




Zlepšení značení čínské sémantické role

Veronika Kalivodová

- 
- Předchozí práce byla zaměřena především na to, jak implementovat metody, které jsou úspěšné v angličtině.
 - Informace syntaktických stromů není plně využita, proto se vyžaduje větší úsilí v oblasti výzkumu.
 - Nové funkce, z nichž některé jsou určeny pro lepší zachycení strukturální informace.
 - S pomocí nových vlastností systém dosahuje úspěšnosti 93,49%.
 - Dvojitý efekt v syntaktickém analyzování v čínské SRL.

Chinese Semantic Role Labeling (SRL)

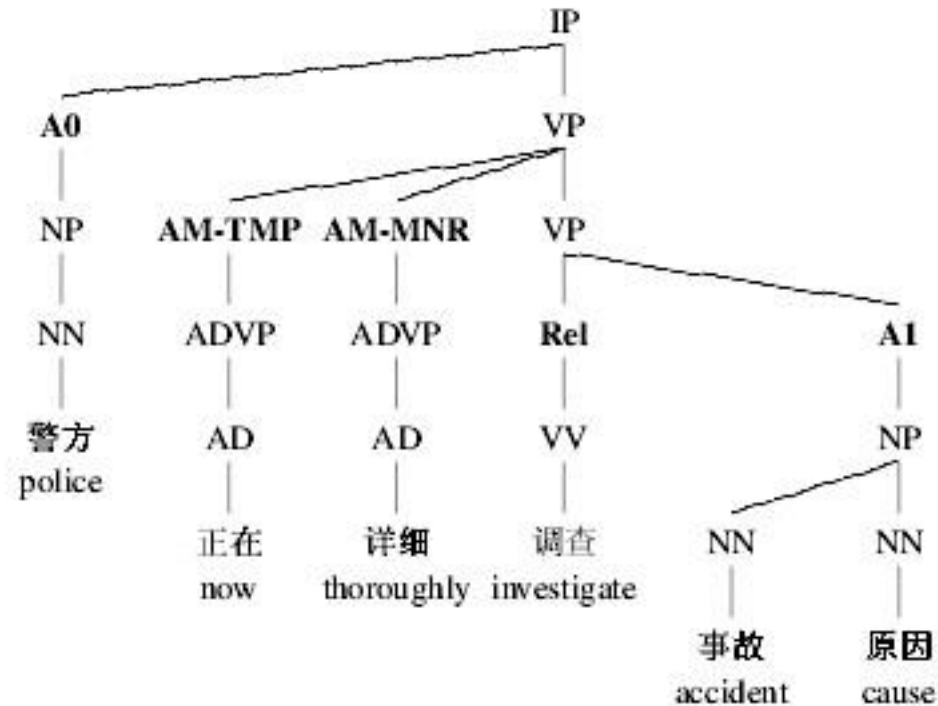


Figure 1: An example sentence: *The police are thoroughly investigating the cause of the accident.*

Chinese Semantic Role Labeling (SRL)

- Argumenty výpovědi jsou označeny ve formě AN (N je přirozené číslo).
- Doplnky jsou označovány jako AM, který reprezentuje sémantickou klasifikaci doplňku.

- Popis obrázku:

Výrok je sloveso vyšetřovat. Příčina nehody je označena jako AI, což znamená, že se jedná o patient.

- Metody SRL jsou rozděleny do tří pod-úloh:
 - 1) odvětvení s heuristickým pravidlem
 - 2) identifikace důvodu (AI) k rozpoznání argumentu
 - 3) sémantická klasifikace role (SRC)
- AI a SRC jsou formulovány jako klasifikační problémy.
- SRC – u každého argumentu by se mělo nejprve určit, zda je to základní argument nebo přívlastek, a poté roztrždit do jednoznačné kategorie.

Celkový výkon

Test	P(%)	R(%)	F/A
AI	98.56	97.91	98.24
SRC	--	--	95.04
(Ding and Chang, 2008)	--	--	94.68
AI + SRC	93.80	93.18	93.49
(Xue, 2008)	93.0	91.0	92.0

Table 1: SRL performance on the test data with gold standard parses.

Tabulka ukazuje výkon nově navržených systémů. Výsledky ukazují významné zlepšení oproti předchozímu systému.

Dvojitý efekt analýzy v SRL

1. Systémy by měly uskupit slova jako kandidáty argumentu.
 2. Systémy mohou určit, zda se jedná o argument a dále předpovědět detailní sémantické typy.
- Pro dva klasifikační problémy analýza může poskytnout komplexní syntaktické informace, jako třeba trajektorii vlastností.

Dvojitý efekt analýzy v SRL

Task	Parser	Bracket	Feat	P(%)	R(%)	F/A
AI	--	Gold	W	82.44	86.78	84.55
	CTB	Gold	W+S	98.69	98.11	98.40
	Bikel	Bikel	W+S	77.54	71.62	74.46
SRC	--	Gold	W	--	--	93.93
	CTB	Gold	W+S	--	--	95.80
	Bikel	Gold	W+S	--	--	92.62

Table 2: Classification performance on development data. In the *Feat* column, *W* means word features; *W+S* means word and syntactic features.

Zdroj

- <http://aclweb.org/anthology//P/P10/P10-2031.pdf>

Děkuji za pozornost 😊