

에너지 기반 모델을 이용한 순차 패턴 부호화 방법

허민오⁰¹ 김권일² 이상우¹ 장병탁^{1,2}

서울대학교 컴퓨터공학부¹ 서울대학교 뇌과학 협동과정²

{moheo, kikim, slee, btzhang}@bi.snu.ac.kr

An Encoding Method of Sequential Patterns using Energy-based models

Min-Oh Heo⁰¹ Kwonill Kim² Sang-Woo Lee¹ Byoung-Tak Zhang^{1,2}

School of Computer Science & Engineering, Seoul National University¹

Interdisciplinary Program in Neuroscience, Seoul National University²

요 약

시계열 데이터 모델링은 시간 간격의 길이에 따라 단기적인 패턴이 주로 반영된다. 특히, 모델에 마코프 가정을 적용하였을 경우 이전 시간의 값에 따라 현재값이 결정된다. 시계열 데이터의 장기적인 변화를 다루기 위해, 특정 길이의 순차적 패턴을 부호화 하고, 이를 상위 모델의 입력으로 사용하는 과정을 통해 추상화를 시도하고자 한다. 실제로 사람의 감각기억은 200~500 밀리초 가량의 짧은 기억 유지기간을 갖는데, 이 기간의 정보를 상위 처리기의 입력 단위로 보고자 하는 것이다. 이에 본 고에서는 에너지 기반 모델링 기법을 이용하여 반복적으로 나타나는 순차적 패턴을 부호화 하는 방법을 제안한다. 이 부호화 방법은 시간 순서에 따른 패턴의 유사도를 이용하여 확률적으로 다음 패턴과의 관계를 표현할 수 있으며, 이는 향후 시계열 데이터를 간략하게 표현하여 분석 및 시각화에 도움을 줄 수 있다.

1. 서 론

자연계에는 음성이나 비디오와 같이 다양한 형태, 다량의 시계열 형태의 순차적 데이터가 존재한다. 이러한 데이터를 모델링 하는 것은 매우 중요한 문제이다. 흔히 사용하는 모델링 방법으로 재귀적 뉴럴 네트워크(recurrent neural network (RNN))[1], 은닉 마코프 모델(hidden Markov model (HMM))[2]이 사용되는데, 주로 시간 간격의 차이가 가까운 단기적인 패턴이 학습된다. 특히, HMM과 같이 모델에 마코프 가정을 적용하였을 경우, 이전 시간의 값에 따라 현재값이 결정된다.

시계열 데이터의 장기적인 변화를 다루기 위해서는, 고차원 마코프 가정을 도입하여, 비교적 먼 시간 간격의 관계까지 모델에 표현하도록 하는 것이 필요하다[2,3]. 하지만, 단순히 고려해야 할 이전 시간 길이를 늘리게 될 경우, 모델의 복잡도가 지수적으로 증가한다. 이러한 현상을 회피하기 위해, 반복적으로 나타나는 순차적 패턴은 부호화가 가능하다는 가정을 도입한다. 실제로 사람의 기억 구조에서 가장 외부 입력 정보에 가까운 부분인 감각기억(sensory memory)은 200~500 밀리초 가량의 짧은 기억 유지기간을 가지는데[4], 이 기간의 정보를 인지 과정상 상위 처리기의 입력 단위로 보고자 한다.

본 고에서는 이러한 가정을 바탕으로 에너지 기반 모델링 기법을 이용하여 반복적으로 나타나는 순차적 패턴을 부호화 하는 방법을 제안한다. 에너지 기반 모델은 데이터를 표현하는 변수와 은닉 변수 사이의 관계를 에너지 형태로 표현하고, Gibbs 분포를

취함으로써 확률분포로 표현할 수 있는 장점이 있다. 또한, 시간 순서에 따른 패턴의 유사도를 이용하여 확률적으로 다음 패턴과의 관계를 표현할 수 있으며 시계열 데이터를 간략하게 표현하여 분석 및 시각화에 도움을 줄 수 있다.

이후, 본 논문은 다음과 같이 구성된다. 2 장에서는 시계열 데이터의 부호화 방법과 관련 기반 모델을 설명하고, 소개한 방법론의 특성을 설명한다. 3 장에서는 실험을 통해 부호화의 성능을 보인다.

2. 시계열 데이터의 부호화 방법

2.1 모델 구조

데이터를 부호화 하고 그 결과를 다음 단계를 위한 입력으로 사용하고자 하는 계층적 접근방법은 사람의 인지과정과 유사하다는 점에서 많은 관심을 받고 있다[3,4]. 이러한 접근은 영상 데이터에 주로 적용되었으며[3], 시계열 데이터에도 도입된 사례[5,6]가 있으나 제한된 데이터에 국한되었다.

특히, 시계열 데이터는 순서가 존재하므로 일정 시간 구간 내에 나타나는 시간 순서에 따른 값의 차이를 기준으로 반복적인 패턴이 나타날 수 있으며, 이는 사람이 감각기억(sensory memory) 내에 저장된 데이터를 인식을 위한 기초정보로 활용하는 것과 유사하다. 이러한 점을 반영하기 위해, 에너지 기반 모델링 방법을 이용하여 시계열 데이터를 표현할 수 있도록 모델을 설계하고, 각 코드에 해당하는 은닉 변수를 정의하였다. 에너지 기반 모델링은 에너지에 대한 Gibb 분포를

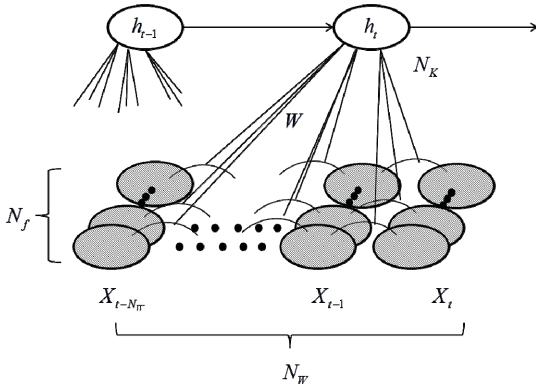


그림 1. 제안 모델의 구조. $t-N_w$ 시점에서 t 시점 사이의 각 시점 별 인접한 시점과의 패턴의 차를 W 로 표현하고 가장 가까운 패턴을 나타내는 h_t 가 선택된다.

취하면 확률을 얻을 수 있다는 장점이 있다[7].

임의의 시간 t 에 대해 길이 N_w 인 패턴과 은닉변수 h_t 와의 관계를 다음과 같이 에너지를 이용하여 정의한다. h_t^k 는 h_t 의 k 번째 값을 나타내고, $X_{t-N_w:t}$ 는 $t-N_w$ 시점부터 t 시점까지의 관측변수들을 나타낸다.

$$E(\mathbf{h}_t, \mathbf{X}_{t-N_w:t}) = \sum_{k=1}^{N_k} h_t^k \sum_{f=1}^{N_f} \sum_{\tau=0}^{N_w-1} ((X_{t-\tau}^f - X_{t-\tau-1}^f) - W_{\tau+1}^{f,k})^2$$

그림 1에서 볼 수 있듯이 $t-1$ 시점의 패턴과 t 시점 사이의 패턴의 차를 W 로 표현하고, 이 차이가 작아질수록 전체 에너지도 작아지며, 그러한 패턴을 N_k 개 갖게 된다. N_f 는 인자의 개수이다.

임의의 시계열 $X_{1:T}$ 에 대한 에너지는 아래와 같다.

$$E(\mathbf{h}_{1:T-N_w+1}, \mathbf{X}_{1:T}) = \sum_{t=1}^{T-N_w+1} \sum_{k=1}^{N_k} h_t^k \sum_{f=1}^{N_f} \sum_{\tau=0}^{N_w-1} ((X_{t-\tau}^f - X_{t-\tau-1}^f) - W_{\tau+1}^{f,k})^2$$

시계열 입력값이 주어졌을 때, 은닉변수 h_t^k 에 대한 조건부 확률은 h_t^k 가 0일 때의 에너지와 1일 때의 에너지를 이용하여 확률식을 유도하면, 아래와 같은 sigmoid 함수 형태를 갖는다.

$$P(h_t^k = 1 | \mathbf{h}_t^{-k} = 0, \mathbf{X}_{t-N_w:t}) = \frac{\exp(-E(\mathbf{h}_t, \mathbf{X}_{t-N_w:t}))}{1 + \exp(-E(\mathbf{h}_t, \mathbf{X}_{t-N_w:t}))}$$

이 확률로부터 h_t^k 중에서 가장 큰 확률을 갖는 h_t^k 만 1이 되고 나머지는 0이 되도록 하는 연산인 max-pooling을 수행하여 실제 h_t 를 얻으며 부호와 대응된다. 즉, h_t 는 원소중 하나만 1이고 나머지는 0이 되는 벡터이다.

또한, h_t 가 주어졌을 경우, 관측 시계열 값에 대한 조건부 확률은 다음과 같다.

$$P(\mathbf{X}_{t-N_w:t} | \mathbf{h}_t) = \frac{1}{Z} \exp(-\sum_k h_t^k \sum_f \sum_{\tau} ((X_{t-\tau}^f - X_{t-\tau-1}^f) - W_{\tau+1}^{f,k})^2)$$

이 식은 h_t^k 값이 1인 항만 남고 나머지는 사라지며, 완전 제곱식이므로 multivariate Gaussian 분포가 된다.

확률식을 다시 작성하면 아래와 같다.

$$P(\mathbf{X}_{t-N_w:t} | h_t^k = 1) = \mathcal{N}(\mathbf{W}^k, \alpha \mathbf{I})$$

이전 부호가 주어졌을 때 다음 시점에 부호에 대한 확률은 유사도 S 를 이용하여 다음과 같이 정의한다.

$$P(h_t^j = 1 | h_{t-1}^k = 1) = \frac{S(\mathbf{W}^k, \mathbf{W}^j)}{\sum_j S(\mathbf{W}^k, \mathbf{W}^j)}$$

여기서 이전 시점의 부호가 갖는 패턴 W^{prev} 와 현재 시점의 부호가 갖는 패턴 $W^{current}$ 사이의 유사도는 아래와 같이 정의한다.

$$S(\mathbf{W}^{prev}, \mathbf{W}^{current}) = 1 / (\epsilon + dist(\mathbf{W}_{2:N_w}^{prev}, \mathbf{W}_{1:N_w-1}^{current}))$$

두 패턴 간의 거리를 구하는 함수 $dist$ 는 Euclidean distance를 사용하였다. ϵ 은 두 패턴 간의 거리가 0일 경우 계산적인 문제를 해결하기 위한 상수이다.

2.2 파라미터 학습 방법

상기 모델에서 파라미터 W 를 학습시키기 위해 학습 데이터 집합의 각 인스턴스에 대한 에너지의 합을 최소화 해야 하는 목적함수로 삼았으며, 식으로 표현하면 아래와 같다. (T_i 는 i 번째의 시계열 데이터의 길이이다.)

$$L = \sum_{i=1}^{|D|} E(\mathbf{h}_{1:T_i-N_w+1}, \mathbf{X}_{1:T_i})$$

목적함수 L 을 최소화 하기 위해 위 식을 각 파라미터에 대해 미분하여 목적함수가 작아지는 방향으로 이동하는 방법인 gradient descent 방법을 이용한다. L 은 2차 미분가능하며, Hessian이 diagonal이고 각 파라미터의 2차 미분값이 0보다 크거나 같아 positive semi-definite이 되므로 local minimum에 항상 수렴한다. 아래는 L 의 파라미터에 대한 미분항이다.

$$\frac{\partial L}{\partial W_{\tau+1}^{f,k}} = -2 \left\langle \sum_t \sum_k h_t^k \sum_f \sum_{\tau} ((X_{t-\tau}^f - X_{t-\tau-1}^f) - W_{\tau+1}^{f,k}) \right\rangle_D$$

2.3 시계열 데이터의 부호화 방법

본 절에서는 앞서 소개한 모델과 파라미터 학습 방법을 이용하여 시계열 데이터를 부호화 하는 방법을 아래에 제시한다. (그림 2)

- 1) 초기의 패턴 수 N_k 를 1 이상의 값으로 정한다.
- 2) 시계열 데이터를 $T_{시작}=1$ 에서 $T_{종료}=N_w$ 시점까지 모델에 입력으로 제시하고 파라미터 학습을 한다.
- 3) 모델로부터 입력데이터에 대한 부호 시계열 $h_{1:T_{종료}-N_w+1}$ 를 얻고 이로부터 시계열 데이터를 생성(복호화)한다.
- 4) 원본 데이터와 생성한 시계열 데이터의 오차가 δ 보다 크면, N_k 를 하나 늘리고, 최근 시계열로부터

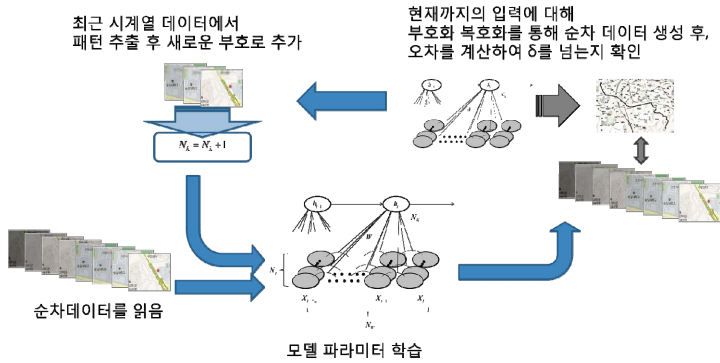


그림 2. 시계열 데이터로부터 순차 패턴 부호화 학습 과정. 현재까지의 입력데이터와 그의 부호화 및 복호화를 통해 생성된 데이터와의 차이가 δ 보다 크면 가장 최근 시계열 데이터에서 패턴을 추출하여 추가된 h 를 위한 w 의 초기값으로 설정 후 모델 학습을 한다.

패턴을 추출하여 h^{N_k+1} 에 대한 w 의 초기값으로 설정한 후, 파라미터 학습을 수행한다.

5) $T_{종료}$ 를 1 늘리고 모든 데이터에 대해 3)~5)를 반복한다.

학습과정에서 부호를 늘릴 때, 데이터로부터 패턴을 추출하여 파라미터 값에 초기값으로 주는 것을 통해, global minimum 근처에서 파라미터 학습을 시작하게 되어 시간 비용을 줄이고 학습된 부호의 정확도는 증가한다.

3. 실험 결과

실험을 통해, 유사한 규칙성과 반복성을 지닌 시계열 데이터가 지속적으로 입력 될 때, 제시한 방법이 이력적은 오차 이내로 잘 표현할 수 있는지 확인하고, 부호의 수가 어떻게 변화하는지 보인다. 이를 위해 [8]에서 사용된 데이터 일부에 대해 부호화를 수행하였다. 이 데이터는 피험자가 실외에서 주요 거점 간을 이동한 라이프로그를 스마트폰으로 1초 간격으로 수집한 것이며, 이 중 8 건의 GPS 궤적(총 23118초)을 이용하였다. 또한, N_w 는 3으로 설정하였다.

그림 3은 부호화/복호화 과정을 거쳐 얻은 결과와 원본을 지도상에 오차값과 함께 도시한 것이며, 그림 4 (좌)는 데이터를 지속적으로 입력함에 따라 총 부호 수의 변화를 보인다. 하나의 시계열 데이터를 부호화 하여 얻은 부호의 수보다 다수의 시계열 데이터를 함께 부호화 할 경우, 중복되어 표현되는 패턴은 기존의 것을 이용하므로 부호의 수가 크게 늘어나지 않는 것을 확인할 수 있다. 또한, 그림 4 (우)에서 볼 수 있듯이 각 부호와 연관되는 패턴을 확인할 수 있어 분석에 용이하다.

감사의 글

이 논문은 지식경제부 및 한국산업기술평가관리원의 산업원천기술개발사업(10035348, mLife), 교육과학기술부 국가연구재단의 지원을 받아 수행된 연구(No. 2012-0005643) 및 BK21-IT사업에 의해 일부 지원되었음.

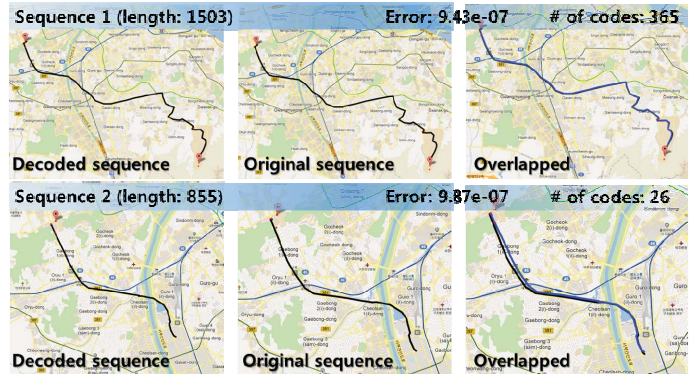


그림 3. 부호화/복호화를 거친 시계열 데이터 예. (좌: 복호화를 통해 생성된 시계열, 중앙: 원본, 우: 두 시계열을 겹쳐 표시)

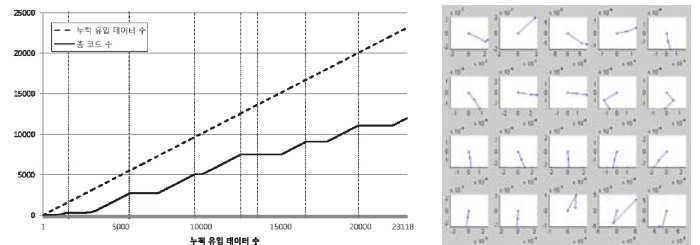


그림 4. 시퀀스 총 길이를 학습함에 따른 부호 수 변화(좌). 세로 보조선은 다른 궤적들이 입력되는 시점을 표시한 것임. 부호와 연관된 패턴 예 20 개(우). 각 패턴들은 이전 시점과의 차이를 나타내며, N_w 가 3이므로 $t-2$, $t-1$, t 시점 간의 정보를 길이와 방향으로 표현하고 있다.

참고문헌

- [1] Simon Haykin, Neural Networks and Learning Machine, Prentice Hall, 2009.
- [2] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
- [3] G. E. Hinton and R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, *Science*, vol. 313, no. 5786, pp.504-507, 2006.
- [4] M. S. Gazzaniga, R. B. Ivry, G. R. Mangun, Cognitive Neuroscience: The Biology of the Mind (3rd ed.), Norton, 2009.
- [5] I. Sutskever, G. E. Hinton, and G. W. Taylor, The Recurrent Temporal Restricted Boltzmann, *Proc. in 22nd Advances in Neural Information Processing Systems (NIPS 2009)*, 2009.
- [6] G. W. Taylor, G. E. Hinton, S. Roweis, Modeling Human Motion using Latent Variables, *Proc. in 19th Advances in Neural Information Processing Systems (NIPS 2006)*, 2006.
- [7] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. Jie Huang, A Tutorial on Energy-based Learning, in book "Predicting Structured Data", MIT Press, 2006.
- [8] 허민오, 강명구, 임병권, 황규백, 박영택, 장병탁, 확률 그래프 모델을 이용한 스마트폰 사용자의 이동경로 학습 및 실시간 예측 기법, *한국정보과학회 가을학술발표 논문집*, 제38권, 2(B), pp. 248-251, 2011.