말인식, 손글씨 인식, polyphonic music modeling.

Lstm 8가지 변종으로

fANOVA 평가지표 사용

1. Vanilla LSTM

gers et al에 의해 변경됨

peepholes 연결부 존재

gate activation -> sigmoid

input & output activation -> tanh

초기버전 lstm은 peepholes, forget gate 존재 x

Forget gate 도입으로 reber문법? 과 같은 지속적인 과제의 배움 허락?

정확한 타이밍을 배우기 쉽게 하기 위해 peephole connection 추가

Gru ->peephole연결 & 출력 활성화함수-> 입력 & forget 게이트 -> update게이드에 결합

검증된 Timit 말뭉치 데이터셋 사용

Iam online handwriting database 사용(train1, vali 2, test 1)

Piano role

2개의 히든레이어로 하나 입력, 하나 이전 시간,

Softmax 단일 출력 레이어 이동

timit, iam online tasks 사용

cross entropy error 사용, 정규분포, stochastic gradien descent인 nesterov momentum 사용

early stopping 150

파생된 7가지 변이 lstm들

NIG: No Input Gate:

NFG: No Forget Gate

NOG: No Output

NIAF: No Input Activation Function

NOAF: No Output Activation Function

CIFG: Coupled Input and Forget Gate

NP: No Peepholes

FGR: Full Gate Recurrence:

Jsb -> one layer lstm + sigmoid output layer

Timit & iam -> bidirectional lstm, 각 layer one layer of softmax output layer에 connection

Jsb, timit -> cross entropy loss

Iam -> connectionist temporal classification

Timit : classification error

Iam : character error rate

Jsb : negative log-likelihood

Grid search 말고 random search 사용

9가지 변형과 3가지 데이터셋 -> 27개의 무작위 검색 수행

5400번 시험, 각 200번

• number of LSTM blocks per hidden layer: log-uniform samples from [20, 200];

learning rate: log-uniform samples from [10−6 , 10−2 ];

• momentum: 1 − log-uniform samples from [0.01, 1.0];

• standard deviation of Gaussian input noise: uniform samples from [0, 1].

티밋 데이터셋의 경우 2가지 bool hyperparameter 고려됨

1. 일반 모멘텀 vs 네스테로프 모멘텀
2. Gradient clipping 할지 말지

티밋의 경우 cifg가 29.6% 제일 좋음

Iam -> NP가 제일 좋음

Figure3 파란색 박스들 vanilla lstm과 많이 다른것들

출력 활성화 기능(NOAF) 또는 망각 제거  
게이트(NFG)는 세 데이터셋 모두에서 성능을 크게 떨어뜨립니다.

CEC와는 별도로 오래된 정보를 잊어버리는 능력  
그리고 세포 상태의 찌그러짐은 매우 중요한 것으로 보인다

실제로 출력 활성화 기능이 없으면 블록 출력은 원칙적으로 무한대로 증가할 수 있습니다.  
입력과 망각 게이트를 결합하면 이 문제를 피할 수 있으며 출력 비선형성의 사용이 덜 중요해질 수 있으므로 GRU 없이도 성능이 좋은 이유를 설명할 수 있다.

Cifg 유의하지 않음, 음악 모델링 성능 약간 up

Np 는 peephole 연결부 제거하지도 않았는데 최고성능 가짐

Fgr 유의하지 않음 timit, iam 성능 up, jsb는 성능 down 저자는 사용안하는거 추천

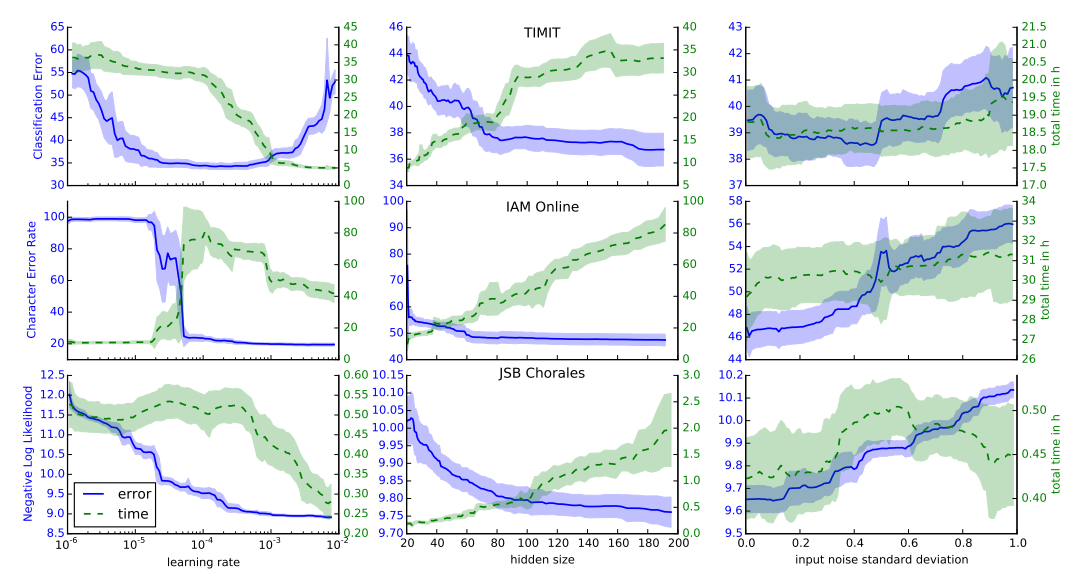
Nig, nog, niaf는 음성 및 필기인식 성능 떨어트림 하지만 음악 모델링에는 영향 x

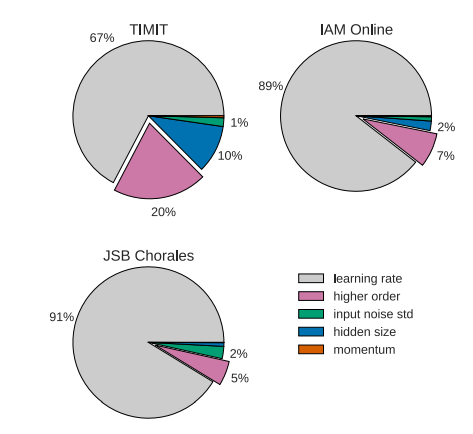
Nig, niaf 음악 모델링 성능 약간 up

연속 실측값 데이터에 대한 지도학습의 경우 입출력 게이트, 입력 게이트의 activation function 중요

Fanova 는 hyper parameter 평가 지표(regression tree base?)

1. Learning rate
2. Hidden layer size
3. Input noise 0.2~0.5
4. Momentum -> 중요하지 않다?





Timit 학습 속도에 따라 크게 달라진다.

망각 게이트(CIFG) 또는 핍홀 연결부 제거(NP)  
성능이 크게 저하되지 않으면서 LSTM을 단순화했습니다.

Forget 게이트와 출력 활성화 기능은 LSTM 블록의 가장 중요한 구성 요소

이들 중 하나를 제거하면 성능이 크게 저하됩니다. 우리는 무한 셀 상태가 네트워크를 통해 전파되는 것을 방지하고 학습을 불안정하게 하기 위해 출력 활성화 함수가 필요하다고 가정한다.

1. Hyper parameter
2. Network size
3. Momentum 쓸모없었다.
4. 가우스 노이즈는 timit에 도움됐지만, 다른건 도움 x
5. 실용적인 hyper parameter들은 상호 독립적이다.
6. 학습속도 작은 사이즈 네트워크로 먼저 조정해라