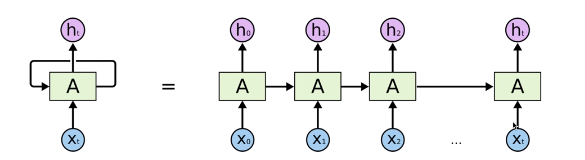
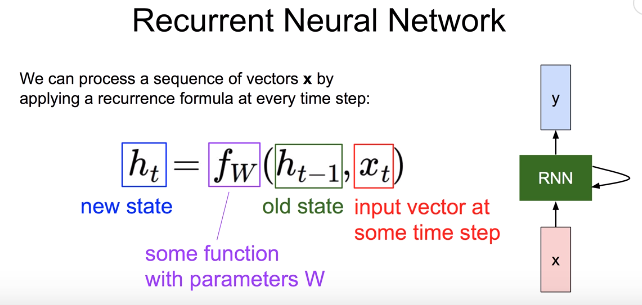
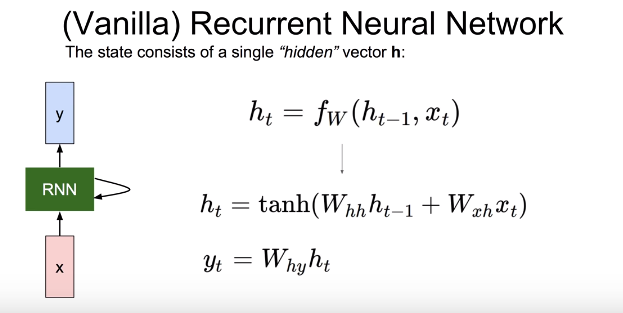
\*RNN – sequence data



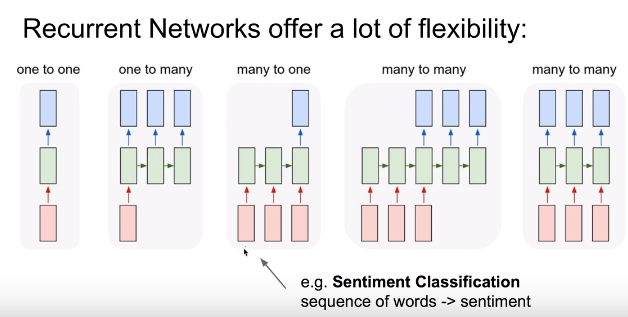
입력 x0, 출력 h0(y에 해당)



같은 함수와 같은 파라메터 셋이 매 time step마다 이용된다.



hello 예측 – h->e / e->l / l->l / l->o



**2. Weight Initialization**

- Restricted boltzmann machine (RBM) 방법으로 network를 초기화 시킨 것을 Deep Belief Nets이라고 부른다.

- 'Restricted'라는 단어는 같은 layer의 노드들끼리는 연결이 되지 않는 것을 뜻하기 위해서 사용되었다.

- RBM은 'recreate input'의 철학이 학습이 되는데, feed-forward로 입력에 대한 출력 결과를 다시 feed-backward로 입력에 대한 예측값을 만들어 내어 이 예측값이 실제 입력값과 일치하도록 weight를 학습시킨다. (auto-encoder/auto-decoder라고도 불림.)

- RBM 과정은 입력 layer에서부터 출력 layer까지 차례대로 2 layer씩 묶어서 'recreate input'이 되도록 관련 weight들을 학습시키고, 실제 데이터로 fine-tuning하는 과정으로 나눌 수 있다.

- RBM이 보여준 것은 초기 weight만 잘 줘도 쉽게 학습이 된다는 것이고, RBM까지는 복잡하지 않아도 간단한 방법으로도 초기값을 잘 줄 수 있다는 것을 xavier (2010) He (2015)가 밝혀 냈다.

출처: [http://eyeofneedle.tistory.com/category/모두를 위한 머신러닝과 딥러닝의 강의 정리](http://eyeofneedle.tistory.com/category/%EB%AA%A8%EB%91%90%EB%A5%BC%20%EC%9C%84%ED%95%9C%20%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D%EA%B3%BC%20%EB%94%A5%EB%9F%AC%EB%8B%9D%EC%9D%98%20%EA%B0%95%EC%9D%98%20%EC%A0%95%EB%A6%AC) [Technology worth spreading]

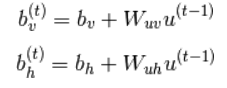
Modeling and generating sequences of polyphonic music with the RNN-RBM

\*RBM

RNN\_RBM을 이용한 연속적인 다음음악의 모델링과 생성

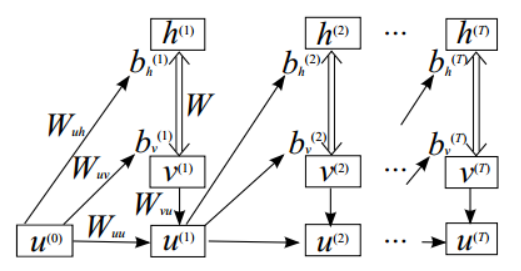
1. RNN\_RBM

RNN\_RBM은 시간단위 t에서의 특성벡터 v(t)가 고차원일 때의 시간적 시퀀스의 밀도 추정을 위한 에너지에 기반한 모델이다. 이것은 v(t)|A(t)의 다양한 조건적 분포를 묘사하도록 해준다. 이때, A(t) ={vT|T<t}는 파라메터 bv(t), bh(t)가 숨겨진 단위 u(t)에서의 결정적 RNN의 결과에 의존하는 (각각이 하나의 시간 단위인) 일련의 조건적 RBM들을 통해 시간 t에서의 시퀀스 히스토리를 나타낸다.

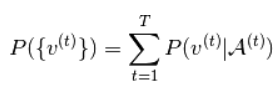
, 그리고 single-layer RNN recurrence relation은



결과적으로 모델을 시간에 따라 펼치면 다음과 같다.



전체적인 확률 분포는 T time step의 합에 의해 다음과 같이 주어진다.



여기에서 우변의 곱해지는 수는 t번째 RBM의 한계 확률이다.

구현의 명확성을 위해 [BoulangerLewandowski12]에 대비하여 우리는 웨이트 매트릭스에 대해 분명한 네이밍 규칙을 사용하는데, 우리는 recurrent hidden unit에 대해 h(t)대신 u(t)를 사용한다.

구현

우리는 두개의 Theano 함수를 만들고자 한다. 하나는 RNN-RBM을 트레인하기위해, 그리고 다른 하나는 그것으로부터 샘플 시퀀스를 생성하기 위해.

1. 트레이닝을 위해, 즉 {v(t)}가 주어졌을 때, RNN hidden state {u(t)}와 관련된 {bv(t), bh(t)} 파라메터들은 결정적이고 각 트레이닝 시퀀스에서 쉽게 계산될 수 있다. 그리고 나면 파라메터 들에 대한 SGD 업데이트는 각각의 트레이닝 예시가 보통의 RBM에서 mini-batch로 다루어지는 것과 같은 방식으로 각 시퀀스의 각각의 time step에 대해 contrastive divergence(CD)를 통해 추정될 수 있다.

Sequence generation은 v(t)가 recurrence와 시퀀스 히스토리로 넘겨지기 전에 분리된 (배치가 아닌)깁스체인을 가지고 각 타임 스텝에 연속적으로 추출될 수 있다.

RBM 레이어

RBM – weight initialization

고차원 시퀀스에서의 시간 의존성 모델링 – 다음음악 생성과 편곡에의 적용

Abstract

우리는 매우 일반적인 piano-roll 형식으로 다음음악의 symbolic sequence를 모델링하는 문제를 연구한다. 우리는 고차원 시퀀스에서 시간의존성을 발견할 수 있는 RNN을 가정하고 분포 추정에 기반한 확률 모델을 도입한다. 우리의 접근은 다양한 현실의 데이터셋에서 다음음악의 다양한 전통적 모델을 능가한다. 우리는 어떻게 우리의 음악 언어 모델이 다음음악 편곡의 정확성을 향상시키는데 상징적인 역할을 하는지 보인다.

1. Introduction

Modeling sequence는 많은 음악, 연설, 인간의 움직임과 같은 자연적으로 발생하는 현상들이 본질적으로 연속적이기 때문에 머신러닝에서 중요한 영역이다. 복잡한 시퀀스는 시간에 국한된 요소의 영향이 임의의 긴 시차에 의해 느려질 수 있다는 점에서 non-local이다. 예를 들어, 초반의 곡조는 음악 패턴이나 주제는 끝까지 반복된다. RNN은 원칙적으로 전체의 sequence history를 요약할 수 있는 내부 메모리를 포함한다. 이 특성은 장기간의 의존성을 표현하는데 적합하다. 그럼에도 gradient-based optimization을 가지고 효율적으로 트레인하는 것은 도전이다. 최근에는 HF최적화를 통해 RNN을 트레이닝시키는 것이 이러한 어려움을 줄이는 것을 도와줄 수 있다. 많은 관련된 시퀀스는 비디오에서의 이미지, 오디오 음악에서의 단기간의 스펙트럼, 악보의 음의 튜플이나 텍스트의 단어 같이 고차원 이상이다. 이러한 경우에서 단순히 이전 시간 단위에서 관측된 값에 기반하여 다음 시간 단위의 값을 예측하는 것은 만족스럽지 않다. 그러한 각 시간 단계의 고차원적인 object를 가지고 조건적 분포는 매우 자주 다양하고, 우리는 그러한 시퀀스에 대한 우리의 모델을 이전의 시간 단계를 가지고 다음 시간 단위의 조건적 분포를 예측하는 것 보다 매우 선호한다. 다음음악의 경우에, 특정한 음이 특정 시간에 나타나는 것은 다른 음들이 동시에 나타나는 확률을 크게 수정한다. 다시말해서, 음들은 다층 분류 작업을 위해 고안된 전형적인 RNN 구조에 의해 알맞게 묘사될 수 없는 상호 연관된 패턴으로나 동시에 나타난다. 예를 들면 예측할 변수의 모든 설정을 enumerate하는 것은 매우 고비용이기 떄문이다. 이러한 어려움은 RBM이 악명높아진 중에 임의의 에너지 함수에 의해 주어진 설정의 negative log-likelihood를 표현할수 있는 energy based model의 동기가 되었다.

이와 관련하여, 우리는 각 시간 단계의 복잡한 분포를 표현하기 위하여 이전의 파라메터에 의존하는 파라메터들을 가지고 RBM의 능력, 시간적 RBM을 가지고 제시된, 을 활용하려고 한다. 이것은 heuristic 과정을 통해 훈련된다. RNN과 RBM의 바람직한 특성을 조합하는 것은 사소하지 않다는 것이 알려져있다. 반복시간RBM은 contrastive divergence에 의해 정확한 추론과 효율적인 훈련을 허락한다. 그것은 간단하지만 성공적으로 많은 흥미로운 시퀀스를 설명한다. echo state network에 기반한 비슷한 아키텍쳐도 최근에 개발되었다. 이 연구에서 우리는 RTRBM이 다음음악의 많은 전통적인 모델을 능가한다는 것을 보였고, 우리는 RTRBM을 일반화한 RNN-RBM을 소개했다. 그것은 시간적 의존성이 연관된 것들을 묘사하는 데 더 용이하다.

더 정확하게, 우리는 symbolic music, 즉 더 복잡하고 음향적으로 풍부한 음향 신호가 아닌 전형적으로 한 악보나 한 MIDI 파일에 포함된 명시된 시간, 음조, 속도, 악기의 정보로 표현되는 것들의 시퀀스를 고려할 것이다. 음악 모델은 대부분 서양 음악의 기본적요소와 조화, 리듬에 집중하고, 이전 시간 구간을 고려하여 그 다음에 함꼐 연주될 음의 패턴(동시)를 예측하도록 트레인된다. 두가지 요소가 모델의 양적 퍼포먼스를 대표한다 : 시간적 의존성과 화음의 조건적 분포. 대부분의 흥미로운 모델들은 미리 정의된 화음이나 다른 차원이 줄어든 표현과 함께 하나의 음이 그 결과인 반면, 우리는 제한되지 않은 piano-roll 모델을 목표로한다. 즉 각 시간 단계에 어떤 음이 나타나는지 정확히 표시한 이진 행렬로서. 동적성과 다른 악보의 주석들을 무시함에도 불구하고, 이 작업은 머신 러닝 알고리즘을 향상시킬 잘 정의된 프레임워크를 대표하고, 다음음악의 편곡에 직접적으로 적용될 수 있다.

다음음악편곡의 목표는 기저한 다음음악의 음향 신호의 음들을 그 악보에 접근하지 않고 결정하는 것이다. 인간 전문가는 이 어려운 문제를 현재 시그널에 무엇이 존재하는가보다는 그들이 듣기를 기대하는 것에 더 중요도를 두어 적븐한다. 대부분의 존재하는 편곡 알고리즘은 프레임에 기반하고 대부분 음향 신호에만 의존한다. (어떠한 접근들은 근본적인 음악적 제약을 이용하지만). 자연어 모델이 음성 인식 시스템의 성능을 획기적으로 개선시켰던 것과 같은 방식으로 음악적 언어 모델이 음악적 정보 재현에 순수한 음향적 접근을 향상시켰다는 것은 알려져있따. 그러나 이러한 두 정보를 합하는 것은 사소한 일이 아니다. 그 결과 HMM을 가지고 시간적 smooting은 종종 최신 편곡에 post-processing이 포함된다. 우리는 어떻게 symbolic 지퀀스에 훈련된 전문가의 조언을 포함하는 임의의 편곡 알고리즘을(기본적 가정하에서) 풍부하게 하는지를 보여줄 것이다. 우리의 이러한 혼합적 접근을 통해서 우리는 편곡 정확성을 높일 수 있다(유명한 HMM접근보다 훨씬 더)

p – 특정 time step에서 note가 분포할 확률

1. RBM

v: input vector(visible) h: hidden vector

RBM – 주어진 v에 대해 hi추론 / 주어진 h에 대해 vj 샘플링

contrastive divergence?(Hinton, 2002)

RBM > RTRBM(Contrastive divergence) > RNN-RBM