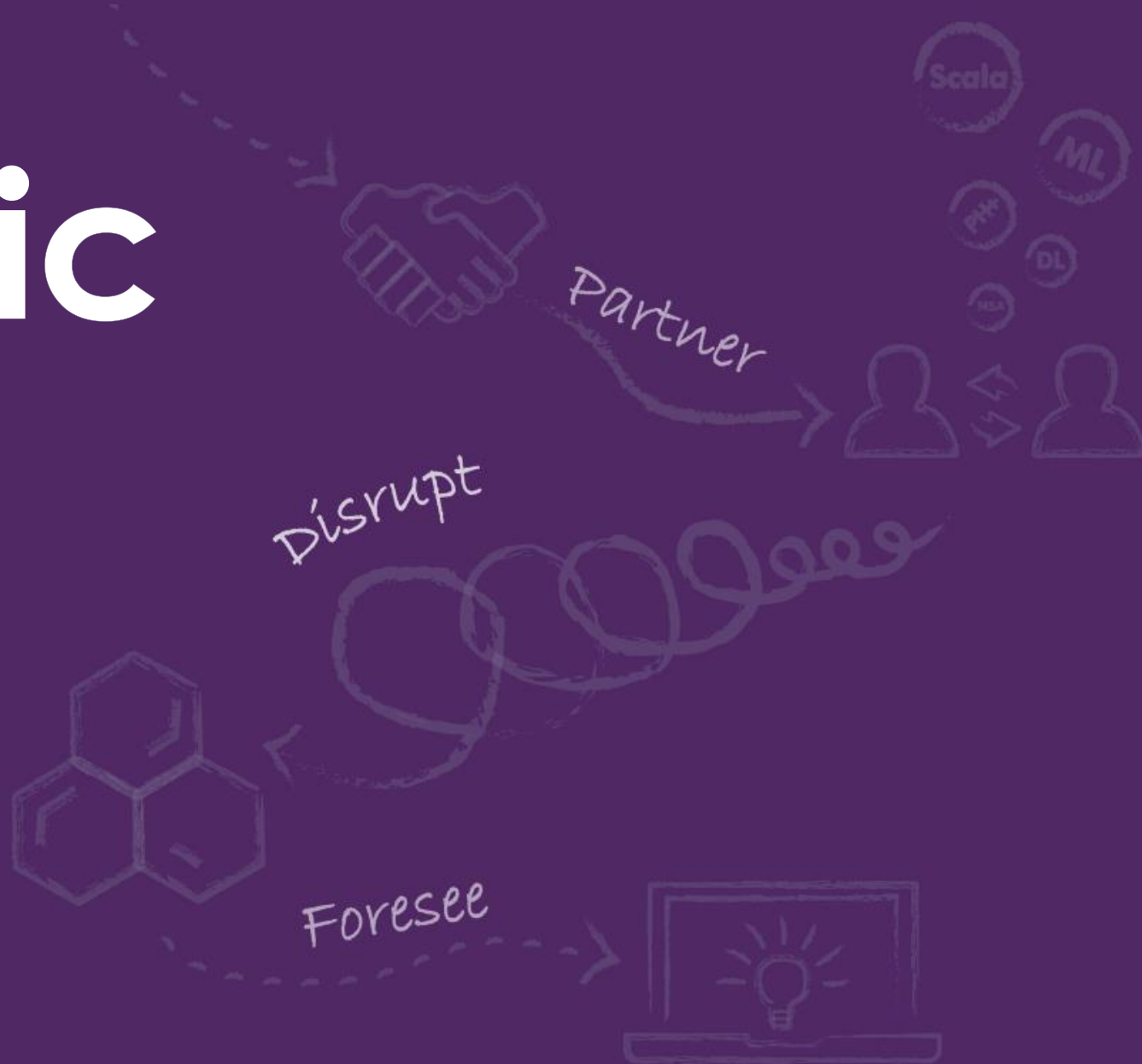


Techtonic 2018

-
Thu . Nov 15

-
SAMSUNG SDS Tower
West Campus B1F
Magellan Hall /Pascal Hall



좌충우돌!

딥러닝으로 부정맥찾기

삼성SDS 윤용근 프로



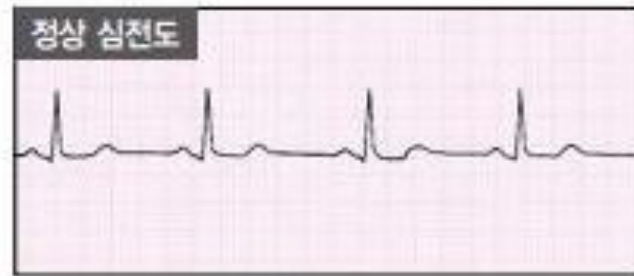
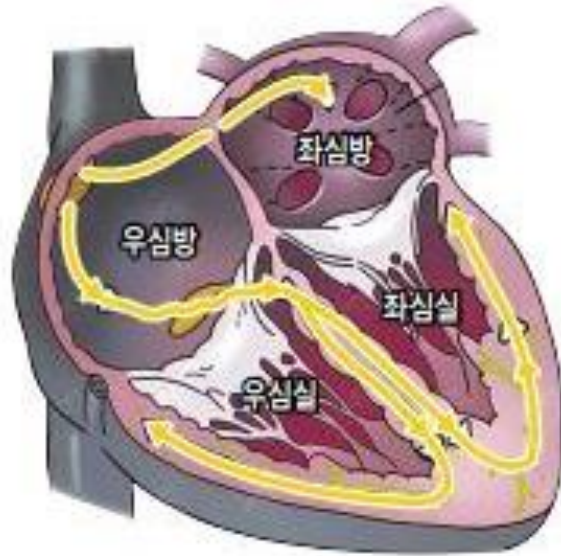
- 부정맥이란?
- 딥러닝 적용 과정
- 딥러닝 적용 결과

딥러닝으로 부정맥 찾기

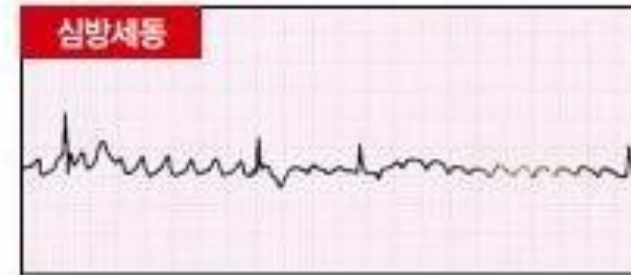
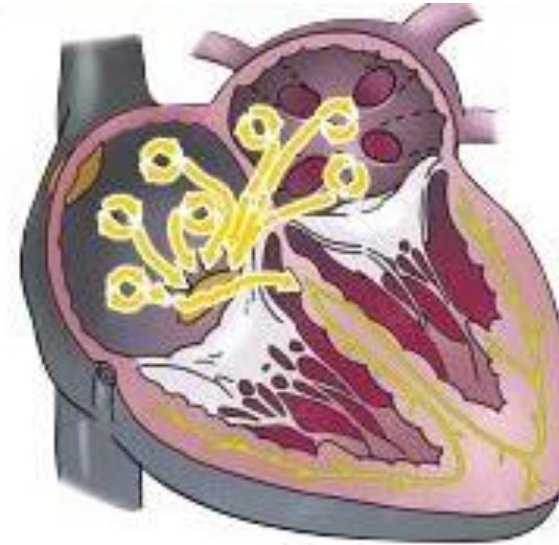
부정맥이란?

부정맥

심장의 수축과 팽창을 위해 필요한 전기적 신호가 비정상적일 때 나타나는 질환



정상



부정맥

부정맥의 위험성

KBS NEWS

2018. 11. 4. (일) 08:00~09:00 / 16:00~17:00

진행: 박광식 KBS 의학전문기자

출연: 정동섭 삼성서울병원 심장외과 교수

[박광식의 건강365] 심장 회로 합선 ‘부정맥’...늦게 뛰면 심정지, 빨리 뛰면 혈전 위험 ↑

▷박광식:

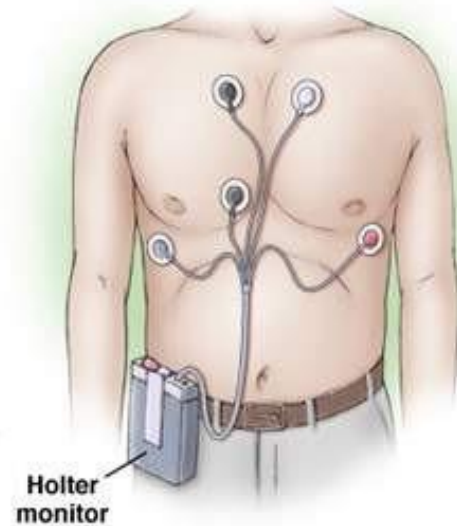
부정맥 이 부분이 가장 궁금할 것 같아요. 이게 위험한 질환입니까?

▶정동섭:

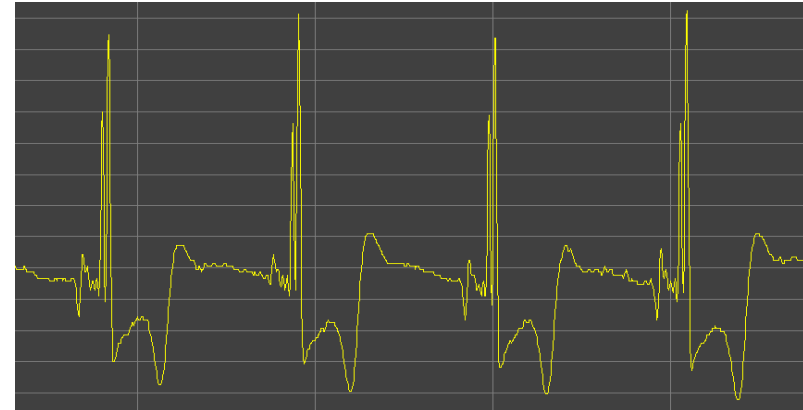
네, 그렇습니다. 부정맥은 심실부정맥인 경우에는 그 자체로 생명이 위험할 수 있습니다. 심방부정맥, 심방세동이나 심방조동의 경우에는 그 자체로 생명에 지장이 없을 수는 있지만, 중풍 뇌졸중이나 실신 그다음에 심장기능이 떨어지는 심부전증 등의 합병증을 유발하게 되어 삶의 질이 떨어지거나 호흡부전 또는 마비로 삶을 살아가게 될 수도 있습니다. 심방세동의 경우에 뇌졸중 위험은 고령인 경우 정상인의 7배에서 10배까지도 보고되고 있습니다.

부정맥 확인방법 - 심전도

정해진 시간 동안에 심장의 전기적 활동을 기록하는 것



© 2004 NorthPoint Domain



24시간 데이터 기준 , 8만~11만(60bpm~80bpm)

부정맥의 진단



판독 전문가가 판독하고 심장전문의를 진단
* 24시간 데이터기준 판독에 2~3시간 소요

부정맥 분석의 자동화 필요성

심전도가 Normal인 상태가 약 97.5%

2.5%의 이상을 찾기 위해
8만 beat이상을 분석

해당 구간을 빠르게 찾아주면
판독전문가의 업무에 도움이 될 수 있음



딥러닝이 필요한 이유

룰 베이스

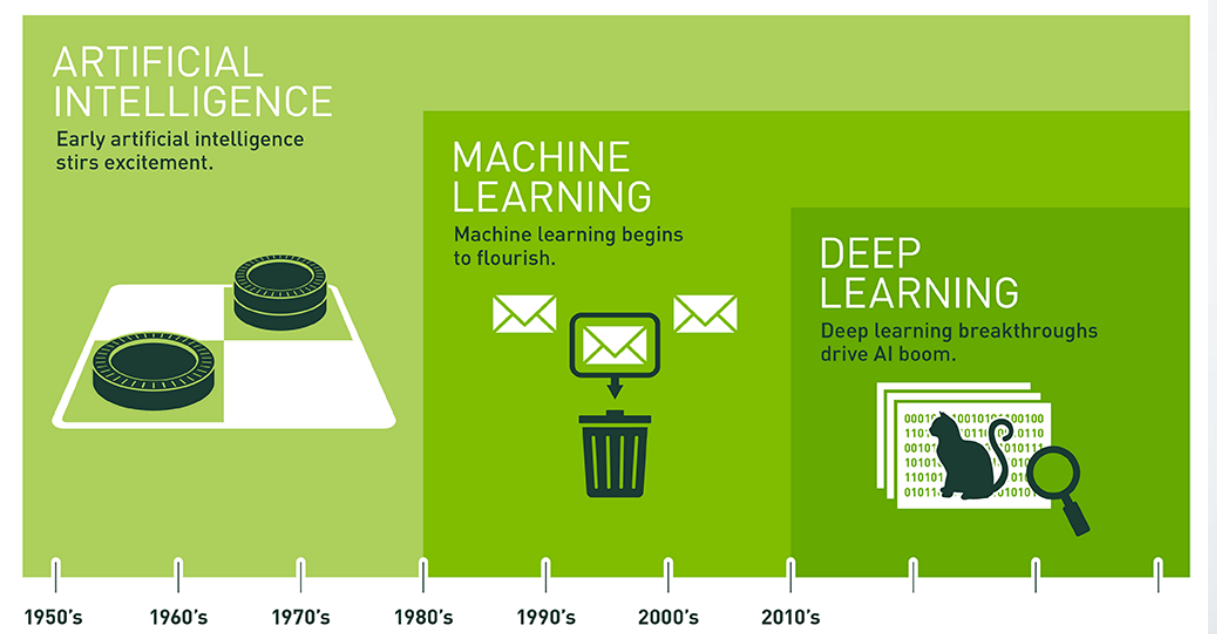
전문가의 지식수준에 따라 달라짐
한번 만들어진 룰은 변경하기 힘들

머신 러닝

Feature Extraction의 수작업이 필요함

딥러닝

Feature Extraction의 자동화
비 정형 데이터 대응 가능
새로운 데이터 발생시 모델 추가 학습 가능



딥러닝으로 부정맥 찾기

딥러닝 적용 과정

목표

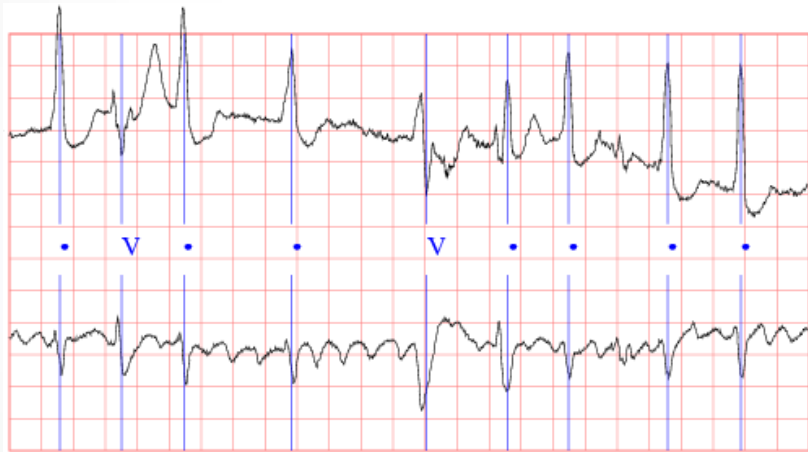
딥러닝을 이용하여 부정맥을 판독하자

제한 조건 1 : ONLY CPU

제한 조건 2 : 판독전문가들의 실제 현업 소요시간보다 월등히 빠를 것

제한 조건 3 : 판독전문가들이 인정할 만한 정확도를 달성할 것

학습 데이터 모으기



← → ↻ <https://www.physionet.org/lightwave/>

Choose input View/edit Tables Settings Help

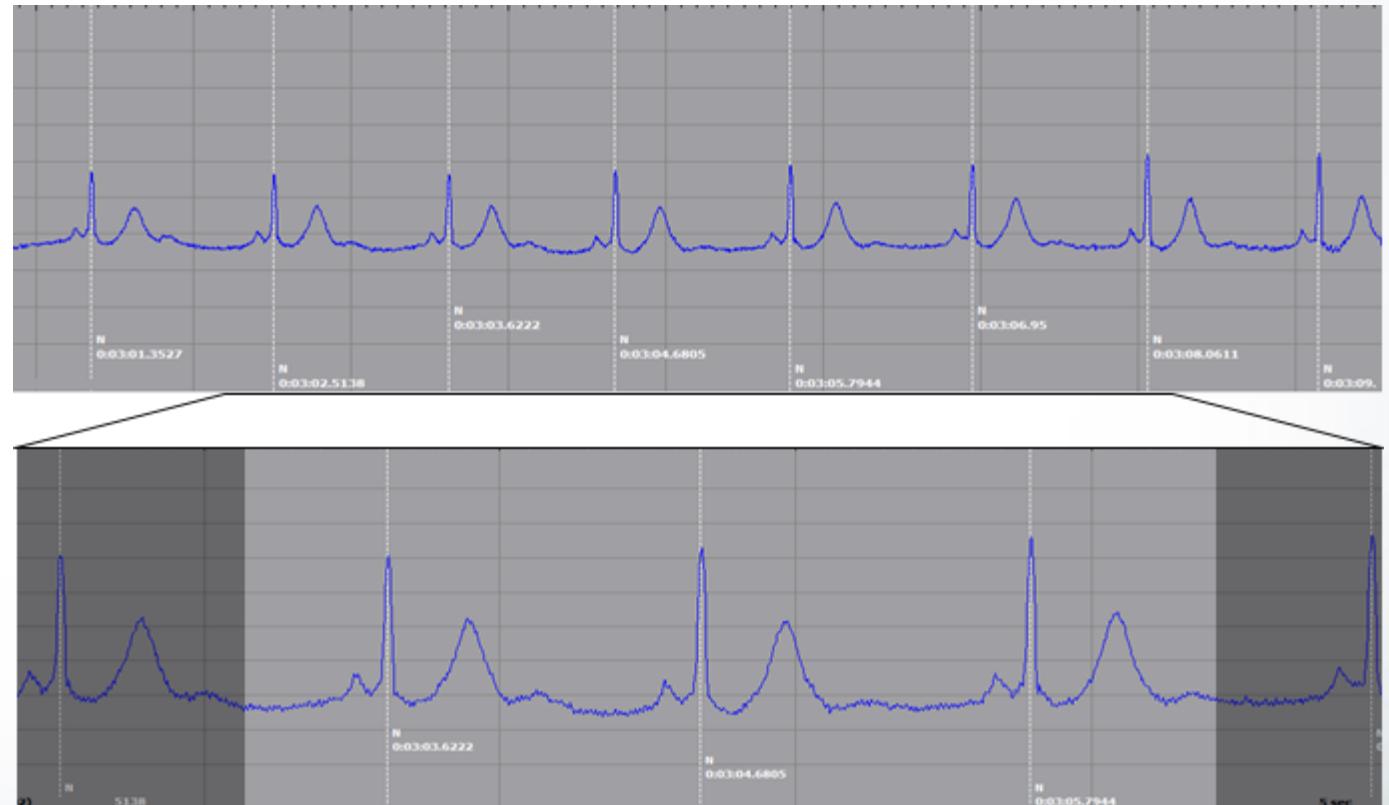
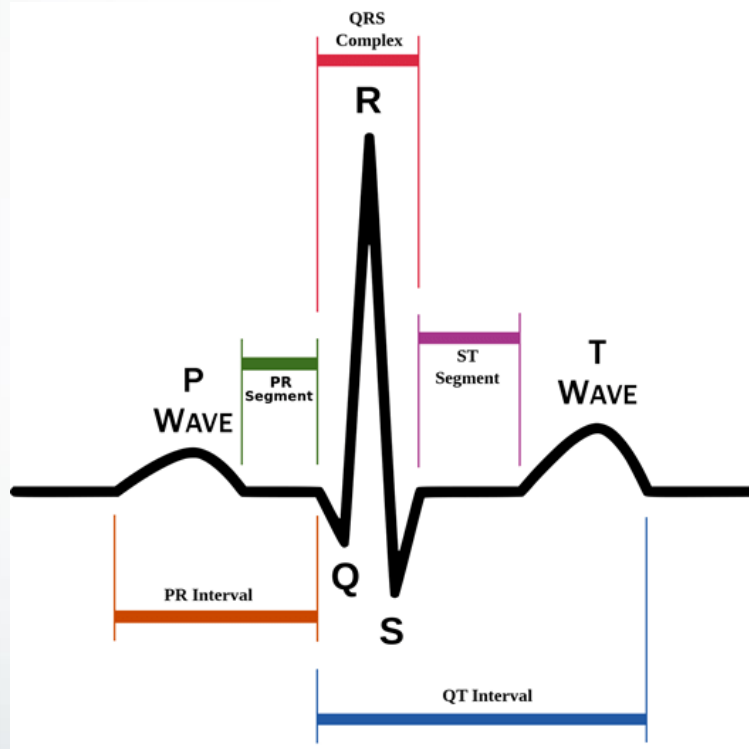
Database: --Choose one--

- Choose one--
- ANSI/AAMI EC13 Test Waveforms (aami-ec13)
- Abdominal and Direct Fetal ECG Database (adfecgdb)
- MIT-BIH Atrial Fibrillation Database (afdb)
- PAF Prediction Challenge Database (afpdb)
- AF Termination Challenge Database (aftdb)
- AHA Database [sample excluded record] (ahadb)
- Apnea-ECG Database (apnea-ecg)
- Blood Pressure in Salt-Sensitive Dahl Rats (bpsrat)
- CAP Sleep Database (capslpdb)
- MIT-BIH ECG Compression Test Database (cdb)
- Combined measurement of ECG, Breathing and Seismocardiograms (cebsdb)
- Challenge 2009 Test Set A (.../test-set-a)
- Challenge 2009 Test Set B (.../test-set-b)
- Challenge 2010 Training Set A (.../set-a)
- Challenge 2010 Test Set B (.../set-b)
- Challenge 2010 Test Set C (.../set-c)
- Challenge 2011 Training Set A (.../set-a)
- Challenge 2011 Test Set B (.../set-b)
- Challenge 2011 Pilot Set (.../sim)

physionet에서 MIT-BIH, AHA, CU 등 open data를 구할 수 있음

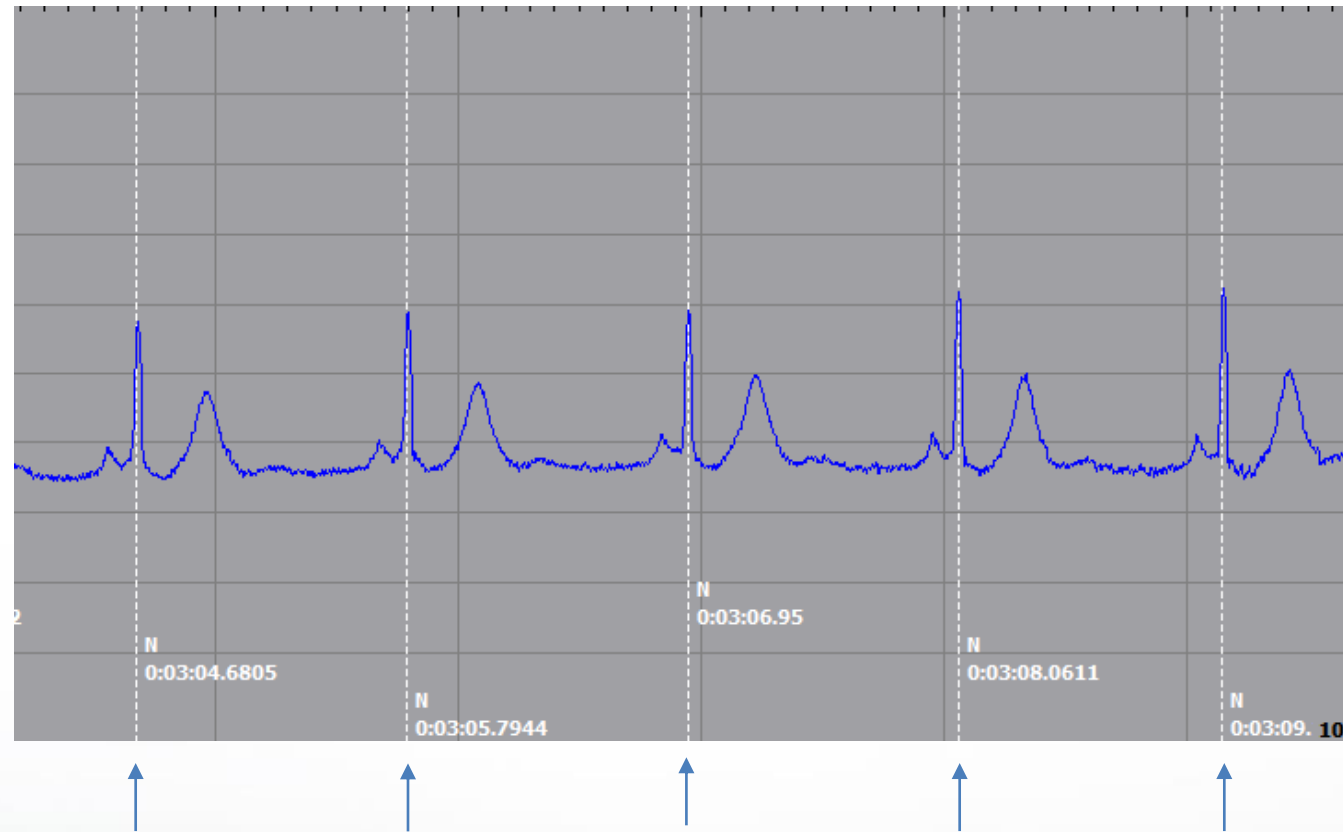
전처리 – 입력 데이터 Segmentation

연속된 심전도를 처리하기 위한 입력 단위 설정 필요



전처리 - Labeling

데이터 학습을 위해 Labeling 진행



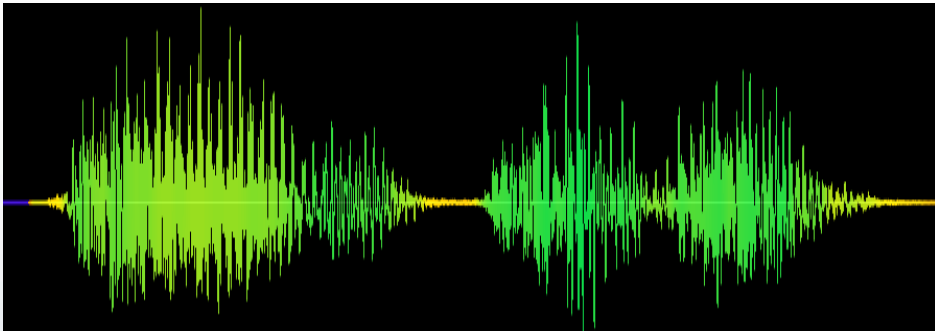
학습 모델 구상하기

데이터가 끊임 없이 들어온다

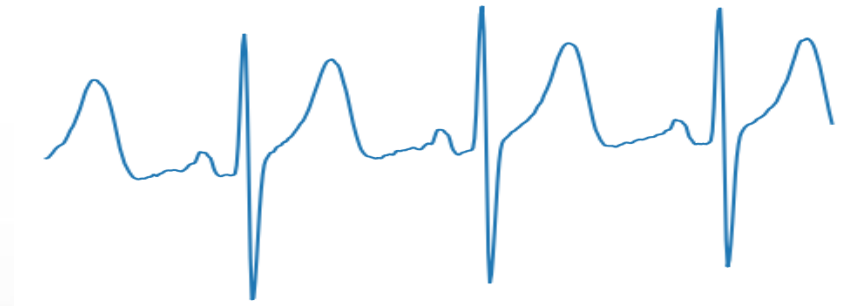
단위데이터의 길이가 일정하지 않다

데이터가 파형을 가지고 있다

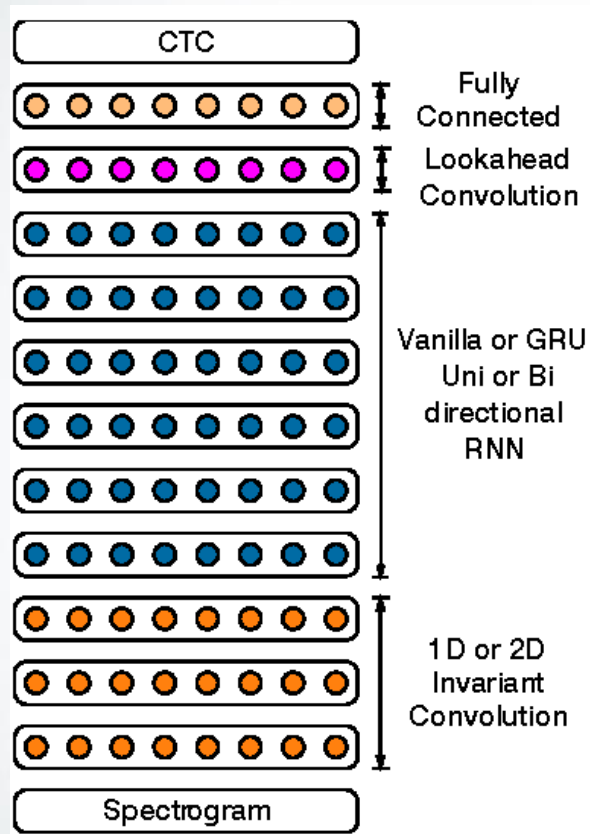
음성데이터랑 비슷하다



\approx



음성인식 모델



Deep Speech 2

[samsungsds-rnd / deepspeech.mxnet](#) Watch 4

Code Issues 2 Pull requests 0 Projects 0 Wiki Insights Settings

A MXNet implementation of Baidu's DeepSpeech architecture

mxnet warp-ctc speech baidu deepspeech arch stt speech-recognition speech-to-text Manage topics

78 commits 2 branches 0 releases 2 contributors

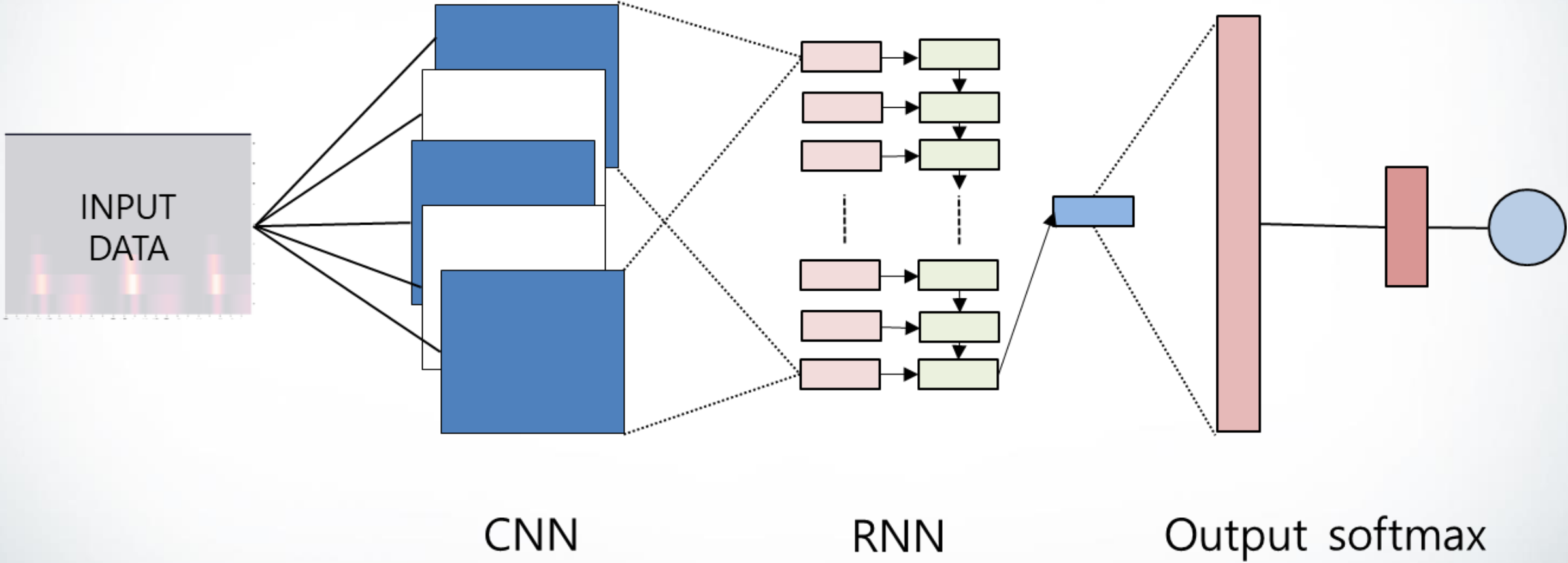
Branch: master New pull request Create new file Upload files

Soonhwan-Kwon fix duplicated drop out layer

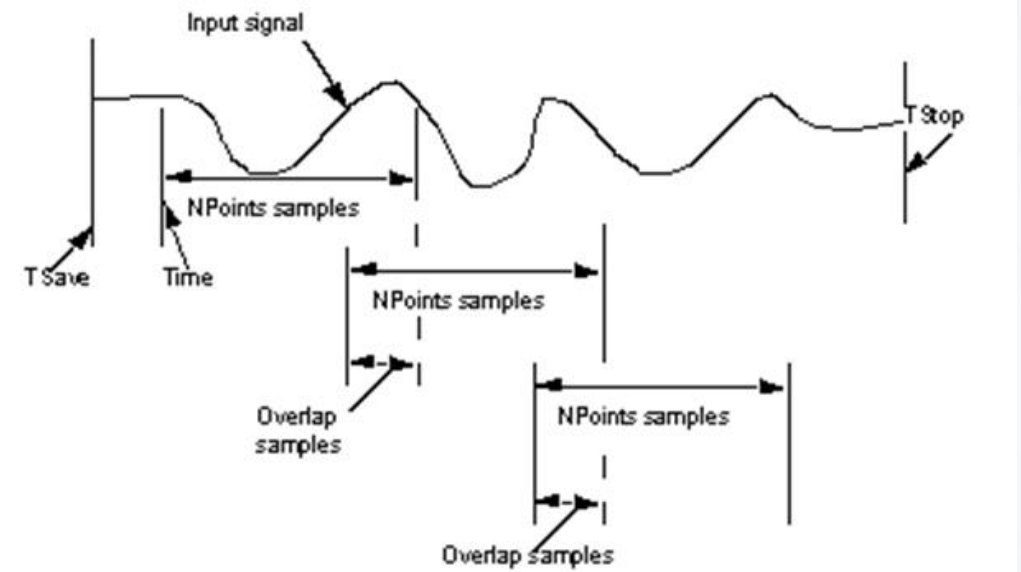
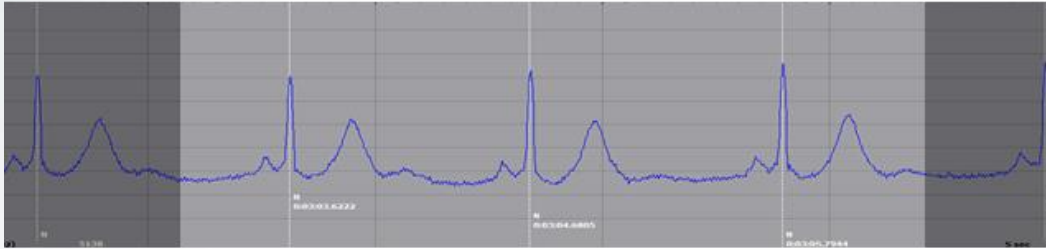
- Libri_sample first commit
- layer fix duplicated drop out layer

[incubator-mxnet / example / speech_recognition /](#)

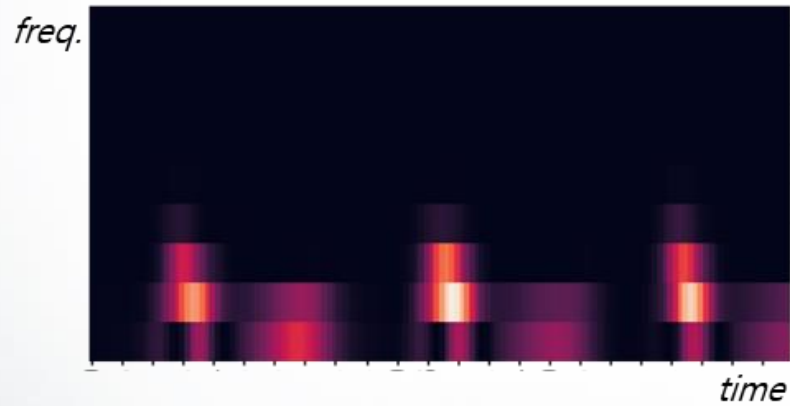
딥러닝 학습 모델 설계



CNN 적용 – Feature Extraction (1)

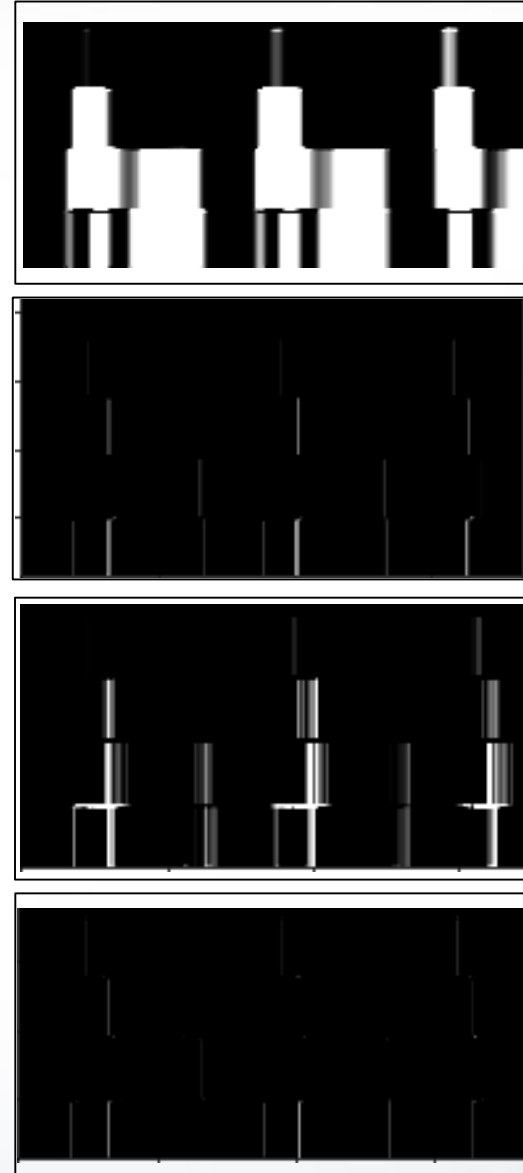
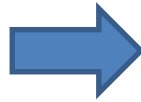
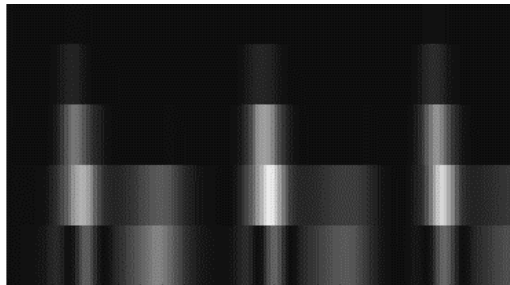


There are 3 segments here, i.e., $NS = 3$, to be used in averaging the spectrum.

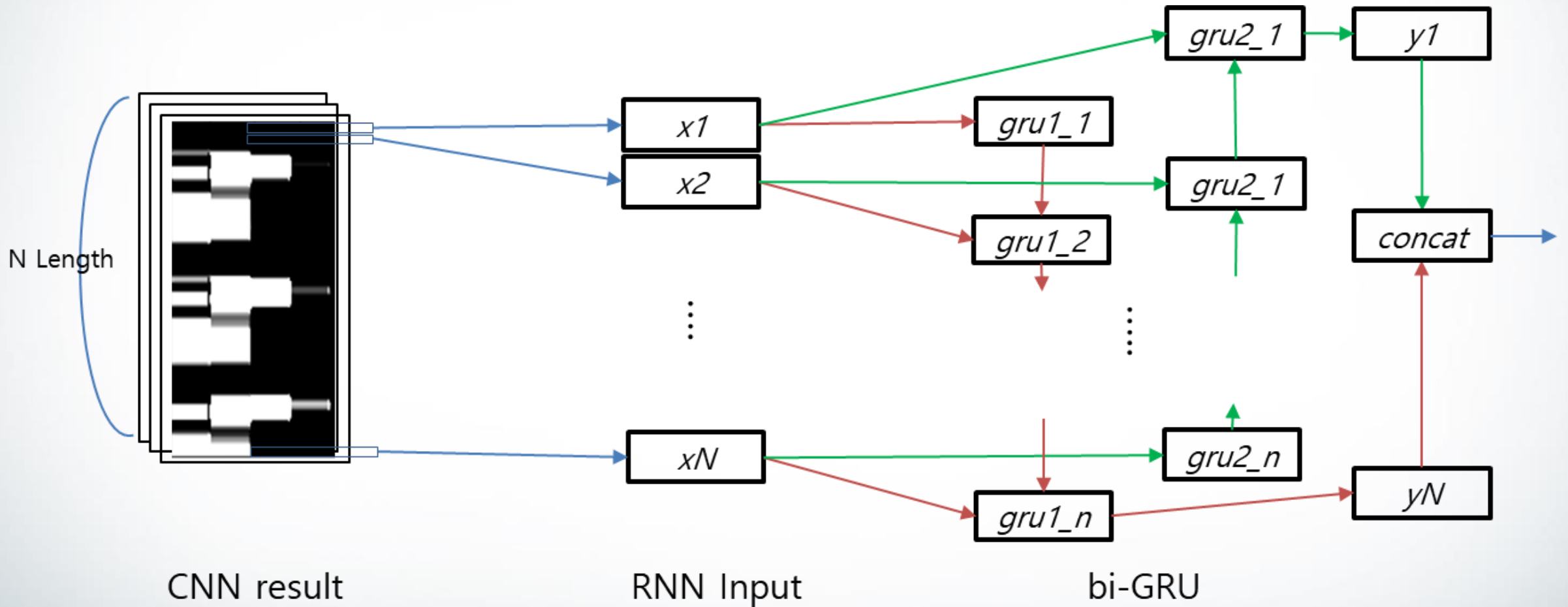


<http://edadocs.software.keysight.com/display/ads2009U1/FFTAnalyzer>

CNN 적용 – Feature Extraction (2)



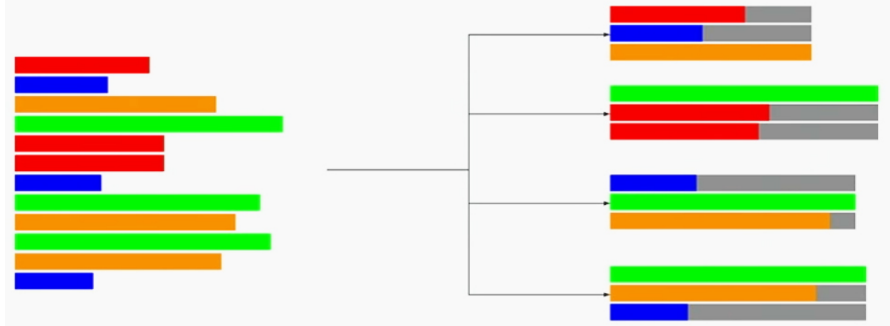
RNN 적용



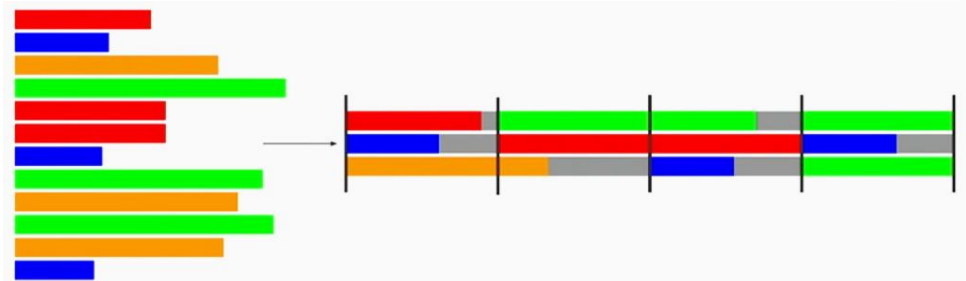
RNN 입력데이터 최적화

학습데이터를 특정 길이에 맞게 분배

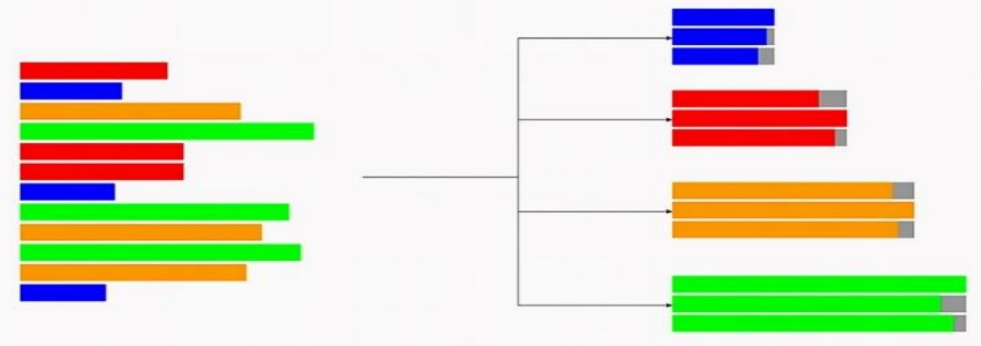
Dynamic Padding



State Saver



Bucketing



딥러닝으로 부정맥 찾기

딥러닝 적용 결과

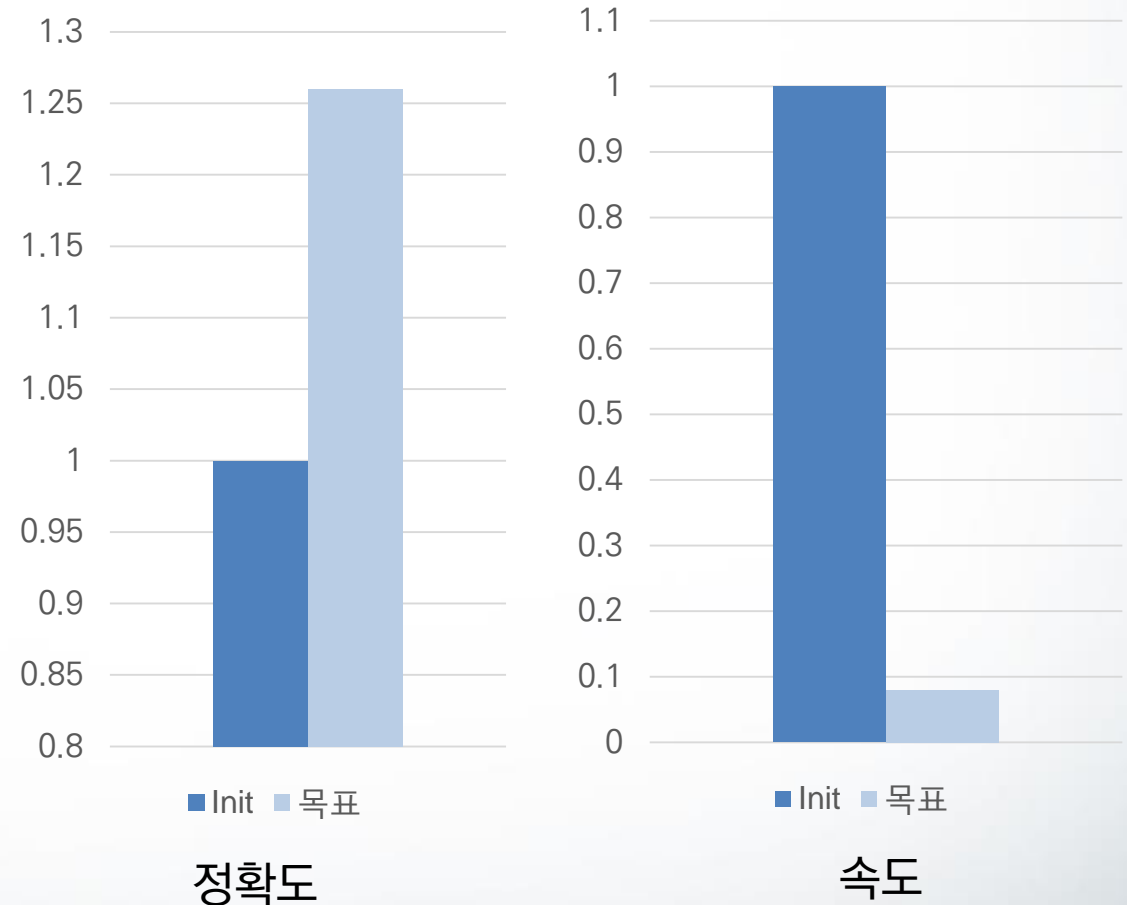
Initial State

대상장비 : 8 core CPU

대상데이터 : 24시간 ECG data

목표정확도 : Init 대비 26%이상 향상필요

목표속도 : Init 대비 8%수준 필요



개선 방안

학습 방법 개선

학습 데이터의 길이 분포를 고려하여 분류한 후 학습

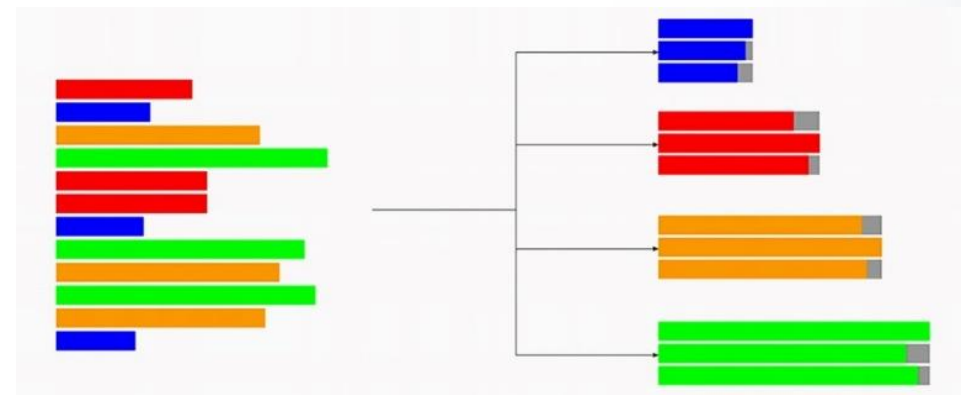
학습 데이터 정제

학습 데이터 중 정답 오류, 레이블링 오류 등 정제작업

판독 속도 개선

데이터 압축

Matrix 연산량 감소 (via Intel MKL)



INTEL® MATH KERNEL LIBRARY

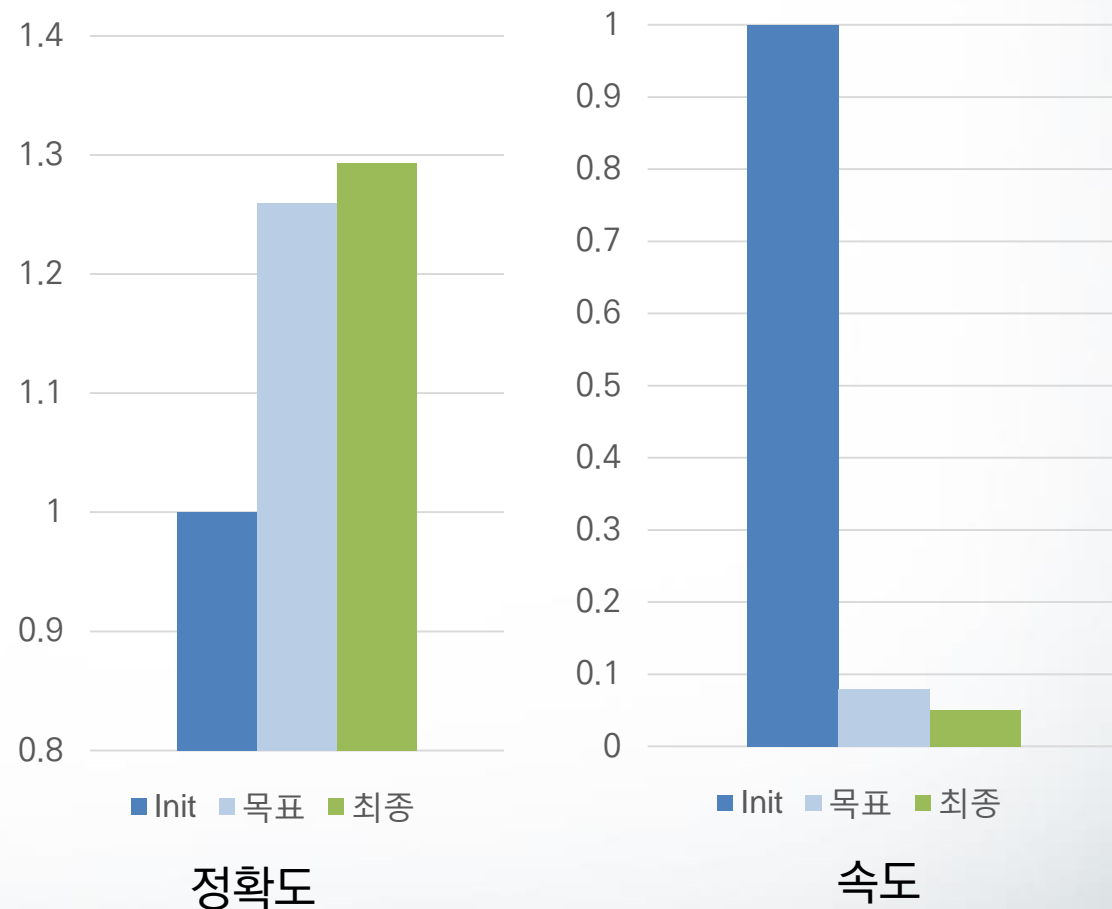
최종 결과

대상장비 : 8 core CPU

대상데이터 : 24시간 ECG data

정확도 : Init 대비 29.3% 상승

속도 : Init 대비 5% 수준



사실 오늘은 잘 된 것만 말씀 드렸어요

카테고리 별 삽질 리스트

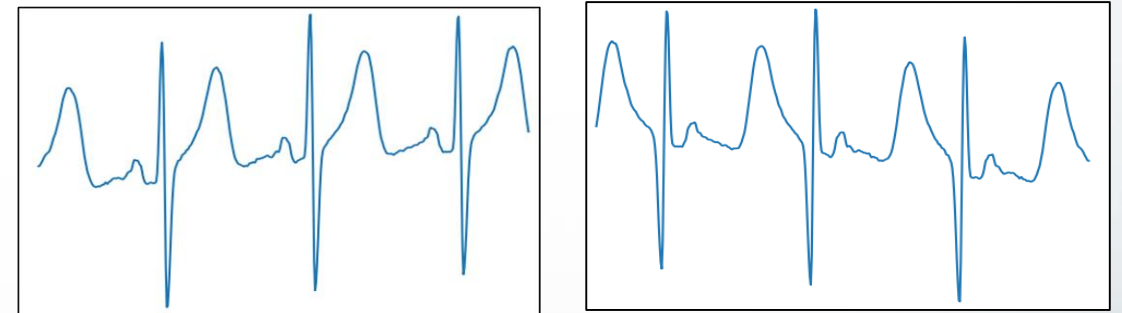
전처리

각 시그널 값을 one-hot encoding 하기
Inverse Data 로 학습하기

color	color_red	color_blue	color_green
red	1	0	0
green	0	0	1
blue	0	1	0
red	1	0	0

Hyper Parameter

CNN, RNN Hidden Unit 수 바꾸기
CNN, RNN Layer 수 바꾸기
Output Class 개수 바꾸기



그 외 나열할 수 없는 수많은 삽질들

Q & A

Partner

Disrupt

Foresee



Thank you

