

# Topographic non-negative matrix factorization에 기반한 텍스트 문서로부터의 토픽 가시화

장정호\*                                  엄재홍\*\*<sup>o</sup>                                  장병탁\*\*  
\*Fraunhofer IAIS.KD                                  \*\*서울대학교 컴퓨터공학부  
jeongho.chang@gmail.com                                  jheom@bi.snu.ac.kr<sup>o</sup>                                  btzhang@bi.snu.ac.kr

## Topographic Non-negative Matrix Factorization for Topic Visualization from Text Documents

Jeong-Ho Chang\*                                  Jae-Hong Eom\*\*<sup>o</sup>                                  Byoung-Tak Zhang\*\*  
\*Fraunhofer IAIS.KD, \*\*School of Computer Science and Engineering, Seoul National University

### 요 약

Non-negative matrix factorization(NMF) 기법은 음이 아닌 값으로 구성된 데이터를 두 종류의 양의 행렬의 곱의 형식으로 분할하는 데이터 분석기법으로서, 텍스트마이닝, 바이오인포매틱스, 멀티미디어 데이터 분석 등에 활용되었다. 본 연구에서는 기본 NMF 기법에 기반하여 텍스트 문서로부터 토픽을 추출하고 동시에 이를 가시적으로 도시하기 위한 Topographic NMF (TNMF) 기법을 제안한다. TNMF에 의한 토픽 가시화는 데이터를 전체적인 관점에서 보다 직관적으로 파악하는데 도움이 될 수 있다. TNMF는 생성모델 관점에서 볼 때, 2개의 은닉층을 갖는 계층적 모델로 표현할 수 있으며, 상위 은닉층에서 하위 은닉층으로의 연결은 토픽공간상에서 토픽간의 전이확률 또는 이웃함수를 정의한다. TNMF에서의 학습은 전이확률값의 연속적 스케줄링 과정 속에서 반복적 파라미터 갱신 과정을 통해 학습이 이루어지는데, 파라미터 갱신은 기본 NMF 기반 학습 과정으로부터 유사한 형태로 유도될 수 있음을 보인다. 추가적으로 Probabilistic LSA에 기초한 토픽 가시화 기법 및 희소(sparse)한 해(解) 도출을 목적으로 한 non-smooth NMF 기법과의 연관성을 분석, 제시한다. NIPS 학회 논문 데이터에 대한 실험을 통해 제안된 방법론이 문서 내에 내재된 토픽들을 효과적으로 가시화할 수 있음을 제시한다.

### 1. 서 론

인터넷의 발달과 이에 따른 정보량의 폭발적 증가로 온라인 텍스트나 전자화된 문서의 양이 크게 증가함에 따라 자동화된 텍스트 분석에 대한 요구가 증대되고 있으며, 이를 위해 기계학습이나 통계적 알고리즘에 기반한 방법론에 연구가 많이 진행되고 있다. 특히 잠재 토픽 모델(latent topic model)은 단어들의 공기(cooccurrence) 특성에 기반하여 텍스트 문서 데이터에 내재된 특징적 패턴 또는 단어간의 의미있는 상관관계를 자동학습에 의해 추출함으로써, 대규모 데이터에 대한 직관적 이해 및 손쉬운 정보 파악을 가능케 한다. 대표적으로, LSA (latent semantic analysis), PLSA (probabilistic latent semantic analysis)[8], NMF (non-negative matrix factorization)[11] 등이 제안되었으며, 이 중 NMF와 PLSA는 단어들에 대한 양의 가중치만의 집합으로 토픽을 정의함으로써, 무제한(음의 가중치도 존재) 선형 분석에 기반한 LSA에 비해 보다 해석 가능한 형태의 토픽들을 추출할 수 있다는 장점이 있다.

특히, NMF 기법은 음이 아닌 값으로 구성된 데이터를 저차원의 두 양의 행렬로 분할하는 기법으로 토픽 추출 면에서의 유용성[11]과 더불어 추출된 토픽을 기반으로 한 문서군 집화 성능은  $k$ -means 군집화 등의 일반적인 군집화 모델에 비해 보다 우월할 수 있다는 연구결과도 제시되었다[14]. 또한 바이오인포매틱스 분야에 대한 비슷한, 즉 특징적 유전자 발현 패턴 추출[10] 및 샘플 군집화[2, 5] 응용을 통해서도

그 유용성이 제시되고 있다. 하지만, NMF 기법에서는 추출된 토픽에 대한 가시화 기능은 제공하지 않는데, SOM (self-organizing map)에 의한 텍스트 문서 분석[9]에서처럼 2-3차원 상에 잠재 토픽들의 가시화가 가능하다면 대규모 텍스트 문서 집합 내에 내재된 의미 있는 주제들을 직관적이고 효율적으로 파악하는데 도움이 될 수 있다.

본 논문에서는 이를 위하여 기본 NMF 기법에 기초한 텍스트 문서로부터 토픽을 추출과 동시에 이를 가시적으로 도시하기 위한 Topographic NMF (TNMF) 기법을 제안하며, 그 학습 알고리즘은 기본 NMF에서와 유사한 방식으로 유도할 수 있음을 보인다. 또한 기존 PLSA에 기반한 토픽 추출/가시화 모델인 ProbMap과의 연관성을 그 목적함수 및 학습 알고리즘 면에서 분석하며, 행렬 분할 면에서 동일한, sparse NMF의 한 종류인 nsNMF와의 관련성을 제시한다. 그리고, 학회 논문 데이터를 대상으로 제안하는 TNMF 기법의 잠재 토픽 가시화 면에서의 유용성을 검증한다.

### 2. Non-negative matrix factorization

Non-negative matrix factorization (NMF)[10]은 음이 아닌 값으로 구성된 데이터를 두 개의 저차원 양의 행렬의 곱으로 분할한다. 전체  $M$ 개의 문서로 구성된 문서 집합  $V = [v_1, v_2, \dots, v_M]$ ,  $v_i = (v_{ij})_{j=1}^N$ 에 대해 NMF는  $V = WH$ 의 분할에 의해  $V$ 를 근사화한다.  $N$ 은 어휘집 크기이며  $W$ 와  $H$ 는 각각  $N \times K$ ,  $K \times M$ 행렬이다.  $K$ 는 보통  $K \ll N, M$ 으로

설정되는데, 텍스트마이닝 관점에서 텍스트 문서 집합으로부터 추출되는 토픽의 개수라고 할 수 있다. 텍스트 문서에 대한 잠재토픽모델로서의 NMF에 대해,  $W$ 의 각 열은 하나의 토픽을 정의하는 단어들의 가중치 리스트이며,  $H$ 의 각 열은 단어 공간상에 표현된 문서의  $K$ 차원 토픽 공간상으로서의 사상에 해당한다. 그림 1은 이 절차를 역으로 토픽 공간에서 단어공간으로서의 사상, 즉 생성 모델의 관점에서 NMF를 네트워크 구조로 제시한다.

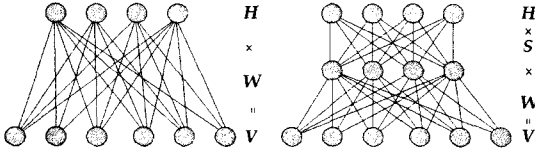


그림 1 생성모델 관점에서의 NMF(좌)와 TNMF(우)

$W$ 와  $H$ 의 모든 원소의 값이 음이 아닌 값을 갖도록 함으로써, NMF는 텍스트 문서에 대해 하나 이상의 요인(factor), 즉 토픽들의 가합적 생성을 가정한다. 이에 기초하여, [11]에서는 이미지 데이터와 더불어 텍스트 문서 데이터에 대해 NMF 기법이 PCA (principal component analysis)나 VQ (vector quantization)와는 달리 데이터 집합으로부터 전체적 특징이 아닌 부분적 특징들을 요약적으로 추출할 수 있음을 제시하였다.

NMF에서의 학습은 I-divergence (generalized KL divergence) 또는  $L_2$ -norm(행렬에 대해서는 Frobenius norm)에 기반한 목적함수를 최소화하는 것으로 정의되며 [12], 확률 모델에서의 EM 알고리즘과 유사한 방식으로  $W$ 와  $H$ 를 반복적으로 추정함으로써 학습이 이루어진다. 목적함수가 I-divergence 인 경우, NMF는 다음의  $J_{nmf}$ 를 최소화한다.

$$J_{nmf} = \sum_{m=1}^M \sum_{n=1}^N \left( v_{nm} \log \frac{v_{nm}}{(WH)_{nm}} - v_{nm} + (WH)_{nm} \right)$$

$J_{nmf}$ 의 최소화로 정의될 때 NMF는 비슷한 시기에 제안된 PLSA (probabilistic latent semantic analysis)와 유사하며 (3.3절 참조), 학습은 다음과 같이 이루어진다[11].

$$W \leftarrow W \odot [(V \odot (WH))H^T] \oslash (1_{N \times M} H^T) \quad (1)$$

$$H \leftarrow H \odot [W^T (V \odot (WH))] \oslash (W^T 1_{N \times M}) \quad (2)$$

위 식에서  $\odot$ 과  $\oslash$ 는 각각 두 행렬에서 같은 인덱스를 갖는 원소들 간의 곱과 나누기를 의미한다.  $1_{N \times M}$ 은 모든 원소 값이 1인  $N \times M$ 행렬이다.

### 3. Topographic non-negative matrix factorization

기본적인 NMF 기법은 데이터로부터 토픽들을 추출하지만, 이를 가시화하기 위한 방법은 포함하지 않는다. 토픽 추출과 더불어 저차원 상에 이러한 토픽들을 가시화함으로써, 텍스트 문서 집합 내에 내재된 토픽들의 직관적이고 효과적인 파악에 도움이 될 수 있다. Topographic NMF (TNMF)는 이를

보완하여 토픽 공간상에서의 토픽들 간의 확률적 전이 개념을 도입하여 학습 과정에서 토픽 추출과 더불어 이를 토픽 맵 상에 가시화할 수 있도록 한다.

#### 3.1 모델의 구조

그림 1은 생성모델(generative model) 관점에서 기본 NMF와 대비하여 제안하는 TNMF를 네트워크 구조로 제시한다. 토픽공간상에서 하나의 토픽 구성  $h$ 가 주어질 때, TNMF 구조상에서  $h$ 는 토픽간의 전이확률  $S$ 에 의해  $hs$ 라는 새로운 토픽 구성을 생성하고 각 토픽들의 가합적 조합으로써, 하나의 데이터가 생성된다고 가정한다. 역으로 설명하면, TNMF에서 데이터  $v$ 는  $v = WSH$ 에 의해 근사화된다. 여기서,  $S$ 는 데이터 전송상의 하나의 noisy channel로 가정될 수도 있다. 구조적 측면에서 볼 때 TNMF는 계층적 NMF [1]의 한 예로 볼 수 있으며, 다만 두 은닉층의 개수가 동일하고 은닉층을 연결하는 가중치 행렬이 고정되거나 미리 설계된 스케줄링에 의해 변화되므로, 학습의 대상이 아니라는 점에서 제한된 형태의 모델이라 할 수 있다.

#### 3.2 모델의 학습

TNMF에서 전이확률 행렬  $S$ 는 고정되거나 미리 스케줄링되므로, 학습은 NMF에서와 같이  $W$ 행렬과  $H$ 행렬의 추정으로 정의된다. 최소화하고자 하는 목적함수가 I-divergence인 경우, 즉

$$J_{tnmf} = \sum_{m=1}^M \sum_{n=1}^N \left( v_{nm} \log \frac{v_{nm}}{(WSH)_{nm}} - v_{nm} + (WSH)_{nm} \right),$$

[12]에서 제시된 보조함수(auxiliary function)에 기초한 방법을 적용하여 매개변수  $W$ 와  $H$ 는 식 (3)과 (4)의 반복적/교대 갱신에 의해 추정된다.

$$W \leftarrow W \odot [(V \odot (WSH))(SH)^T] \oslash (1_{N \times M} (SH)^T) \quad (3)$$

$$H \leftarrow H \odot [(WS)^T (V \odot (WSH))] \oslash ((WS)^T 1_{N \times M}) \quad (4)$$

다시 말해, 기본 NMF 기법의  $W$  갱신 식에서  $H$  대신에  $S \times H$ 가 이용되며,  $H$  갱신 식에서  $W$  대신에  $W \times S$ 가 이용되는 것이다.

TNMF가 목적으로 하는 은닉공간에서의 토픽 맵 생성에서 핵심 사항은 전이확률 행렬  $S$ 라고 할 수 있는데, 본 논문에서  $S$ 는 그래프 이론에 기초하여 생성하였다.  $R \times C (= K)$  격자 구조에 의한 2차원 토픽 맵을 가질 때, 이로부터 하나의 그래프  $G$ 를 생성한다. 정점  $a_{ij} (1 \leq i \leq R, 1 \leq j \leq C)$ 는 하나의 잠재 토픽에 해당하며 정점간의 연결은 다음과 같이 설정한다(그림 2).

$$e(a_{ij}, a_{kl}) = \begin{cases} 1 & \text{if } |i-k| + |j-l| = 1 \\ 0 & \text{otherwise} \end{cases}$$

그리고, 이렇게 표현된 그래프  $G$ 의 인접행렬을  $A$ 라 할 때, 이로부터 graph Laplacian  $L = D - A$ 에 대한 graph exponential에 의해 전이함수 행렬  $S = \exp(-\beta L)$ 를 최종적으로 생성한다.  $\beta (\geq 0)$ 는 상수로서 이를 변경함으로써 토픽

간 전이확률 정도를 조정하는, 즉  $S$ 를 변경하는 효과를 얻을 수 있다.  $\beta$ 의 값이 클수록 다른 토픽으로의 전이확률 정도는 커지며,  $\beta$ 의 값이 작을수록 전이확률은 줄어든다. 궁극적으로  $\beta \rightarrow \infty$ 일 때  $S_{ij} = 1/K (\forall i, j)$ 이며,  $\beta = 0$ 일 때  $S=I$ 가 되고 이 경우 TNMF는 기본 NMF와 동일하게 된다.

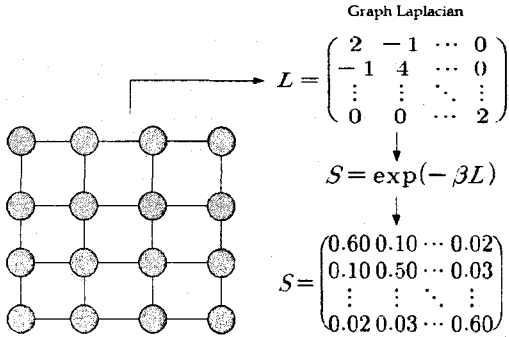


그림 2 Graph Laplacian에 기반한 토픽 맵 상에서의 토픽 간 전이확률 행렬  $S$ 구축

TNMF 학습 과정상에서 토픽맵을 구성하기 위해  $\beta$ 값은 충분히 큰 값  $\beta_{start}$ 에서 시작하여 미리 설정된 스케줄링 ( $\beta^{new} = \tau \beta^{old} (0 < \tau < 1)$ )에 의해 그 값을 줄여나가면서,  $\beta \leq \beta_{final}$ 를 만족할 때까지 진행된다. 이러한 전이확률행렬 스케줄링은 SOM (self organizing map)에서의 이웃함수 (neighborhood function)의 조정 과정과 유사하다. 주어진  $\beta$  값, 즉 전이확률 행렬  $S$ 에 대해서는 목적함수 값이 수렴할 때까지 식 (3)과 (4)를 반복적으로 수행한다. 표 1은 전체 학습 과정을 요약하여 제시한다.

표 1. Topographic NMF의 전체 학습 절차

- \* INPUT:  $V, \beta_{start}, \tau (0 < \tau < 1)$
- \* LEARNING PROCEDURE
- 1. Initialize  $W$  and  $H$ .  
 $\beta = \beta_{start}$
- 2. do  
 $S = \exp(-\beta L)$   
 do  
 update  $W$  according to Eq. (3)  
 update  $H$  according to Eq. (4)  
 until converged  
 $\beta \leftarrow \tau \beta$   
 until ( $\beta \leq \beta_{final}$ )
- \* OUTPUT:  $W, H$

### 3.3 관련 연구

Hofmann[7]은 PLSA (Probabilistic Latent Semantic Analysis)[8]기법에 기초하여 텍스트 문서로부터 토픽을 추

출하고 이를 가시화하는 확률적 토픽 매핑 기법(ProbMap)을 제안하였다. PLSA 기법은 모델의 데이터 생성에 대한 가정과 그 목적함수 면에서 NMF 기법과 상당한 공통점이 있으며, 학습 알고리즘 측면에서 볼 때 KL divergence에 기반한 NMF 학습 결과는 PLSA의 EM 알고리즘에 기초한 (지역) 최대 우도(maximum likelihood) 해의 하나에 해당한다 [4, 6].

여기에서는 ProbMap 기법과 TNMF의 관계를 목적함수 및 모델 매개변수(parameter) 학습 면에서 제시하고자 한다. ProbMap에서는 다음과 같은 로그 우도 형태의 목적함수를 최대화함으로써 모델을 학습한다.

$$J_{pmap} = \sum_{m=1}^M \sum_{n=1}^N n(d_m, w_n) \log p(w_n | d_m)$$

$$= \sum_{m=1}^M \sum_{n=1}^N n(d_m, w_n) \log \sum_k p(w_n | z_k) \sum_k p(z_k | z_k') p(z_k' | d_m)$$

$v_{nm} = c(d_m, w_n) / \sum_{m,n} c(d_m, w_n)$ ,  $p(d_m) = \sum_n v_{nm} / \sum_m \sum_n v_{nm}$ 로 설정할 때,  $J_{pmap}$ 에 기반한 I-divergence 식을 다음과 같이 유도할 수 있다 ( $C = \sum_{m,n} c(d_m, w_n)$ ).

$$J_{div} = -J_{pmap} / C + constant$$

$$= -J_{pmap} / C + \left[ \sum_{m,n} v_{nm} \log v_{nm} - v_{nm} \log p(d_m) - (v_{nm} - p(d_m, w_n)) \right]$$

$$= \sum_{m,n} \left( v_{nm} \log \frac{v_{nm}}{p(d_m, w_n)} - v_{nm} + p(d_m, w_n) \right)$$

$$p(d_m, w_n) = \sum_{z, z'} p(w_n | z) p(z | z') p(z' | d_m) p(d_m)$$

따라서  $J_{pmap}$ 의 최대화는  $J_{div}$ 의 최소화에 해당하며,  $W_{nk} = p(w_n | z_k)$ ,  $S_{lk} = p(z_l | z_k)$ ,  $H_{mk} = p(d_m, z_k)$ 로 설정할 경우, ProbMap와 I-divergence 기반 TNMF의 목적함수는 동일하다고 할 수 있다.

매개변수 추정 과정 면에서 보면, ProbMap 모델은 EM 알고리즘에 의해 다음과 같이  $p(w_n | z_k)$ 와  $p(z_k | d_m)$ 를 반복적으로 추정함으로써 학습되며,

$$p(w_n | z)^{(t+1)} \propto \sum_{m=1}^M v_{nm} p_l^{(t+1)}(z | d_m, w_n) \quad (5)$$

$$p(z | d_m)^{(t+1)} = \frac{\sum_{n=1}^N v_{nm} p_u^{(t+1)}(z | d_m, w_n)}{\sum_{n=1}^N v_{nm}} \quad (6)$$

$p_u(z | d_m, w_n)$ 과  $p_l(z | d_m, w_n)$ 은

$$p_l^{(t+1)}(z | d_m, w_n) = \frac{p^{(t)}(w | z) \sum_y p(z | z') p^{(t)}(z' | d_m)}{\sum_y p^{(t)}(w | y) \sum_y p(y | z') p^{(t)}(z' | d_m)} \quad (7)$$

$$p_u^{(t+1)}(z | d_m, w_n) = \frac{p^{(t)}(z | d_m) \sum_y p(z' | y) p^{(t)}(w | z')}{\sum_y p^{(t)}(y | d_m) \sum_y p(z' | y) p^{(t)}(w | z')} \quad (8)$$

$$1) p(z_k | d_m) p(d_m) = p(d_m, z_k)$$

식 (7), (8)을 식 (5), (6)에 각각 대입하고, 앞서 목적함수의 경우와 같이 TNMF에서의 표현법으로 변환하면,

$$W_{nk}^{(t+1)} \propto \frac{\sum_{m=1}^M W_{nk}^{(t)} v_{nm} \sum_l S_{lk} H_{lm}^{(t)}}{\sum_k W_{nk}^{(t)} \sum_l S_{lk} H_{lm}^{(t)}} \quad (9)$$

$$H_{km}^{(t+1)} = \frac{\sum_{n=1}^N H_{km}^{(t)} v_{nm} \sum_l S_{kl} W_{nl}^{(t)}}{\sum_k H_{km}^{(t)} v_{nm} \sum_l S_{kl} W_{nl}^{(t)}} \quad (10)$$

따라서 TNMF에서 행렬  $W$ 의 각 열에 대해  $\sum_{n=1}^N W_{nk} = 1$ 이 되도록 표준화하고 식 (3), (4)와 식 (9), (10)의 대개변수 갱신을 대비하면,  $W^{(t+1)} \approx W^{(t)}$ ,  $H^{(t+1)} \approx H^{(t)}$ 일 경우 고정된 전이확률  $S$ 에 대해 ProbMap의 지역 해는 TNMF의 지역 해의 하나임을 알 수 있다. 특히,  $S = \exp(-\beta L)$ 에서  $\beta = 0$ , 즉  $S=I$ 일 경우  $W$ 의 열 기반의 표준화(식 (7), (9)에서는 생략된) 면에서도 TNMF와 ProbMap은 같은 학습 식을 따른다.<sup>2)</sup>

결론적으로, 고정된 개수의 잠재 토픽에 대해 I-divergence에 기초한 TNMF는 ProbMap의 최대 우도 기반 학습과 밀접한 관련이 있다고 할 수 있다. 하지만, NMF 기법이 Euclidean 오차에 기초한 목적 함수[12]를 포함한 다양한 Bregman divergence 기반의 목적함수에 의해서도 정의될 수 있듯이, TNMF 역시 ProbMap 기법에서의 다항분포 (multinomial distribution) 기반과는 다른 형태의 목적함수에 대해서나, 데이터 객체/자질에 대한 가중치 도입 또는 전처리 등에 의해 데이터 값이 더 이상 빈도수 형태로 주어지지 않을 경우에도 행렬 기반의 근사화 관점에서 자연스럽게 정의되고 구체화될 수 있다는 이점이 있다.

주어진 데이터의 근사화를 위한 행렬 분할 표현 면에서 볼 때, TNMF는 non-smooth NMF (nsNMF)[3, 13]와 동일하게 표현된다. nsNMF 기법은 토픽 자질( $W$ ) 및 데이터 사상( $H$ ) 면에서 원래의 NMF에 비해 보다 희소한(sparse) 특성을 갖는 해(solution)을 도출하기 위해 제안된 기법으로서  $W = VSH$ 에 의한 데이터 근사화에서 행렬  $S$ 는 다음과 같이 정의된다.

$$S_{ij} = \begin{cases} 1 - \frac{K-1}{K} \lambda & i = j \\ \frac{1}{K} \lambda & i \neq j \end{cases} \quad (1 \leq i, j \leq K, 0 \leq \lambda \leq 1)$$

결국, 이것은 잠재변수  $i$ 로부터 일정 부분을 감하여 이를 다른 잠재 변수들에 동등하게 할당하는 것이며,  $\sum_j S_{ij} = 1$ ,  $S_{ij} \geq 0$ 을 만족하므로 잠재변수들 간의 전이확률로 생각할

2) 엄밀하게는,  $W$  갱신에서의  $H$  이용과  $H$  갱신에서의  $W$  이용 방식에서 ProbMap은 이전 갱신에서의 결과를, TNMF는 현재 갱신에서의 결과를 즉시 이용한다. 따라서 지역해 근처에서는 각 갱신 결과가 비슷하게 도출되나 [4]에서 제시한 PLSA와 NMF 학습 결과의 실험적 분석에서도 유추할 수 있듯이, 비록 같은 초기치(지역 해가 아닌)로부터 학습이 진행된다 하더라도 해에 이르는 경로상의 차이로 인해 ProbMap 기법과 TNMF는 서로 다른 지역 해에 이를 수 있다.

수 있다. 이를 통해  $W$ 와  $SH$ 를 평활(smooth)하게 함으로써 NMF 기법이 가합적(additive) 해를 탐색하는 학습과정상의 내재적 특성을 이용하여  $H$ 와  $W$ 에 대한 보다 희소한(sparse) 해를 얻고자 하는 것이 그 목적이다. 하지만, TNMF 상에서의 행렬  $S$ 는 학습 과정에서 체계적으로 변화되며 맵 상에서 보다 이웃한 잠재변수가 다른 이웃 변수보다 보다 높은 할당치를 갖도록 설정됨으로써 이를 통해 최종적으로 토픽들에 대한 topographic map을 구축하기 위한 것이라는 점에서, 고정된 값의 평활 행렬  $S$  도입으로 희소한 특성의 해를 얻는 것을 목적으로 하는 nsNMF 기법과 구별된다.

#### 4. 실험 및 결과

NIPS (Neural Information Processing Systems) 학회 논문 데이터<sup>3)</sup>를 대상으로 TNMF의 토픽 가시화에 관한 실험을 하였다. 이 텍스트 집합은 1987년부터 1999년까지 13년간의 NIPS 학회에 발표된 1,740개의 문서로 구성되어 있으며, 어휘집 크기는 13,649이다. 토픽은  $9 \times 9 (=81)$  크기의 2차원 맵 상에 표현되도록 하였고,  $\beta$ 값은 초기값을 10.0으로 하여  $\beta^{new} = 0.9 \times \beta^{old}$  스케줄링에 의해  $\beta^{new} > 10^{-3}$ 을 만족할 때까지 학습을 진행하였다. 각  $\beta$ 값에 대해 TNMF의 학습은 목적함수 값의 상대적 변화량이  $5 \times 10^{-6}$  이하이거나 식 (3)과 (4)에 의한 학습 반복수가 700에 이를 때까지 진행하였다.

그림 3은 TNMF 모델 학습 진행에 따른 목적함수 값의 변화, 즉 학습 곡선을 보인다. 목적함수 값이 계단식으로 다소 급격하게 감소하는 지점(-o-)은  $\beta$ 값의 스케줄링에 의해 전이확률행렬  $S$ 가 새롭게 조정된 지점을 의미한다. 이러한 계단형 변화 지점의 등장은, 토픽간 전이확률 폭이 좁혀짐으로써 이전 단계에 비해 보다 세분화된 토픽 추출이 가능해지고 결과적으로 주어진 학습데이터에 대해서 보다 더 적합한 학습 환경이 갖추어지기 때문인 것으로 사료된다.

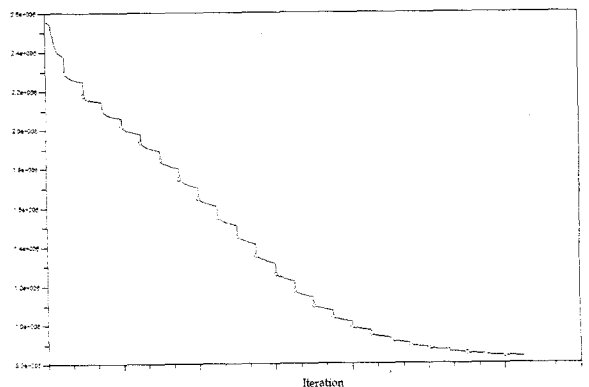


그림 3. NIPS학회 논문데이터에 대한 TNMF의 학습곡선

그림 3은 TNMF 학습이 종료된 후의 2차원 토픽 맵을 제시한다. 하나의 토픽당 가중치가 높은 상위 5개의 단어들을 나열하였다. 토픽 맵의 구성을 보면 관련 있는 연구 주제들은

3) <http://www.cs.toronto.edu/~roweis/data.html>

대체로 맵 상에서 인접 위치에 배열되어 있음을 알 수 있다. 예를 들어 오른쪽 하단의 "강화학습(reinforcement learning)"에 관한 토픽을 살펴보면, 그 인접 셀에 이론 면에서 "마코프 모델(Markov model)", 응용 면에서 "로봇 제어", "게임" 등의 관련 주제들이 배열되어 있음을 알 수 있다. 다시 "로봇 제어" 관련 주제는 그 좌측에 "vision" 관련, 다시 "vision" 관련 토픽은 그 주위에 "시각관련 뇌신경" 연구에 관련된 토픽들이 포진됨을 볼 수 있다. "인공 신경망" 관련 주제들은 주로 토픽 맵의 좌측 상단에 위치하며, pca와 ica 관련 토픽은 "신호 처리(signal processing)", "이미지 처리" 관련 연구 주제와 인접해서 배열된다. 이와 같은 실험 결과는 Topographic NMF 기법이 기본 NMF의 텍스트 문서에 대한 토픽 추출 기능을 바탕으로 동시에, 저차원 공간상에 의미 있는 토픽 맵을 효과적으로 구축, 제시함을 보인다.

5. 결론

본 논문에서는 Topographic NMF 기법을 이용한, 텍스트 문서 집합으로부터의 토픽 가시화 기법을 제시하였다. TNMF에 의한 토픽 맵의 구성 기능은 전이확률 행렬의 도입과 이로부터의 순차적 학습에 의해 달성되며, 학습 과정은 기본 NMF에서와 동일한 방식으로 유도 가능하다. 또한 목적함수 및 데이터 분할형식 측면에서 ProbMap 및 nsNMF와의 관련성을 제시하였으며 실제 데이터에 대한 실험을 통해 토픽 가시화 측면에서 실효성을 보였다. 텍스트마이닝이나 바이오인포매틱스 분야에서는 최근 데이터를 정의하는 서로 다른 요인들을 통합 분석하기 위한 연구가 활발한데, 향후 다중 데이터를 동시에 가시적으로 분석 가능한 방향으로 모델을 확장하고자 한다.

structure	state	units	training	function	probability	algorithm	kernel	bayesian
rules	recurrent	unit	error	functions	bound	learning	support	distribution
language	input	hidden	generalization	neural	learning	loss	vector	gaussian
representation	time	learning	learning	threshold	bounds	examples	set	posterior
connectionist	network	weights	weight	networks	theorem	function	svm	prior
representations	node	network	learning	dynamics	convergence	optimal	function	model
model	nodes	input	gradient	system	stochastic	function	data	data
memory	network	layer	algorithm	point	equation	approximation	model	models
code	input	output	error	equations	theorem	error	method	likelihood
representation	activation	training	weight	networks	function	functions	estimate	em
memory	network	network	time	energy	graph	data	data	data
capacity	time	neural	neural	optimization	problem	clustering	local	experts
associative	routing	networks	dynamics	function	algorithm	cluster	regression	variance
patterns	fault	results	system	bolzmann	local	algorithm	model	prediction
stored	traffic	net	state	problem	path	clusters	function	set
parallel	neuron	control	model	matrix	space	rbf	genetic	tree
processor	neurons	system	time	vector	points	basis	algorithms	set
neural	learning	controller	nonlinear	linear	point	function	search	data
processors	neural	model	state	vectors	dimensional	functions	algorithm	trees
bit	pattern	neural	prediction	space	mapping	radial	problem	training
circuit	analog	noise	signal	learning	vector	distance	classification	rules
input	chip	input	signals	pca	vectors	tangent	set	set
voltage	visi	optical	noise	principal	map	vectors	classifier	learning
weight	neural	noisy	net	input	kohonen	transformations	training	rule
figure	digital	output	linear	rule	input	set	data	data
system	circuit	channel	source	basis	level	recognition	training	hmm
delay	figure	time	sources	sparse	matching	digit	detection	recognition
neuron	voltage	template	separation	data	model	digits	false	system
phase	output	signal	blind	projection	hierarchical	network	region	context
circuit	current	channels	independent	learning	hierarchy	neural	set	speech
neurons	spike	auditory	filter	image	feature	character	word	speech
neuron	information	sound	filters	wavelet	features	characters	words	recognition
firing	rate	frequency	ica	images	image	network	recognition	speaker
time	noise	model	natural	disparity	images	recognition	letter	phoneme
model	time	time	information	figure	resolution	set	character	acoustic
synaptic	stimulus	eeg	phase	image	face	system	learning	state
cell	response	time	model	images	images	user	network	markov
cells	stimuli	activity	visual	visual	faces	hand	game	model
neurons	population	data	oscillators	surface	image	road	training	states
synapses	information	brain	oscillator	color	recognition	tracking	tasks	time
model	cells	motion	visual	object	motor	robot	learning	state
orientation	cell	eye	attention	objects	model	environment	reinforcement	learning
cortex	model	model	motion	model	control	goal	model	policy
dominance	orientation	visual	figure	recognition	learning	learning	system	action
ocular	visual	velocity	direction	view	movement	position	time	function

그림 4. NIPS 학회 논문 데이터에 대한 TNMF 학습 결과 (토픽 맵의 크기는 9×9)

## 참고문헌

- [1] Ahn, J. H., Choi, S. and Oh, J. H., A multiplicative up-propagation algorithm. In *Proceedings of the International Conference on Machine Learning*, pp. 17-24, 2004.
- [2] Brunet, J.-P., Tamayo, P., Golub, T. R., and Mesirov, J. P., Megagenes and molecular pattern discovery using matrix factorization, *Proceedings of the National Academy of Sciences of the USA (PNAS)*, vol. 101, no. 12, 2004
- [3] Carmona-Saez, P., Pascual-Marqui, R. D., Tirado, F., Carazo, J. M. and Pascual-Montano, Biclustering of gene expression data by non-smooth non-negative matrix factorization, *BMC Bioinformatics*, vol. 7, no. 78, 2006.
- [4] Ding, C., Li, T., and Peng, W., Nonnegative matrix factorization and probabilistic latent semantic indexing: equivalence, chi-square statistic, and a hybrid method, In *Proceedings of AAAI 2006*, 2006.
- [5] Gao, Y. and Church, G., Improving molecular cancer class discovery through sparse non-negative matrix factorization, *Bioinformatics*, vol. 21, no. 21, pp. 3970-3975, 2005.
- [6] Gaussier, E. and Goutte, C., Relation between PLSA and NMF and implications, In *Proceedings of SIGIR'05*, pp. 601-602, 2005.
- [7] Hofmann, T., ProbMap - a probabilistic approach for mapping large document collections, *Journal of Intelligent Data Analysis*, vol. 4, pp. 149-164, 2000.
- [8] Hofmann, T., Unsupervised learning by probabilistic latent semantic analysis, *Machine Learning*, vol. 42, pp. 177-196, 2001.
- [9] Kaski, S, Honkela, T., Lagus, K., and Kohonen, T., WEBSOM-self-organizing maps of document collections. *Neurocomputing*, vol. 21, pp. 101-117, 1998.
- [10] Kim, P. M. and Tidor, B., Subsystem identification through dimensionality reduction of large-scale gene expression data, *Genome Research*, vol. 13, pp. 1706-1718, 2003.
- [11] Lee, D. D. and Seung, H. S., Learning the parts of objects by non-negative matrix factorization, *Nature*, vol 401, pp. 788-791, 1999.
- [12] Lee, D. D. and Seung, H. S., Algorithms for non-negative matrix factorization, In *Advances in Neural Information Processing Systems 13*. 2001.
- [13] Pascual-Montano, A., Carazo, J. M., Kochi, K., Lehmann, D., and Pascual-Marqui, R. D., Non-smooth non-negative matrix factorization (nsNMF), *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 403-415, 2006.
- [14] Xu, W., Liu, X., and Gong, Y., Document-clustering based on non-negative matrix factorization, In *Proceedings of SIGIR'03*, pp. 267-273, 2003.