

# 독립성분분석과 Box-Cox 변환을 이용한 동사 개념 클러스터링

Chagnaa Altangerel, 이 창 범, 옥 철 영  
울산대학교 컴퓨터정보통신공학부  
{goldenl, chblee1225, okcy}@mail.ulsan.ac.kr

## Verb concept clustering using Independent Component Analysis and Box-Cox transformation

Altangerel Chagnaa, Changbeom Lee, Cheolyoung Ock  
School of computer engineering and information technology, University of Ulsan

### 요 약

본 논문에서는 한국어 동사의 개념적 클러스터링 방법을 제안한다. 사용되는 기법은 독립 성분분석, Box-Cox 변환, 상관분석 등이다. 독립성분분석은 잠재적인 성분을 통계적 독립 (statistical independence)에 기반하여 추출하는 분석 방법이다. 그런데, 독립성분분석에서는 mixture(동사)의 분포는 정규 분포(가우시안 분포)에 따른다고 가정한다. 따라서, 동사의 분포를 보다 정규 분포화 할 필요가 있다. 이에 본 논문에서는 Box-Cox 변환을 이용하여 동사의 분포를 정규 분포에 근사한다. 또한, 독립성분분석에서는 추출할 적당한 성분의 개수를 결정할 수가 없다. 이에 본 논문에서는 주성분분석의 결과로 획득되는 고유치의 누적 기여율을 이용하여 독립성분의 수를 결정한다. 그리고, 추출된 독립성분 벡터와 동사 벡터간의 상관관계수에 이용하여 독립성분(개념)에 밀접하게 관련있는 동사들을 하나의 클러스터로 구성한다. 한국어 동사를 대상으로 클러스터링한 결과, Box-Cox 변환을 적용한 경우가 더 좋은 성능을 보였다.

### 1. 서론

“국어를 가르치다”와 “영어를 배우다”라는 두 개의 문장에서, “가르치다”와 “배우다”는 어떤 행위를 나타내는 동사이다. 비록 서로 상반되는 개념이지만 이들이 같은 클러스터에 속할 가능성은 크다. 이러한 클러스터링 작업을 자동으로 수행할 수 있다면, WSD(Word Sense Disambiguation)와 시소러스(thesaurus) 자동 구축 등에 도움이 될 수 있다.

본 논문에서는 독립성분분석(ICA: Independent Component Analysis)을 기반으로 동사를 클러스터링하는 모델을 제안한다. 독립성분분석[4,5]은 신호처리 특히 BSS(Blind Source Separation) 분야에서 광범위하게 응용되어 왔다. 하지만, 텍스트 처리 분야에서는 독립성분분석의 응용이 상대적으로 미비하다. 다만, [1, 7]에서는 용어-문서 행렬을 독립성분분석의 mixture로 고려하면서 문서 분석을 시도하였다. 이때 구성된 의미 공간은 직교하지 않는 용어 빈도 히스토그램의 형태를 취하고 있다. 그리고, [3]에서는 언어적 특징 추출 (linguistic feature extraction)에 독립성분분석을 응용하였는데, 말뭉치로부터 구해진 문맥 정보를 이용하고 있다.

독립성분분석은 관측된 분포에서 알려지지 않은 또는 내재되어 있는 새로운 성분을 추출할 수 있다. 본 논문에서는 동사 분포에 내재되어 있는 새로운 성분(개념)을 추출하는데 독립성분분석을 이용한다. 그리고, 추출된 독립성분과 통계적으로 밀접한 관련이 있는 동사들은 같은 개념을 표현할 수 있다고 가정한다. 이러한 가정 하에, 동사 벡터를 사전 정의문에서 직접목적어 관계에 있는 명사와의 상호정보(Mutual Information) 값으로 표현하여, 동사 클러스터링을 시행함으로써 독립성분분석의 그 응용 가능성을 확인하고자 한다.

독립성분분석에서는 mixture 분포(동사 분포)는 가우시안분포 또는 정규분포를 따른다고 가정한다. 따라서, 본 논문에서는 mixture 분포를 역 변환(power transformation) 방법 중의 하나인 Box-Cox 변환 방법 [6]에 의하여 보다 가우시안 분포를 따르도록 변환한 후 독립성분분석을 시행한다. 변환 전후의 성능을 비교함으로써, 변환의 필요성 유무를 확인하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 독립성분분석에 대해서 간략하게 설명하고, 3장에서는 Box-Cox 변환 방법에 대해서 언급한다. 그리고, 4장에서는 제안한 모델을, 5장에서는 실험 결과에 대해 기술한다. 마지막으로, 제 6장에서는 결론 및 향후 연구에 언급한다.

## 2. 독립성분분석

독립성분분석[4]의 모델은 전통적으로 다음 식(1)과 같이 표현된다.

$$x = As \quad (1)$$

여기서  $x = (x_1, x_2, \dots, x_n)^T$  는 관측된 신호 즉, mixture를 나타내며,  $s = (s_1, s_2, \dots, s_n)^T$  은 독립성분 벡터로 가정한다. 그리고,  $A$  는 알려지지 않은 상수 행렬 즉,  $n \times n$  mixing matrix이다. 식 (1) 에서 단지  $x$  만이 관측할 수 있으며,  $A$  와  $s$  는 추정을 해야 한다.

독립성분분석 결과에 대해 다음의 두 가지를 고려해야 한다. 첫째, 독립성분의 분산(설명력)을 결정할 수 없다. 그렇기 때문에, 추출할 독립성분의 개수를 결정할 수 없다. 본 논문에서는 주성분분석의 결과인 고유치의 누적비율을 이용하여 추출할 독립성분의 개수를 결정한다. 둘째, 독립성분  $s_i$  는 서로 간에 독립이며, 가우시안 분포를 따르지 않아야만 한다.

본 논문에서는 독립성분분석을 수행하기 위해 FastICA[1] 패키지를 사용하였다.

## 3. Box-Cox 변환

Box-Cox변환[6]은 멱변환(power transformation) 방법의 일종이며, 원자료를 정규 분포에 근사하도록 변환하는 방법이다.

하나의 확률 변수  $x$  의 관측치가  $x_1, x_2, \dots, x_n$  라고 주어진다면, 식 (2)에 의해 원자료  $x_j$  를 구해진  $\lambda$  에 의해 변환한다.

$$x_j^{(\lambda)} = \begin{cases} \frac{x_j^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln x_j & \lambda = 0 \end{cases} \quad (2)$$

이때,  $\lambda$  는 아래의 식 (3)을 최대로 하는 값으로 구해진다. 그리고, 식 (4)는 변환된 자료의 산술 평균이다.

$$\ell(\lambda) = -\frac{n}{2} \ln \left[ \frac{1}{n} \sum_{j=1}^n (x_j^{(\lambda)} - \overline{x^{(\lambda)}})^2 \right] + (\lambda - 1) \sum_{j=1}^n \ln x_j \quad (3)$$

$$\overline{x^{(\lambda)}} = \frac{1}{n} \sum_{j=1}^n x_j^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n \left( \frac{x_j^\lambda - 1}{\lambda} \right) \quad (4)$$

Box-Cox 변환 시 상기할 사항은 멱변환이 그러하듯이 Box-Cox 변환은 양수에 대해서만 정의된다는 점이다. 하

지만, 이것은 변환 시 큰 문제가 되지는 않는다. 예를 들어, 관측치 중 양수가 아닌, 0 또는 음수가 있을 경우 단순히 하나의 상수를 모든 관측치에 더함으로써 이러한 제약 사항을 해결할 수 있다.

본 논문에서는 통계패키지 중의 하나인 미니탭(MINITAB)을 이용하여 적당한  $\lambda$  를 구하였다.

## 4. 동사의 개념 클러스터링

동사 클러스터링 위해 크게 자료 변환, 독립성분 추출 그리고 클러스터 형성 과정 등 세 단계로 나눌 수 있다.

### 4.1 원자료의 Box-Cox 변환

전술한 바와 같이 독립성분분석에서는 그 분석 대상(mixture)의 분포를 가우시안 또는 정규 분포로 가정하고 있다. 따라서, 본 논문에서는 그 가정에 보다 충실하기 위해 mixture에 Box-Cox 변환을 적용한 후, 정규성 검정(normality test)를 시행한다.

본 논문에서는 mixture를 동사로 간주하며, 각 동사를 표현하기 위해 사전 뜻풀이 상에서 직접목적어 관계에 있는 명사를 이용한다. 그리고, 동사와 명사와의 연관 정도는 상호 정보(mutual information)로 표현하였다 [2,8]. 다시 말하면, 각 동사  $x_i$  는 벡터

$(x_{i1}, x_{i2}, \dots, x_{im})$  로 표현되며,  $x_{ij}$  는 명사와의 상호 정보 값이다.

각각의 확률 변수(동사)에 대해서 Box-Cox 변환을 시행하여 각각의  $\lambda$  를 구한 다음, 식 (2)에 따라 원자료를 변환하여야 한다[6]. 그런데, 동사 벡터의 대부분의 값이 0이기 때문에, Box-Cox 변환을 시행할 수 없다. 따라서, 모든 값에 1을 더한 후에 Box-Cox 변환하였다.

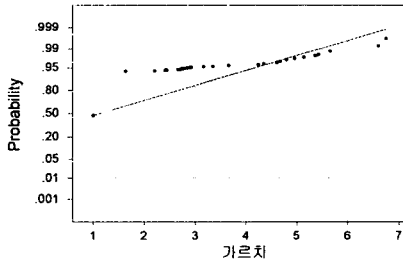
(그림 1)은 미니탭에서 제공되는 Ryan-Joiner 방법에 의해 Box-Cox 변환 전의 정규성 검정을 수행한 정규 확률 플롯을 나타낸다. 그래프에서 점들이 직선 상에 가까이 배치될수록 정규 분포에 가깝다고 판단할 수 있다. 하지만, (그림 1)에서는 대부분의 점들이 직선에 근접하지 않고 있다. 따라서, 동사 '가르치' 는 정규 분포를 따르지 않고 있다고 말할 수 있다. (그림 2)는 Box-Cox 변환 결과를 보이고 있는데,  $\lambda = -0.5$  와 같이 추정되었다. (그림 3)은 추정된  $\lambda$  를 이용하여 변환된 자료에 대한 정규성 검정 결과인데, 변환 전과 비교하여 크게 개선된 점이 없다. 오히려 정규 분포에서 더 멀어졌다고 볼 수 있는데, 이는 원자료의 대부분의 값이 0

1) <http://www.cis.hut.fi/projects/ica/fastica/>

으로 동일하기 때문이라고 생각된다.

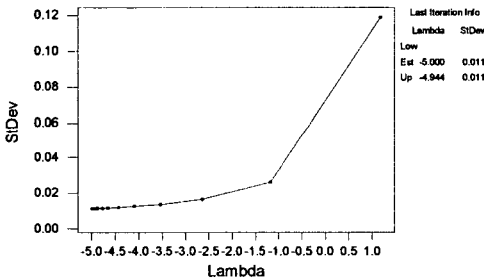
따라서, 본 논문에서는 동사 각각에 대해  $\lambda$  를 구하는 대신에, 원자료의 평균을 Box-Cox 변환하여 구한  $\lambda$  를 구한 다음, 식 (2)에 의거하여 원자료를 변환한다. 여기서, 원자료의 평균은 행이 명사이고, 열이 동사인 행렬에서 행의 평균을 의미한다.

Normal Probability Plot



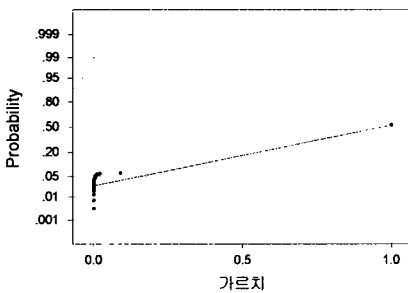
(그림 1) 변환 전의 정규 확률 플롯

Box-Cox Plot for 가르치



(그림 2) '가르치'의 Box-Cox 변환

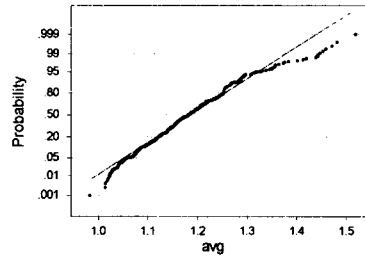
Normal Probability Plot



(그림 3) 변환 후의 정규 확률 플롯

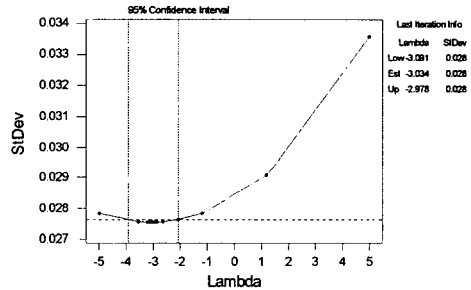
(그림 4)는 평균에 대한 정규 확률 플롯이며, (그림 5)는 평균에 대해 Box-Cox 변환 결과를 보이고 있는데, 추정된  $\lambda$  는  $-3.034$  임을 알 수 있다. (그림 6)은 평균에 대해  $\lambda = -3.034$  을 이용하여 변환한 결과이다. (그림 4)에 비해 (그림 6)의 점들이 보다 직선 상에 근접해 있음을 볼 수 있다. 이는 보다 정규 분포에 근사되었다고 판단할 수 있다.

Normal Probability Plot



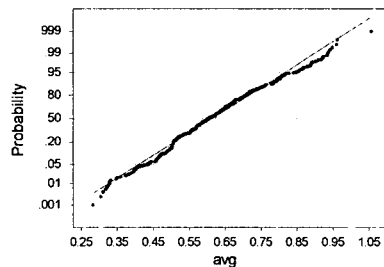
(그림 4) 변환 전의 정규 확률 플롯 (평균)

Box-Cox Plot for avg



(그림 5) 평균의 Box-Cox 변환

Normal Probability Plot



(그림 6) 변환 후의 정규 확률 플롯 (평균)

결국, 각 동사에 대해 Box-Cox 변환을 수행하여 각각

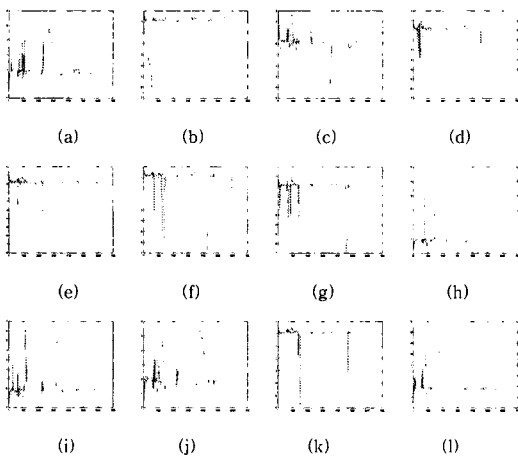
의  $\lambda$  을 구하는 대신에, 평균에 대해서만  $\lambda$  을 구하게 된다. 그리고, 이 평균에 대한  $\lambda$  를 이용하여 원자료의 모든 값은 식 (2)을 사용하여 변환하게 된다. 본 논문에서는  $\lambda = -3.034$  을 사용한다.

#### 4.2 독립성분 추출

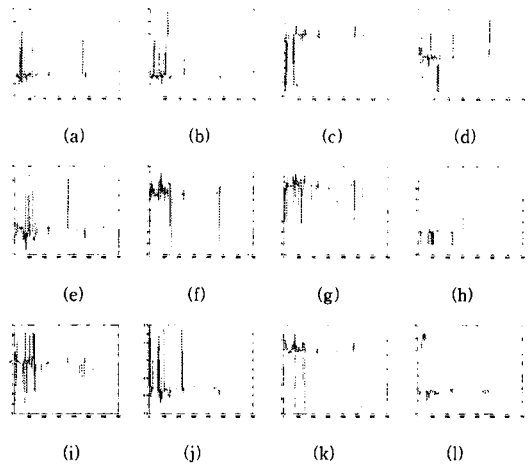
앞서 언급한 바와 같이 독립성분분석에서는 몇 개의 독립성분을 추출할 것인지 결정할 수 없다. 따라서, 본 논문에서는 추출할 독립성분의 수를 결정하기 위해 먼저 주성분분석을 수행하였다. 주성분분석은 변량간의 분산-공분산 정보를 기반으로 분석하는 다변량 분석 방법이다 [9, 10]. 주성분분석의 결과로 고유치들이 주어지는데 고유치의 누적 기여율에 따라 주성분의 개수를 선택할 수 있다. 본 논문에서는 0.7 과 0.8 을 기준값으로 선택하였다.

결정된 주성분수와 동일하게 독립성분분석에서 독립성분을 추출한다. 12개까지의 독립성분을 추출할 때의 누적기여율은 약 72.3% 이고, 15개까지의 경우에는 약 82.7% 였다.

(그림 7)은 Box-Cox 변환 전의 12개의 독립성분 벡터를, (그림 8)은 평균  $\lambda$  를 이용하여 Box-Cox 변환 후의 독립성분 벡터를 나타내고 있다. 여기서, 각 벡터의 차원은 실험에 사용된 명사의 수와 동일하다.



(그림 7) 12개의 독립성분과 명사와의 관계 (변환 전)



(그림 8) 12개의 독립성분과 명사와의 관계 (변환 후)

(그림 7)(b)와 (그림 8)(l)은 그 부호가 반대지만 비슷한 모습을 보이고 있다. 하지만, 그 외 대부분은 변환 전과 변환 후의 모습이 다름을 나타내고 있다.

#### 4.3 클러스터 형성

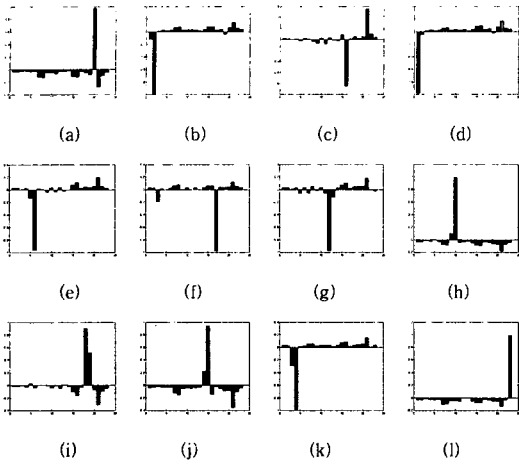
독립성분분석에 의해 추출된 성분을 새로운 개념으로 볼 때, 각 개념과 밀접한 상관성이 있는 동사들은 같은 클러스터에 포함될 수 있다. 독립성분 벡터  $IC_i$  와 동사 벡터  $x_k$  사이의 상관계수는 두 변수 사이의 공분산과 각 변수들의 표준편차에 대한 곱의 비율

$$\rho(IC_i, x_k) = \frac{Cov(IC_i, x_k)}{\sqrt{Var(IC_i)}\sqrt{Var(x_k)}} \quad (5)$$

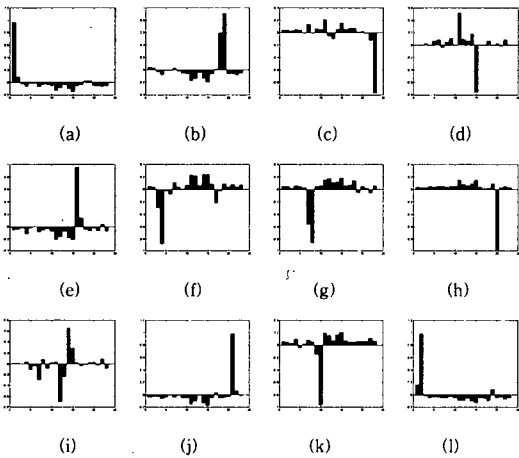
이다[10].

본 논문에서는 독립성분과 동사간의 상관계수를 식 (5)과 같이 계산을 하였으며, 그 값이 유의수준 0.05 하에서 통계적으로 유의한 동사들을 하나의 클러스터로 형성하였다. 같은 클러스터 있는 동사들은 같은 개념(독립성분)을 표현한다고 볼 수 있다.

(그림 9)와 (그림 10)은 12개의 독립성분과 동사와의 상관계수를 나타낸 그래프이다. (그림 9)는 Box-Cox 변환 전의 상황이며, (그림 10)은 변환 후의 모습이다. (그림 9)(d)와 (그림 10)(a)는 부호는 반대이지만 비슷한 양상을 보이고 있으나, 그 외의 독립성분에서는 다소 차이가 있다.



(그림 9) 독립성분과 동사와의 상관계수 (변환 전)



(그림 10) 독립성분과 동사와의 상관계수 (변환 후)

## 5. 실험 및 결과

동사-명사 패턴 정보를 추출하기 위해 품사와 의미 태깅된 금성 사전을 이용하였다. 이용한 사전에는 총 24,932개의 유일한 명사와 8,172개의 중복되지 않은 동사가 표제어의 뜻풀이 상에 출현하였다(총 296,442 동사-명사 쌍). 이러한 동사-명사 쌍 중에서 38,724쌍이 명사가 동사의 목적어로 쓰인 경우였다(3,820 개의 유일한 동사와 11,904의 유일한 명사). 명사가 동사의 목적어로 쓰인 동사-명사 쌍 중에서 그 빈도(frequency)가 명사는 20번 이상, 동사는 40-1000번 내인 쌍만을 선택

하였다. 그리고, 동사와 명사와의 관련성 측도는 상호 정보를 사용하였다[2]. 이렇게 추출된 동사-명사 쌍 중에서 본 실험에는 의미적으로 서로 상반되는 23개의 동사(11쌍)와 그 동사들과 관련이 있는 621개의 명사를 사용하였다. <표 1>은 실험에 사용된 동사 리스트를 나타내고 있다.

<표 1> 실험에 사용된 23개의 동사 리스트

#	동사	#	동사
1	가르치 배우	7	뭉
2	닫_2 열_2	8	사 팔
3	당기_1 밀	9	막 뭉
4	더하_2 뻗_1	10	모르 알
5	벗_2 입	11	잇 붙이
6	신		떼_1 자르_1

<표 2>는 12개의 독립성분과 유의수준 0.05 하에서 상관있는 동사들을 Box-Cox 변환 전과 후의 결과를 보여주고 있다. 밑줄로 표시된 동사들은 같은 독립성분(클러스터)에 포함된 <표 1>의 동사 쌍을 의미한다. 이러한 동사 쌍의 개수를 비교한다면, 변환 전에는 18쌍, 변환 후에는 36쌍이다. 결국 변환 후에 보다 많은 동사 쌍이 같은 클러스터에 포함되었다고 할 수 있다.

<표 2> 12개의 독립성분과 상관있는 동사 (유의수준 0.05)

IC	동사 (변환 전)	IC	동사 (변환 후)
1	더하_2, 뻗_1, 팔, 막, 알, (잇), 붙이, 떼_1	1	(가르치), 배우, 신, 뭉, 사, 팔
2	가르치, (배우), 뻗_1, 팔, 막, 붙이	2	신, 뭉, 사, 팔, 모르, (알), 떼_1
3	신, 팔, (막), 붙이, 떼_1	3	더하_2, 신, 풀, 팔, 떼_1, (자르_1)
4	(가르치), 팔, 막, 붙이	4	밀, 벗_2, 신, 뭉, 사, (팔), 떼_1
5	당기_1, (밀), 막, 붙이	5	열_2, 신, 뭉, 사, (팔), 막, 뭉
6	닫_2, 뻗_1, (뭉), 붙이	6	닫_2, (열_2), 더하_2, 신, 뭉, 사, 팔, 막, 뭉, 알
7	(뭉), 풀, 팔, 막, 붙이	7	더하_2, (뻗_1), 신, 뭉, 풀, 사, 모르
8	더하_2, 뻗_1, 벗_2, (입), 팔, 막, 잇, 붙이	8	신, 뭉, 사, 팔, 알, (잇)
9	팔, 막, (모르), 알, 붙이, 떼_1	9	당기_1, 더하_2, 뻗_1, 벗_2, (뭉), 풀, 사, 팔, 떼_1, 자르_1
10	더하_2, 뻗_1, 사, (팔), 막, 알, 잇, 붙이, 떼_1	10	신, 뭉, 사, (팔), 붙이
11	닫_2, (열_2), 팔, 막, 붙이	11	열_2, 벗_2, (입), 신, 뭉, 사, 팔, 떼_1
12	더하_2, 뻗_1, 팔, 붙이, (자르_1)	12	가르치, (배우), 신, 뭉, 사, 팔, 모르, 알

<표 3>은 15개의 독립성분과 유의수준 0.05 하에서 상관있는 동사들이 쌍을 <표 2>와 같은 방식으로 보여주고 있다. 같은 클러스터에 포함된 <표 1>의 동사 쌍은 변환 전과 후 모두가 22쌍이었다. 하지만, 동사의 개수와 같은 23개의 독립성분을 추출하여 비교한 결과, 변환 전에는 6쌍이, 변환 후에는 11쌍이 같은 클러스터에 포함됨을 확인할 수 있었다.

<표 3> 15개의 독립성분과 상관있는 동사  
(유의수준 0.05)

IC	동사 (변환 전)	IC	동사 (변환 후)
1	팔, (막), 잇, 붙이	1	열_2, 팔, 막, (붙이)
2	사, 팔, (모르), 알, 잇, 붙이	2	열_2, 사, (팔), 막, 풀
3	(가르치), 팔, 잇	3	당기_1, 밀, (더하_2), 풀, 풀, 모르, 알
4	사, (팔), 잇, 자르_1	4	가르치, (배우), 더하_2, 막
5	닫_2, (열_2), 팔	5	닫_2, (열_2), 막
6	당기_1, (벗_2), 팔	6	열_2, 더하_2, 팔, 막, (모르), 알
7	더하_2, (매_1), 실, 사, 팔, 알, 잇, 붙이	7	열_2, 더하_2, 벗_2, (신), 막, 풀, 모르
8	가르치, (배우), 팔, 잇	8	열_2, 당기_1, 더하_2, 매_1, 벗_2, (뉘), 풀, 막, 풀, 알, 매_1, 자르_1
9	사, 팔, (잇), 붙이	9	열_2, 벗_2, (열_2), 팔, 막, 풀
10	사, 팔, 잇, (붙이), 매_1	10	(열_2), 팔, 막, 풀
11	닫_2, 팔, (뉘)	11	열_2, 더하_2, 풀, 막, 모르, (자르_1)
12	당기_1, (밀), 팔	12	열_2, 더하_2, (매_1), 뉘, 팔, 막, 풀, 알
13	(뉘), 풀, 사, 팔	13	열_2, (사), 팔, 막
14	(신), 사, 팔, 잇, 붙이	14	당기_1, 밀, 더하_2, 매_1, 뉘, 막, 풀, 모르, (매_1)
15	(열_2), 사, 팔, 잇	15	막, 알, (잇)

<표 2>와 <표 3>에서 볼드체와 괄호로 묶인 동사가 각 클러스터와 가장 관련이 깊은, 클러스터의 대표 단어이다. <표 2>와 같이 12개의 클러스터를 추출할 경우, Box-Cox 변환 전에는 대표 단어와 쌍을 이루는 단어가 함께 추출된 경우는 1, 2, 5번 등 9개의 클러스터에서 나타났다. 하지만, 변환 후에는 8번 클러스터를 제외한 모든 경우에서 대표 단어와 함께 다른 단어도 추출되었다. 그런데, <표 3>과 같이 15개의 클러스터를 추출할 경우에는 오히려 변환 전에는 10쌍, 변환 후에는 9쌍으로 근소한 차이를 보였다. 이는, 본 논문에서 실험한 자료에는 고유치의 누적 기여율 임계치로는 0.7 이 보다

더 적당하며, 또한, 독립성분분석을 시행할 경우 추출하는 독립성분의 개수에 따라 그 성능에 차이가 있을 수 있음을 시사하고 있다.

## 6. 결론 및 향후 연구

본 논문에서는 독립성분분석과 Box-Cox 변환을 이용하여 한국어 동사의 개념적 클러스터링 방법을 제안하였다. Box-Cox 변환을 이용하여 mixture(동사)의 분포를 가우시안 또는 정규 분포에 근사시키고자 하였고, 추출된 독립성분과 동사간의 연관정도를 파악하는데 상관계수를 이용하였다. 서로 의미적으로 상반되지만 같은 개념을 표현한다고 볼 수 있는 23개의 동사에 대해서 실험한 결과, 12개의 독립성분(클러스터)을 추출하는 경우에 제안한 방법이 더 좋은 성능을 보이고 있음을 확인하였다.

본 논문에서 사용된 실험 자료는 단순히 사전 정의문에 기반한 타동사와의 직접목적어 관계에 있는 명사만을 이용하였다. 향후, 동사의 논항구조, 선택제약, 목적어 종류에 따른 처리도 감안하여 연구를 진행할 예정이다. 그리고, 서로 상반되는 동사, 예를 들어 “가다”와 “오다”와 같이 두 단어 모두 “움직임”을 나타내지만 그 방향은 다르다. 향후에는 이러한 방향성 정보까지 포함된 클러스터 구축에 관한 연구를 진행할 예정이다. 또한, 동사 클러스터링 결과를 기반으로 명사의 개념적 클러스터링 방법에 대해서도 연구를 진행할 예정이다.

## 감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 육성·지원사업의 연구결과로 수행되었음. 또한, 이 논문은 2006년 한국과학기술정보연구원에서 시행하는 위탁연구과제로 수행되었음.

## 참고 문헌

- [1] Bigham, E., Kuusisto, J. and Lagus, K., "ICA and SOM in Text Document Analysis," Proceedings of the 25th ACM SIGIR 2002 International Conference on Research and Development in Information Retrieval, Tampere, Finland, pp. 361-362, 2002.
- [2] Church, K. and Hanks, P., "Word Association Norms, mutual Information and Lexicography," Computational Linguistics, vol. 16, pp. 22-29, 1990.

- [3] Honkela, T., Hyvarinen, A. and Vayrynen, J., "Emergence of Linguistic Features: Independent Analysis of Contexts," In Proc. of the Neural Computation and Psychology Workshop 9, Plymouth, UK, 2005.
- [4] Hyvarinen, A., "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," IEEE Tr. on Neural Networks, Vol. 10, No. 3, pp. 626-634, 1999.
- [5] Hyvarinen, A., Karhunen, J. and Oja, E., "Independent Component Analysis," John Wiley & Sons, 2001.
- [6] Johnson, R. A. and Wichern, D. W., "Applied Multivariate Statistical Analysis," NJ: Prentice Hall, 1992.
- [7] Kolenda, T., Hansen, L. and Sigurdson, S., "Independent Components in Text," In M. Girolami, editor, Advances in Independent Component Analysis, Springer-Verlag, pp. 235-256, 2002.
- [8] Manning, C. and Schütze, H., "Foundations of Statistical Natural Language Processing," Cambridge, MA: MIT Press, 1999.
- [9] 노형진, "다변량분석 이론과 실제", 형설출판사, 2005.
- [10] 성웅현, "응용 다변량분석", 탐진출판사, 2002.