

k-NN

o

sbpark@nova.snu.ac.kr {btzhang,ytkim}@cse.snu.ac.kr

Expanded Korean Chunking by *k*-NN

Seong-Bae Park^o

Byoung-Tak Zhang

Yung Taek Kim

School of Computer Science and Engineering, Seoul National University

1.

(chunking)

- 가
- [NP 가] [VP] [NP] [VP]
“ ”, “ ”, “ 가”,
“ ”가 NN (instance-based method)

Abney 가
가 [1], Ramshaw

Marcus [2].

[7].
가 Abney

(chunk)
가

[4].
93%

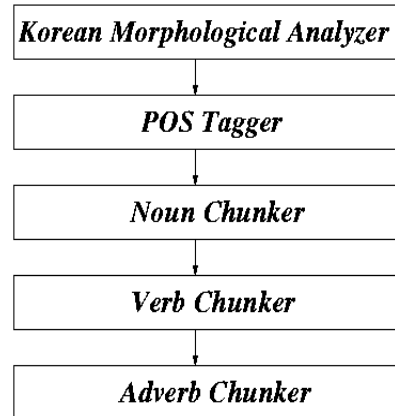
[5].

30

2

2.

1



1.

가

가

2

3. 3.1

[5],

- 1: ()⁵ * ()
- 2: () * ()
- 3: () * ()
- 4: () * ()
- 5: () *

HAM

5

- 5: () * ()
- [6] 가 “[]” 가

3

가

3.2

[3]. [5] [6]

[3]

3.3

가 가

가

가

4. 4.1

가

<p>1. B-NP & I-NP -[] [] [] [] [] ⇒ [] [] [] [] []</p> <p>2. I-NP & I-NP -[] [] 가 [] [] ⇒ [] [] [] 가 [] []</p> <p>3. B-NP & B-NP -[] [] 가 [] [] [] ⇒ [] [] 가 [] [] []</p> <p>4. B-NP, I-NP & B-NP -[] [] [] [] [] ⇒ [] [] [] [] []</p> <p>5. B-ADVP & I-ADVP -[] [] [] [] [] ⇒ [] [] [] [] []</p> <p>6. I-NP & B-NP -[] [] [] [] [] ⇒ [] [] [] [] []</p>

2.

가
 B-NP² I-NP
 B-NP B-NP

4.2 k-NN

w_i

1

w _i	w _i
w _{i-1}	w _{i-1}
w _{i-2}	w _{i-2}
POS _i	w _i
POS _{i-1}	w _{i-1}
POS _{i-2}	w _{i-2}
E _{i-1}	w _{i-1}
R _{i-1}	w _{i-1}
R _{i-2}	w _{i-2}

1.

k-NN

n-

Rⁿ

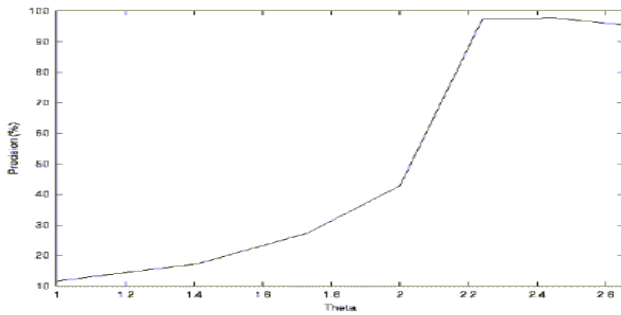
1

2

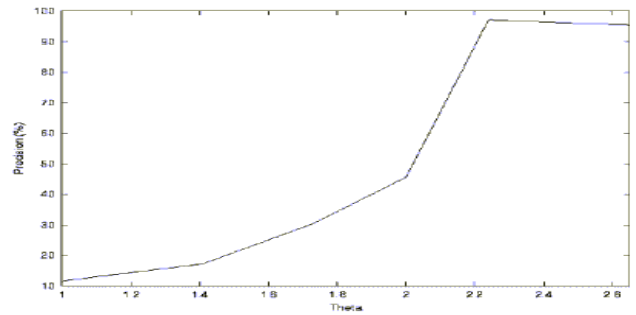
(chunk)

I-VP, B-ADVP, I-ADVP

B-NP, I-NP, B-VP,



(a) $k = 1$



(b) $k = 3$

3. θ

(a)

$k = 1$

(b) $k = 3$

가 가

가

k -NN

,

가

k

θ

k -NN

$$\frac{\sum_{i=1}^k d(x_i, x_q)}{k} \geq \theta$$

x_q

k -NN

가 가

x_1, \dots, x_k

k -NN

5.

1,911

8.9

, 1,273

638

2

. 3

95.5%

, k -NN

2.3%

97.8%

$k = 1$

k -NN	95.3 % N/A	95.5 % 97.8 %

2.

3 k

θ 가 k -NN

(a) $k = 1$

k 가 1

, (b) $k = 3$

3

가

가

가

, θ

가

θ

k -NN

가

6.

k -NN

, k -NN

가

, k -NN

가

k -NN

2.3%

, 2.3%

”(00-102)

[1] S. Abney, "Parsing by Chunks," In *Principle-Based Parsing*, Kluwer Academic Publishers, 1991.

[2] L. Ramshaw and M. Marcus, "Text Chunking Using Transformation-Based Learning," In *Proceedings of the Third ACL Workshop on Very Large Corpora*, pp. 82-94, 1995.

[3] , , , " , 22 , 9 , pp. 1384-1393, 1995.

[4] H.-P. Shin, "The VP-Barrier Algorithm for a Robust Syntactic Parsing in Head-Final Languages," In *Proceedings of Natural Language Processing Pacific Rim Symposium*, pp. 475-478, 1999.

[5] , " , " 11 , pp. 242-247, 1999.

[6] , , , " , 27 , pp. 327-329, 2000.

[7] Erik T.-K. Sang, "Noun Phrase Representation by System Combination," In *Proceedings of ANLP-NAACL 2000*, pp. 50-55, 2000.