≈0.31	

3. Align-smatch

Modify Smatch:

- Two arc triples can match only if the concepts at both ends are same

$location(看-01, city) \neq location(看-04, 北京)$

- Transform top node into arc triple instead of node property triple

$top(a0, 'TOP') \rightarrow top(a0, a0)$

$top(b0, 'TOP') \rightarrow top(b0, b0)$

Add tuples for alignments:

- Concept alignment:
 $anchor(Node, ID)$
- Relation alignment:
(Word, ID, Source Node, Target Node)

Process:

- Linearize CAMR
 $(x3/看-04:location(x1/在)(x2/北京))$

- Set state to symbols
 $\rightarrow 0, (\rightarrow 1, :\rightarrow 2, / \rightarrow 3)$

- Form tuples according to state

Align-smatch Tuples		
Node	'instance'(a0, 看-01) 'instance'(a1, 演出) 'instance'(a2, city) 'instance'(a3, name)	'instance'(b0, 看-04) 'instance'(b1, 北京)
Arc	arg1(a0, a1) location(a0, a2) (在, x1, a0, a2) name(a2, a3) top(a0, a0)	location(b0, b1) (在, x1, b0, b1) top(b0, b0)
Prop.	op1(a3, 北京) anchor(a0, x3) anchor(a1, x4) anchor(a3, x2)	anchor(b0, x3) anchor(b1, x2)
Match Tuples		
/		
Align-smatch scores		
P=0/13, R=0/7, F1=0		

4. Smatch VS Align-smatch

Data:

- 100 sentences/CAMR pairs from CAMR1.0 corpus (LDC2019T07) as data G
- Two annotators re-annotate 100 sentences as data A/B

Metric:

- Smatch, Align-smatch
- Concept-smatch: only put triple of concept alignment into Smatch
- MOD: the version that perfects the shortcoming of Smatch

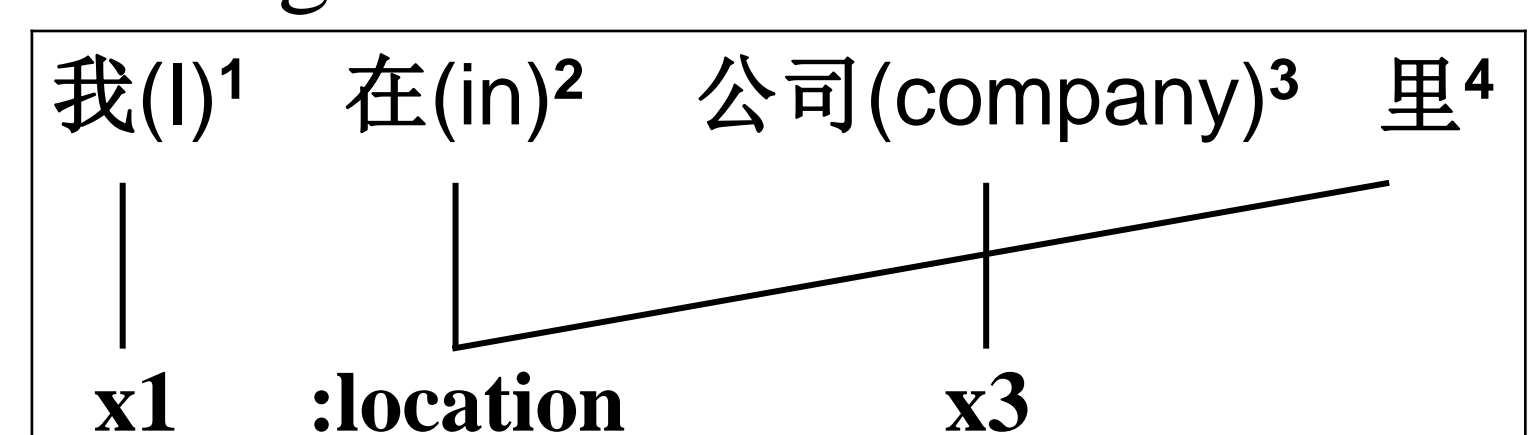
	Group	A-G	B-G
Smatch		0.78	0.82
Concept-smatch		0.81	0.86
Concept-smatch (MOD)		0.78	0.85
Align-smatch		0.79	0.81
Align-smatch (MOD)		0.77	0.81

Results:

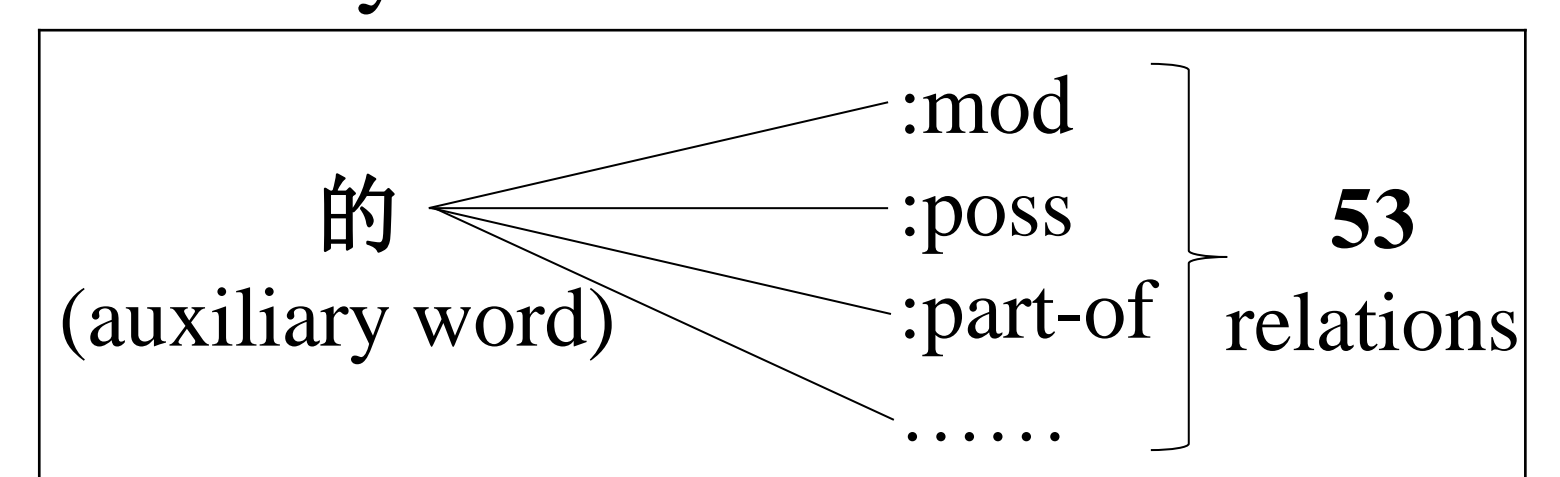
- Performance under MOD version is worse, but MOD version is more tough
- With the help of annotation platform, performance under Concept-smatch is better than Smatch
- Performance under Align-smatch is worse, which indicates the difficulty of annotating relation alignment

Conditions Inducing Errors:

- More than one function words align one relation



- One function word can align many kinds of relations.



References

- [1] Banarescu L, Bonial C, Cai S, et al. Abstract Meaning Representation for Sembanking[C]//Linguistic Annotation Workshop and Interoperability with Discourse. 2013:178-186.
- [2] Li B, Wen Y, Song L, et al. Building a Chinese AMR bank with concept and relation alignments[J]. Linguistic Issues in Language Technology, 2019, 18:1-41.
- [3] Cai S, Knight K. Smatch: an evaluation metric for semantic feature structures[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2013:748-752.