



Cloud Log Data Analysis

Shweta Oak

Mentor: Ricardo Rocha

11/08/2017

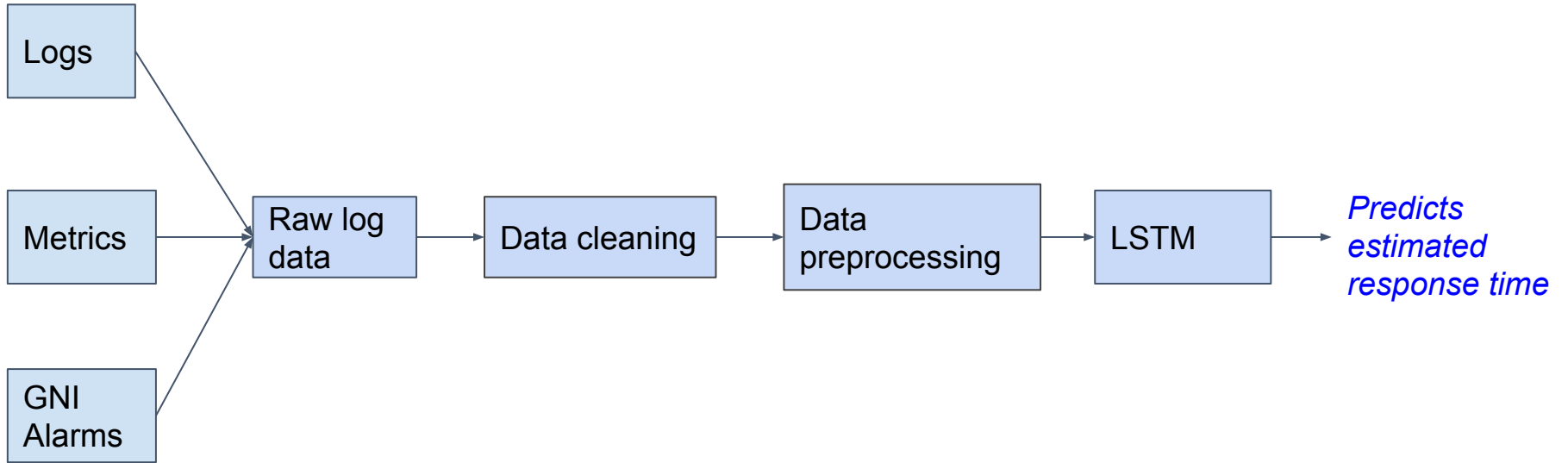
Why

- Difficult to analyse Openstack logs
- Algorithm for analysis
- Better maintenance of services

Methods that can be used

- Linear Regression
- Sliding windows
- Classification

- **Long short term Memory (LSTMs)**
- Non linear regression: complex



Data Collection

3 different data sources are used:

- Logs
- Metrics
- GNI Alarms



Data cleaning

Consider nova logs:

4 parameters are taken:

- host
- request_lb_ip (host group)
- client_ip
- request path

response time(to be predicted)

Data pre processing

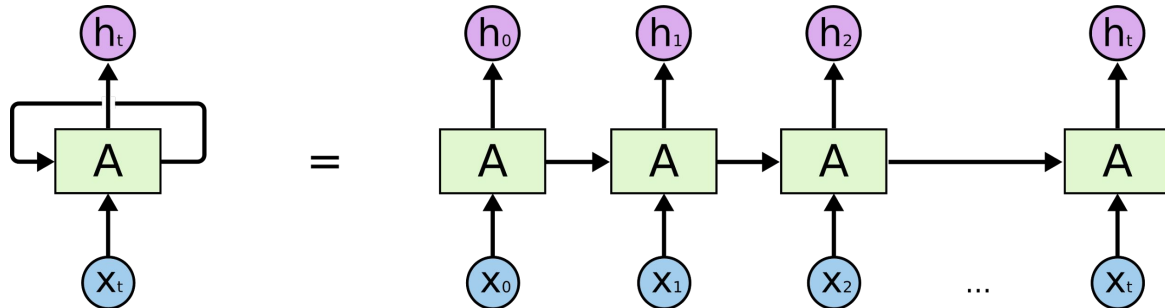
- Pick out unique values
- Convert categorical variables - one hot encoding

LSTM

Long short term memory cells:

“Remembers previous data and on current input data, makes a connection”

- Useful for time series data prediction, to find trends
- Aim is to predict response time



Future scope

- Optimize the model
- Documentation
- Deploy a Tensorflow cluster for model training and serving
- Integrate other OpenStack services and metrics

Thank You!

oakshweta11@gmail.com