TV 드라마 배경 전환 탐지를 위한 베이지안 필터링 방법

유준희○, 석호식, 장병탁
서울대학교 컴퓨터공학부
{jhyoo, hsseok, btzhang}@bi.snu.ac.kr

Bayesian Filtering for Background Change Detection in TV dramas

Jum Hee Yoo○, Ho-Sik Seok, Byoung-Tak Zhang
School of Computer Science and Engineering
Seoul National University

요 약

이미지 인식 기법과 컴퓨팅 능력의 발달과 함께 비디오 동영상 처리를 위해 다양한 기법들이 연구되었다. 그러나 기존 연구 기법들은 특정 이미지의 반복과 같은 처리 데이터에 특화된 사전 지식을 요구하는 경우가 많았기 때문에 현재 발생하고 있는 다양한 비구조 데이터 처리에 한계를 지니고 있다. 또한 데이터의 구조 및 분포에 대한 가정에 오류가 있으면 데이터를 정확하게 근사하지 못하게 되어 정확한 모델을 본 논문에서는 이런 문제점을 해결할 수 있도록 데이터에 대한 사전지식을 최소한만 활용하고 실제 데이터의 분석 결과에 기반하여 기저 분포를 추정하는 방법을 소개한다. 제안 방법은 파티클 필터링(particle filtering)에 기반한 것으로 많은 수의 파티클 및 파티클과 연관된 가중치를 이용하여 관찰된 데이터를 설명할 수 있는 은닉 변수(latent variable) 모델을 설정한 후 은닉 변수가 이미지를 생성할 가능성을 감안하여 이미지 변화를 추정하는 방법이다. 파티클 필터링을 이용하여 데이터 분포를 추정할 경우 데이터 분포 형태를 가정하지 않고도 분포 추정이 가능하므로 다양한 변화가 발생하는 데이터 처리에 매우 유용하다. 본 논문에서는 주어진 드라마 동영상의 배경 이미지 변화 추정에 제안 방법을 적용하였다. 제안 방법을 이용하여 동일 동영상 배경 구간을 설명할 수 있는 모델을 구성한 후 모델에 기반하여 새로 관찰된 장면(Scene)이 새로운 배경을 갖고 있을 가능성을 계산하였으며, 제안 방법의 성능을 확인하기 위하여 모델이 예측한 배경 전환 지점과 인간 실험자의 배경 전환 판단 결과를 비교·분석하였다.

1. 서론

멀티미디어 스트림을 효과적으로 다루려면 멀티미디어 스트림의 이벤트를 분석할 수 있는 방법이 필요하다. 감지(Detection), 구분(Segmentation), 인식(Recognition), 주석(Annotation) 등 다양한 이벤트 분석 업무의 특성에 따라 적합한 분석 방법을 사용할 수 있겠지만[1] 멀티미디어 스트림 분석 과정에 있어 항상 문제가 되는 것은 사전 지식의 사용 및 생성 모델(Generative model)과 분류 모델(Discriminative model)의 선택이다[2]. 기존의 접근 방법에서처럼 메타데이터를 사용하거나, 반복 시퀀스 존재를 활용하거나, 혹은 참조 데이터베이스(Reference Database)를 작성하는 경우, 멀티미디어 스트림 분석 정확도를 높일 수는 있지만 예상하지 못한 새로운 데이터를 처리하기가 힘들고, 메타데이터 및 참조 데이터베이스를 작성해 놓아야 한다는 단점이 있다[3]. 생성 모델은 현상을 설명할 수 있는 장점을 갖추고 있지만 데이터 분포 추정이 어려운 관계로 쉽게 사용할 수 없다는 단점이 있다[2].

본 논문에서는 파티클 필터(Particle filter) [4, 5]를 이용하여 데이터가 속한 분포를 추정한 후 구성된 생성 모델을 통해 멀티미디어 스트림을 처리하는 방법을 소개한다. 멀티미디어 스트림을 구성하는 분포를 사전에 가정하는 경우 추정 성능에 제약이 클 수 있으므로 파티클 필터를 이용하여 데이터 분포를 추정하는 방법은 다양한 가능성을 갖게 된다. 물론 서로 다른 데이터 표현 방법을 이용하여 시간성을 갖는 데이터(Temporal data)의 클러스터를 구하는 것도 가능하지만[6] 이 방법은 온라인 데이터 처리가 어렵기 때문에 멀티미디어 스트림 데이터의 다양한 용도에 적용하기 어렵다.

제안 방법을 다양한 멀티미디어 스트림 관련 작업에서도 특히 배경 구간 구분(Segmentation) 문제에 적용하였다. 배경 구간 구분 문제는 동영상의 배경 변환 시점을 판단하는 문제로 카메라 시점 변화, 조명의 변화 등으로 인해 동일 배경의 특성 값이 달라지기 때문에, 비주얼 워드(Visual word) 등의 하위 수준 특성만을 이용해서 추정하기는 어려운 문제이다. 제안 방법에서는 데이터 분포에 대하여 사전 제약을 설정하지 않고 파티클 필터를 이용하여 데이터의 분포를 추정한 후 추정된 분포에서 현재 관찰한 데이터가 생성될 우도(Likelihood)를 계산하는 방법[7]을 활용하여 배경 변화 시점을 추정하였다.

데이터 스트림에서 새로운 배경의 등장을 찾는다는 점에서 제안 방법론은 새로운 클래스를 감지하는 [8]의 방법과 유사하다고 할 수 있다. 그러나 [8]에서는 기존 클래스 및 해당 레이블의 존재를 가정하고 있는 반면, 제안 방법에서는 어떤 배경이 몇 개나 존재하는지 전혀 모르는 상황에서 새로운 배경의 등장 시점을 추정하므로 문제의 난이도가 한층 높아진다.

본 논문에서는 제안 방법의 추정 결과를 인간 실험자의 배경 전환 판단 결과에 비교하여 성능을 제시하였다.

2. 동영상에서 배경 변환시점 추정

그림 1 배경 전환의 예(회사에서 법정으로 장소 전환)

표 1 인간 실험자가 판단한 배경전환 횟수

	Episode I	Episode II	Episode III
실험자 A	35	33	38
실험자 B	25	27	26
실험자 C	39	37	33
실험자 D	40	40	45

표 2 공통적으로 판단한 배경 변환 시점의 수

	공통변환점 2	공통변환점 3	공통변환점 4
Episode I	4	12	21
Episode II	5	14	18
Episode III	6	15	14

사람은 영화나 TV 드라마를 시청하면서 카메라 시점의 변화, 조명의 변화에 상관 없이 동일한 장소를 쉽게 인식할 수 있다. 그러나 컴퓨터로 스트림 데이터를 처리하는 경우 RGB 값 등의 특성이 변하게 되므로 컴퓨터는 전혀 다른 장소, 개체로 인식하게 된다. 우리는 시점, 조명 등의 변화에 상관 없는 동일 장소 인식 결과를 얻기 위하여, 배경의 정의를 사전에 제한하지 않고 네 명의 인간 실험자에게 주어진 동영상[1]에서 한 스토리가 다른 스토리로 변환되는 시점을 판단하도록 요청하였다.

그림 2와 표 1, 2에서 인간 실험자의 배경 변환 시점 판단 결과를 정리하였다. 표 1에서 동일한 동영상에 대하여 개인별 판단 결과에 편차가 크다는 사실을 알 수 있으며, 표 2를 통해 인간 실험자들이 공통으로 판단한 배경 변환 시점의 수도 다양하다는 것을 알 수 있다. 우리는 인간 실험자가 판단한 배경 변환 결과를 모두 유효한 것으로 간주하고 제안 방법론을 적용하여 배경 변환 시점을 추정하였다.

3. 배경 변환시점 추정 방법과 추정 결과

그림 3에서 제안 방법을 도시하였다. 제안 방법에서는 시점 에서 관찰된 데이터 는 해당 데이터에 대한 은닉 변수 X_t에서 생성되며, 동일 배경 공간에 해당하는 영상의 은닉 변수는 θ로 설명할 수 있다고 가정하고, L번째 배경 공간에 대한 은닉 변수 θ 을 파티클 필터로 추정하고자 한다. 배경 공간 변환시점을 추정하기 위해서는 배경에 해당하는 데이터만을 추출하여 분석할 필요가 있다. 이를 위해 데이터를 사전 처리해야 하는데, 본 논문에서 사용한 데이터는 에피소드 3에 해당하는 이미지에 대하여 1초 당 10장의 이미지를 추출한 후 SIFT (Scale Invariant Feature Transform) [9]를 이용하여 시각 특성이 추출된 데이터이다. 추출된 시각 특성을 이용하여 1000개의 시각 단어를 구성하여 각각의 이미지를 시각 단어의 히스토그램으로 변환하였다. SIFT 방법론이 어느 정도 변화에 강한 특성을 추출하기는 하지만, 예를 들어 동일 장면에서 등장 인물을 줌인(Zoom-in)하게 되면 새로운 특성으로 처리될 수 있으므로 배경 전환이 발생하였다고 판단할 수 있다. 본 논문에서 제안한 방법은 이런 상황까지는 처리할 수 없지만, 특별히 등장 인물이나 특정 사물을 강조하지 않는 상황에서는 SIFT 처리된 데이터에 기반하여 은닉 변수를 추정함으로써 히스토그램의 급격한 변화에도 불구하고 동일 배경 지속 구간 및 배경변환 시점을 추정할 수 있는 방법론이다.

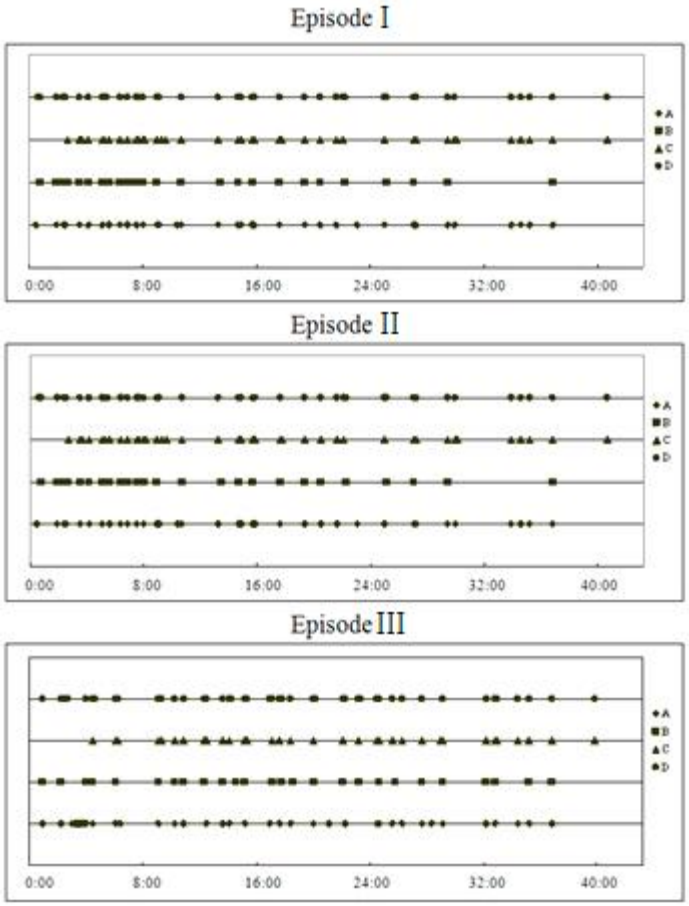

그림 2 인간 실험자의 배경 변환 시점 판단 결과

1) 본 연구에서는 20세기 폭스 텔레비전이 ABC를 위해 제작한 미국의 법률 드라마인 보스턴 리걸(Boston Legal)의 에피소드 3개를 실험용 동영상으로 사용하였다.

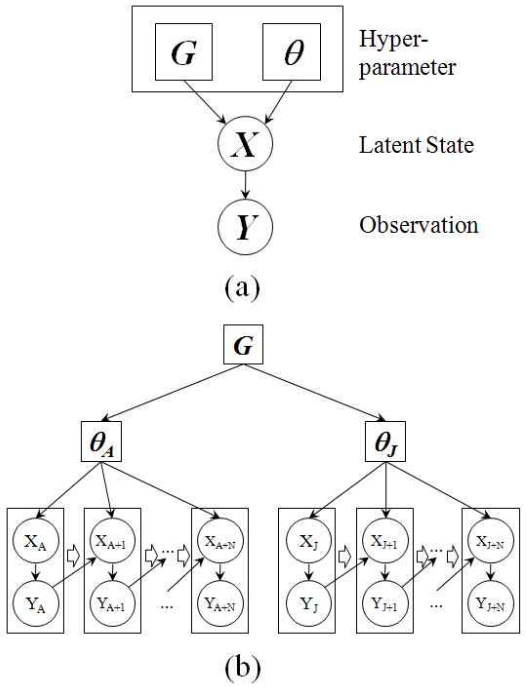

그림 3 배경 변환 시점 판단을 위한 파티클 필터 모델. 는 관찰된 데이터를 의미하며, X는 Y를 생성하는 은닉 변수에 해당한다. 은닉 변수 X는 하이퍼인자 $G,\theta\}$의 영향을 받게 되는데, θ는 동일 배경 공간을 설명하는 은닉 변수를 의미하고 G는 현재 진행 중인 전체 드라마 스트림을 나타낸다.

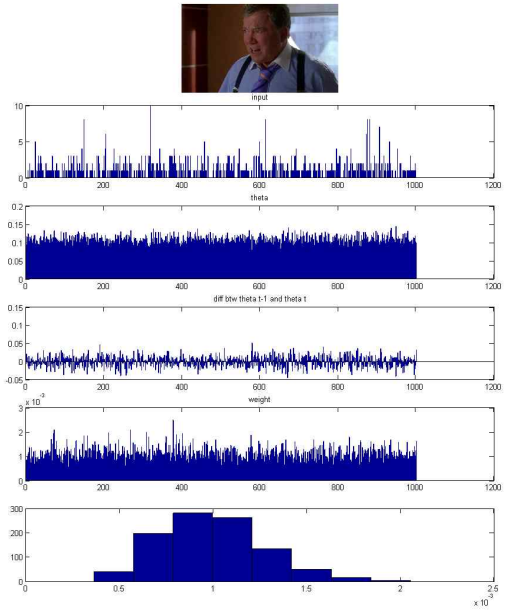

그림 4 실험 화면. 가장 상단의 도표가 연속된 이미지간 히스토그램 차이를 표시. 비주얼 워드 입력 수준에서 빈번한 변화가 발생함을 알 수 있다.

표 3 배경 변환시점 판단을 위한 파티클필터 방법론

은닉 변수 $X_{:} = x_0,\ldots,x_T\}$
관찰 데이터 $Y_{0:T} = \{y_o,\ldots,y_T\}$
사전 하이퍼인자 $\Theta = G,\theta\}$
θ_L: L번째 배경 공간을 설명하는 은닉 변수

1. 초기화 N_S개의 파티클(ξ) 생성(N_s: 파티클 수)
- $\xi_0^i \sim q$ (q: 중요 분포(Importance density))
- 각 파티클에 가중치 부여

단계 2 ~ 5 반복
2. 중요 샘플링(Importance sampling)
- 파티클 샘플 추출$\xi_k^i \sim q(\xi_k|\xi_{k-1}^i, y_k)$
- 가중치 부여

$$w_k^i = w_{k-1}^i \frac{p(y_k|\xi_k^i)p(\xi_k^i|\xi_{k-1}^i)}{q(\xi_k^i|\xi_{S\ :k-1}^i, y_{S_{L\ k}})}$$

3. 반복샘플링(Resampling)
- 가중치 정규화
- 악화정도(Degeneracy)에 따라 파티클 재샘플링

4. 갱신(Update)

$$p(x_k|y_{s_L:k}) = \frac{p(y_k|x_k)\cdot p(x_k|y_{s_L:k-1})}{p(y_k|y_{s_L:k-1})} = \sum_{i=1}^{N_s} w_k^i \cdot \delta(x_{s_L:k} - \xi_{s_{L\ k}}^i)$$

여기서 $\delta(\cdot)$는 dirac delta 함수

5. 추정(Estimate)

$$g(y_{k+1}, y_k, \theta_L) = \ln(y_{t+1}|\theta_L) = \begin{cases} 1 & \text{정기준이상일 때} \\ 0 & \text{일정기준이하일 때} \end{cases}$$

표 3에서 제안 방법을 설명하였다. 파티클 생성에서 파티클 추출 및 갱신까지의 방법은 기본적인 파티클 필터링 방법과 동일하다. 단 시점 $t+1$에서 관측한 새로운 데이터 y_{t+1}이 기존 배경 구간에 속할 가능성을 계속 계산하고 있다가 가능성이 일정 기준 이하이면 새로운 배경이 시작되었다고 간주하고 파티클 초기화 과정부터 다시 시작하여 배경 구간을 설명하는 은닉 변수를 추정한다.

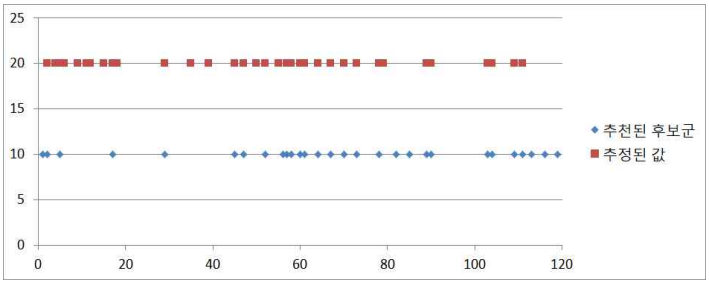

그림 5 100초 구간에 대한 배경 변환시점 추정 결과

본 논문에서는 제안 방법론의 추정 성능을 빠르게 확인해 볼 수 있도록 에피소드 I의 100초 구간을 선정한 후 배경 변환시점을 추정해 보았다. 그림 5는 선정된 구간에 대한 배경 변환시점 추정 결과이다. 인간 실험자들이 실험을 수행할 때 실제 배경변환시점에 대한 반응 지연시간이 존재하게 되므로 컴퓨터가 추정한 변환시점에 대하여 ±1초의 구간에 인간실험자의 변환시점이 존재하면 정확하게 추정하였다고 판단하였다. 실험 대상인 100초 구간에 대하여 인간 실험자는 총 37개 시점에서 장면이 전환되었다고 판단하였으며, 제안 방법이 추정한 장면전환 시점은 120개였다.

$$\text{확도} \quad \frac{\#(\text{인간판단화면전환시점})}{\#(\text{제안방법론판단시점})} \quad ...(1)$$

정확도를 식(1)과 같이 정의할 때 28.3%의 정확도로 변환 시점 후보군을 생성했음을 확인할 수 있었다. 제안 방법이 추정한 장면전환 시점이 실제 사람이 판단한 시점보다 많이 생성된 것은, 사람은 장면 안의 상황을 인지하여 등장인물의 이동이나 사물의 클로즈 업 등에 강건히 판단할 수 있으나 컴퓨터의 경우 동일한 배경에서도 장면을 구성하는 사물들의 이동이 큰 경우 화면이 바뀌었다고 판단하는 것을 볼 수 있었다.

4. 결론 및 토의

본 논문에서는 대상 멀티미디어 스트림 데이터에 대한 최소한의 사전 지식만을 이용하여 데이터의 분포를 추정한 후 추정된 분포에 기반하여 멀티미디어 스트림 데이터를 분할하는 방법을 소개하였다. 파티클 필터에 기반한 데이터 추정 방법은 데이터 분포에 사전 제약을 가하지 않고 분포를 추정하기 때문에, 동영상과 같은 비구조적인 실세계 데이터를 다루기에 매우 적합한 방법이다. 우리는 TV 드라마 에피소드에 제안 방법을 적용하여 배경변환 시점을 추정하는 작업을 통해 제안 방법의 성능을 확인하였다. 인간실험자의 변환판단시점 결과를 모두 참이라고 가정하고 배경 변환시점을 추정한 결과 실험 대상 구간에서 28.3%의 정확도를 갖는 변환 시점 후보군을 생성할 수 있었다.

제안 방법은 비주얼 워드나 픽셀 수준에서의 변화와 같이 하위 수준의 특성 변화에도 불구하고 유지되는 상위 수준의 특성 파악을 위하여 데이터를 설명하는 은닉 변수의 존재를 가정한 후 분포를 추정하는 방법이다. 본 논문에서는 추정한 분포 모델을 이용하여 관찰 데이터를 설명할 수 있는 가능성을 이용하여 상위 수준의 특성을 포착하였으며, 원 데이터(raw data)수준에서 빈번하게 발생하는 특성 변화에 강한 추정 방법을 만들어내기 위해, 현시점에서 확보한 모델이 현 시점에서 관찰된 데이터를 설명할 수 있는 가능성에 기반하여 배경 전환을 판단하였다.

인간 실험자의 판단결과와 비교했을 때 제안 방법을 이용하여 어느 정도 경쟁력이 있는 추정 결과를 확보할 수 있었으나, 추가 성능 개선의 여지가 여전히 많이 남은 상태이다. 특히 본 논문에서는 추정된 모델로 새로 관찰된 데이터를 설명할 수 있는 가능성을 계산할 때 사전 결정된 기준을 활영하였는데, 추후 연구에서는 기준점을 동적으로 결정할 수 있는 방법을 반영해야 할 것이다. 또한 추후 연구에서는 우도 계산 과정의 추세 변화를 변환 시점 추정에 반영하여 추정 정확도를 높일 수 있는 방법을 연구하고자 한다.

감사의 글

이 논문은 교육과학기술부의 재원으로 국가연구재단의 지원을 받아 수행된 연구(２０１１－００１６４８３, Videome)이며, 한국학술진흥재단 (３１４－２００８－１－D00377, Xtran) 및 교육과학기술부의 BK21-IT 사업에 의해 일부 지원되었음.

참고문헌

[1] L. Xie, H. Sundaram, and M. Campbell, “Event Mining in Multimedia Streams“, *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 623 - 647, 2008.

[2] S. J. Gershman, D. M. Blei, and Y. Niv, "Context, Learning, and Extinction", *Psychological Review*, Vol. 117, No. 1, pp. 197 - 209, 2010.

[3] G. Manson and S.-A. Berrani, "Automatic TV Broadcast Structuring", *International Journal of Digital Multimedia Broadcasting*, Vol. 2010, Article ID 153160, 2010.

[4] D. S. Lee and N. K. K. Chia, "A Particle Algorithm for Sequential Bayesian Parameter Estimation and Model Selection", *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, pp. 326 - 336, 2002.

[5] C. Andrieu, A. Doucet, S. S. Singh, and V. B. Tadic, "Particle Methods for Change Detection, System Identification, and Control", *Proceedings of the IEEE*, Vol. 92, No. 3, pp. 423 - 438, 2004.

[6] Y. Yang and K. Chen, "Temporal Data Clustering via Weighted Clustering Ensemble with Different Representations", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 2, pp. 307 - 320, 2011.

[7] I. J. Myung, "Tutorial on Maximum Likelihood Estimation", *Journal of Mathematicla Psychology*, Vol. 47, pp. 90 - 100, 2003.

[8] M. M. Masud, J. Gao, L. Khan, J. Han, and B. Thuraisingham, "Classification and Novel Class Detection in Concept-Drifting Data Streams under Time Constraints", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 6, pp. 859 - 874, 2011.

[9] David G. Lowe, "Object recognition from local scaleinvariant features," Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol.2, no, pp.1150-1157 vol.2, 1999 doi: 10.1109/ICCV.1999.790410.