Representing Text Chunks

Erik F. Tjong Kim Sang
Center for Dutch Language and Speech
Universitaire Instelling Antwerpen
Belgium
erikt@uia.ac.be

overhead sheets: http://pcger34.uia.ac.be/~erikt/talks/

TMR Network

Learning Computational Grammars

Goal

Apply machine learning techniques for recognizing the structure of noun phrases (NPs).

Steps

1. Recognize NP boundaries.
2. Discover syntactic structure within NPs.
3. Find out semantic roles of NP constituents.

NP Chunking

Recognize non-recursive noun phrases (baseNPs): NPs that do not contain another NP.

Ramshaw & Marcus (WVLC95) have used Transformation-Based Learning for training an NP chunker on a subset of the Wall Street Journal corpus.

Their NP chunker achieved a precision and recall rates of approximately 92% while recognizing baseNPs from section 20 of this corpus.

Data representation

RM95: Text chunking can be represented as a tagging task by using three tags: I, O and B

- In $[_N$ early trading $_N]$ in $[_N$ Hong Kong $_N]$ $[_N$ Monday $_N]$, $[_N$ gold $_N]$ was quoted at $[_N$ $ 366.50 $_N]$ $[_N$ an ounce $_N]$.
- In/O early/I trading/I in/O Hong/I Kong/I Monday/B ,/O gold/I was/O quoted/O at/O $/I 366.50/I an/B ounce/I ./O

Advantage: tags are less dependent on each other than brackets. Problems can be solved locally.

Alternative data representation formats

IOB1	O	I	I	O	I	I	B	O	I	O	O	O	I	I	B	I	O
IOB2	O	B	I	O	B	I	B	O	B	O	O	O	B	I	B	I	O
IOE1	O	I	I	O	I	E	I	O	I	O	O	O	I	E	I	I	O
IOE2	O	I	E	O	I	E	E	O	E	O	O	O	I	E	I	E	O
[	.	[	.	.	[	.	[	.	[	.	.	.	[	.	[	.	.
]	.	.	]	.	.	]	]	.	]	.	.	.	.	]	.	]	.

Experiment setup

Memory-based learning (IB1-IG) has been used for performing five-fold cross-validation experiments on section 15 of the Wall Street Journal corpus as prepared by RM95.

Each experiment included four parts. Each part examined different parameter settings.

Results have been measured by examining an average of recall and precision rates:
F = (2×precision×recall)/(precision+recall)

The best parameter settings have been used for processing the complete RM95 data set (sections 15-18 from WSJ as training material and section 20 for testing).

Part 1: determining context size

Example: In/IN early/JJ trading/NN → I

IOB1	R=0	R=1	R=2	R=3	R=4
L=0	81.31	85.15	84.81	84.19	83.87
L=1	87.09	89.12	88.74	88.42	88.22
L=2	86.33	89.17	88.87	88.76	88.59
L=3	85.83	88.96	88.64	88.60	88.43
L=4	85.78	88.65	88.43	88.41	88.34

	Context	F
IOB1	L=2 R=1	89.17
IOB2	L=2 R=1	88.76
IOE1	L=1 R=2	88.67
IOE2	L=2 R=2	89.01
[	L=2 R=1	87.34
]	L=0 R=2	↑

Part 2: determining IOB context size

In/IN/O early/JJ trading/NN/I → I

IOB1	R=0	R=1	R=2	R=3
L=0	89.17	89.77	89.65	89.57
L=1	89.68	90.11	90.12	90.02
L=2	89.64	90.11	89.99	89.98
L=3	89.51	89.96	89.95	89.92

	Context	IOB Context	F
IOB1	L=2 R=1	L=1 R=2	90.12
IOB2	L=2 R=1	L=1 R=0	89.30
IOE1	L=1 R=2	L=1 R=2	89.55
IOE2	L=1 R=2	L=0 R=1	89.73
[	L=2 R=1	L=2 R=0	87.72
]	L=0 R=2	L=0 R=0	↑

Part 3: combine some part 1 results

RM95 use transformation rules with different context sizes. It might help if we combine different results from part 1 in the second processing step.

IOB1	R=0	R=1	R=2	R=3	R=4
L=0	81.31	85.15	84.81	84.19	83.87
L=1	87.09	<u>89.12</u>	88.74	88.42	88.22
L=2	86.33	<u>89.17</u>	88.87	88.76	88.59
L=3	85.83	<u>88.96</u>	88.64	88.60	88.43
L=4	85.78	88.65	88.43	88.41	88.34

	Context	part 1	F
IOB1	L=2 R=1	00 11 22 33	90.53
IOB2	L=2 R=1	21	89.30
IOE1	L=1 R=2	00 11 22 33	90.03
IOE2	L=1 R=2	12	89.73
[	L=2 R=1	21	87.72
]	L=0 R=2	-	↑

Part 4: change k

In IB1-IG, k is the number of nearest neighbors that are considered when determining the most probable output.

Daelemans, Van den Bosch and Zavrel (ML99): abstraction (k>1) is harmful in language learning.

But not for NP chunking?

	k=1	k=2	k=3	k=5
IOB1	90.53	90.10	90.21	89.17
IOE1	90.03	89.35	89.85	88.88

An explanation for the results

By knowing the POS tag we can predict the IOB tags for IOB1 with 94% accuracy. We can use a conditional entropy measure for estimating the difficulty of the representation formats:

$$H = - \Sigma_S P(S) \Sigma_I (P(I|S) * log_2(P(I|S))$$

S = POS tag ; I = IOB tag

WSJ15	H	F
IOB1	0.283	90.53
IOB2	0.561	89.30
IOE1	0.327	90.03
IOE2	0.492	89.73
[	0.401	87.72
]	0.311	↑

Processing the complete RM95 data set

	Context t/w	Context IOB	part 1	k	F	
IOB1	21	11	00 11 22 33	1	91.86	
IOB2	21	10	21	1	91.22	
IOE1	12	12	00 11 22 33	1	92.17	WR
IOE2	12	01	12	1	91.67	
[	21	20	21	1	89.47	
]	02	-	-	1	↑	

Comparison with other work

IOB1	accuracy	precision	recall	F
RM95	97.37	91.80	92.27	92.03
TKS98	97.43	91.70	92.03	91.86
Vee98	97.2	89.0	94.3	91.6
ADK98	-	91.6	91.6	91.6
CP98	-	90.7	91.1	90.9

IOE1	accuracy	precision	recall	F
TKS98	97.55	92.13	92.22	92.17

Concluding remarks

- For a baseNP recognition task it is better to represent data as tags than work with bracket structures.
- It is unclear whether IOB1 or IOE1 is the best tag structure for this task.
- For other related tasks the usability of representation formats can be estimated by computing the conditional entropy of the output given the most important input feature(s).

Future work

- Performing n-fold cross-validation experiments with a larger data set.
- Combining results of different representations in the second processing part.

Address

address:	Center for Dutch Language and Speech Universitaire Instelling Antwerpen Universiteitsplein 1 B-2610 Wilrijk Belgium
phone:	+31 3 8202765
e-mail:	erikt@uia.ac.be
www:	http://ger-www.uia.ac.be/u/erikt/

Overhead sheets

http://pcger34.uia.ac.be/˜erikt/talks/

Software (TiMBL)

http://ilk.kub.nl/

Data

ftp://ftp.cis.upenn.edu/pub/chunker/