ML - Applications and Dataset

Contributors :
Ritesh Kumar and Vipul Bajaj

Kritsnam Tech. Pvt. Ltd.

Topic:

Weather Forecasting

ML in Weather Forecasting-Stanford study

- While much of current forecasting technology involves simulations based on physics and differential equations, many new approaches from *artificial intelligence use mainly machine learning techniques*, mostly *neural networks* while some draw on *probabilistic models such as Bayesian networks*.
- *Neural networks seem to be the popular machine learning model* choice for weather forecasting because of the *ability to capture the nonlinear dependencies of past weather trends and future weather conditions,* unlike the linear regression models that they used. This provides the advantage of not assuming simple linear dependencies of all features over our models.
- Of the two neural network approaches, *one used a hybrid model that used neural networks to model the physics behind weather forecasting* while *the other applied learning more directly to predicting weather conditions.*
- Similarly, the approach *using support vector machines(SVM) also applied the classifier directly for weather prediction* but *was more limited in scope* than the neural network approaches.

ML in Weather Forecasting - Data Mining

Other method for weather forecasting include:

- *KMeans :* The K-Means clustering algorithm is a *partition-based cluster analysis method.* According to the algorithm we firstly select k objects as initial cluster centers, then calculate the distance between each object and each cluster center and assign it to the nearest cluster, update the averages of all clusters, repeat this process until the criterion function converged. Square error criterion for clustering.
- *HMM* : An HMM is a *double implanted stochastic process with two hierarchy levels.* It can be used to model much more complex stochastic processes as compared to a traditional Markov model. *In a specific state, an observation can be generated according to an associated probability distribution.* It is only the observation and not the state that is visible to an external observer.

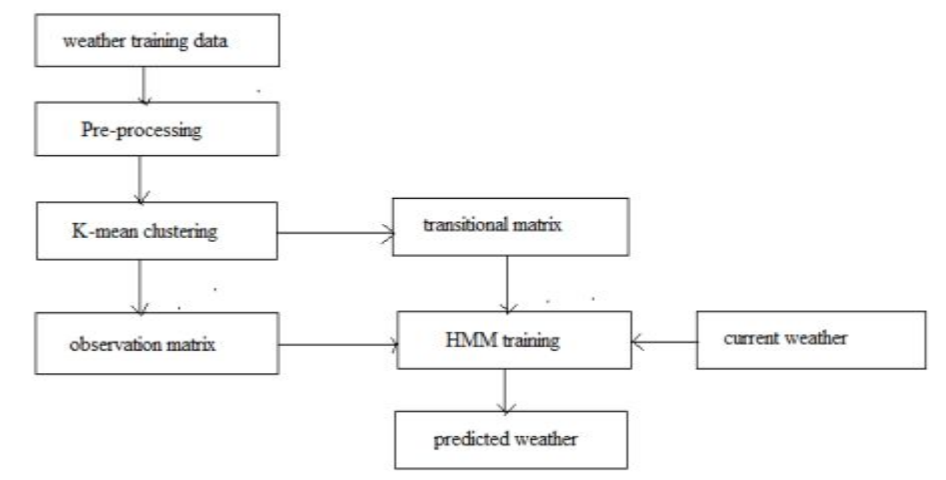


Fig: Block diagram showing the use of Data Mining

ML in Weather Forecasting- Linear Model

If we use basic model of linear regression:

- The first algorithm that was used was *linear regression*, which seeks to predict the *high and low temperatures* as a linear combination of the features.
- Only eight features were used: the *maximum temperature, minimum temperature, mean humidity, and mean atmospheric pressure* for each of the past two days.

Day	Linear Regression	Functional Regression	Professional
1	5.039	5.252	2.612
2	5.157	5.734	3.244
3	5.300	5.914	3.618
4	5.379	6.068	3.708
5	5.446	6.221	4.522
6	5.566	6.211	4.883
7	5.642	6.329	5.062

Table: RMS Error comparison

DATA Collection

Where from it can be collected:

- *Kaggle*
- *Weather Underground*
- *NOAA*
- *DATA.GOV*

Topic:

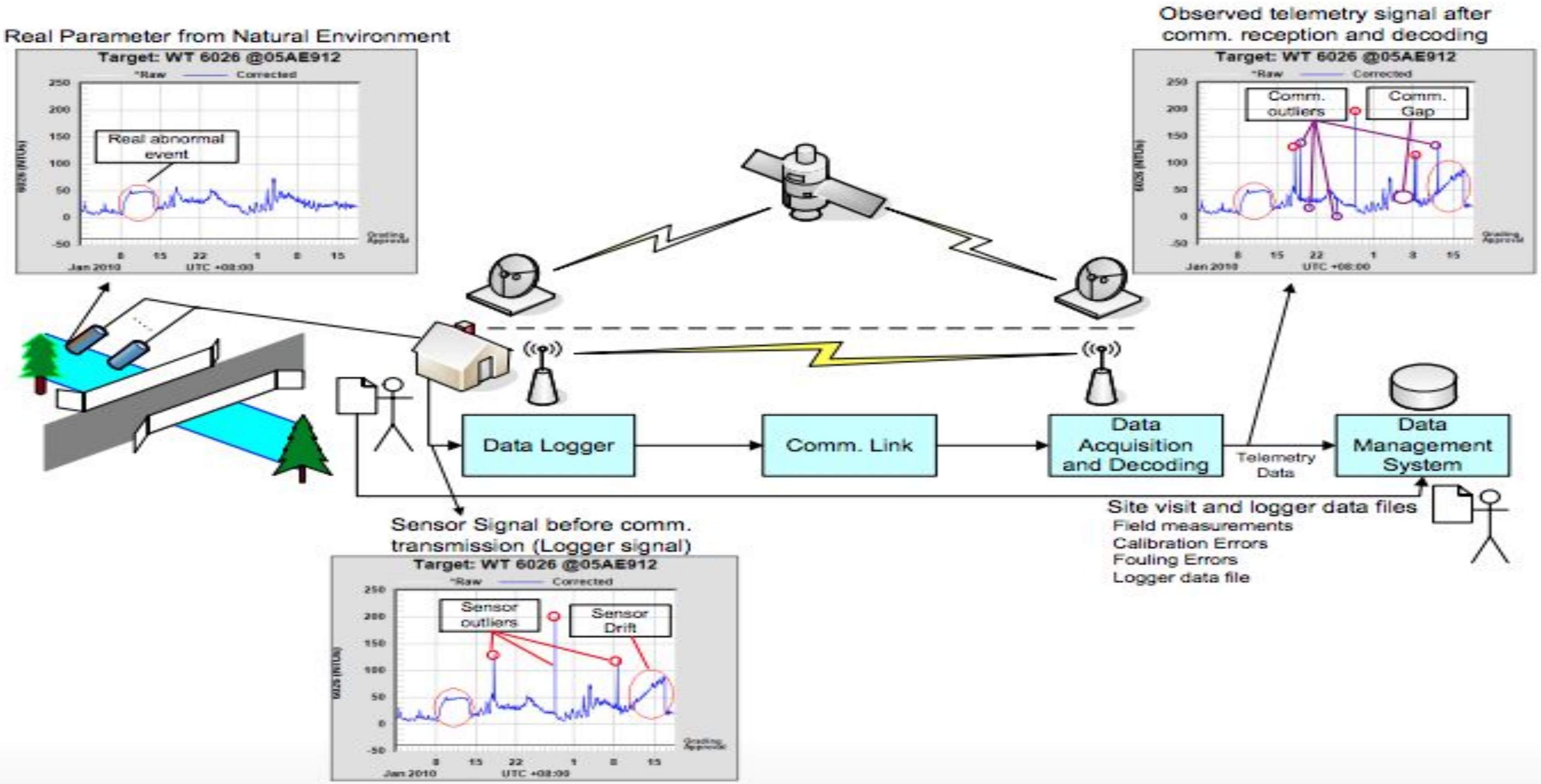
Water Monitoring & Hydrology

Machine Learning for Water Monitoring, Hydrology and Sustainability

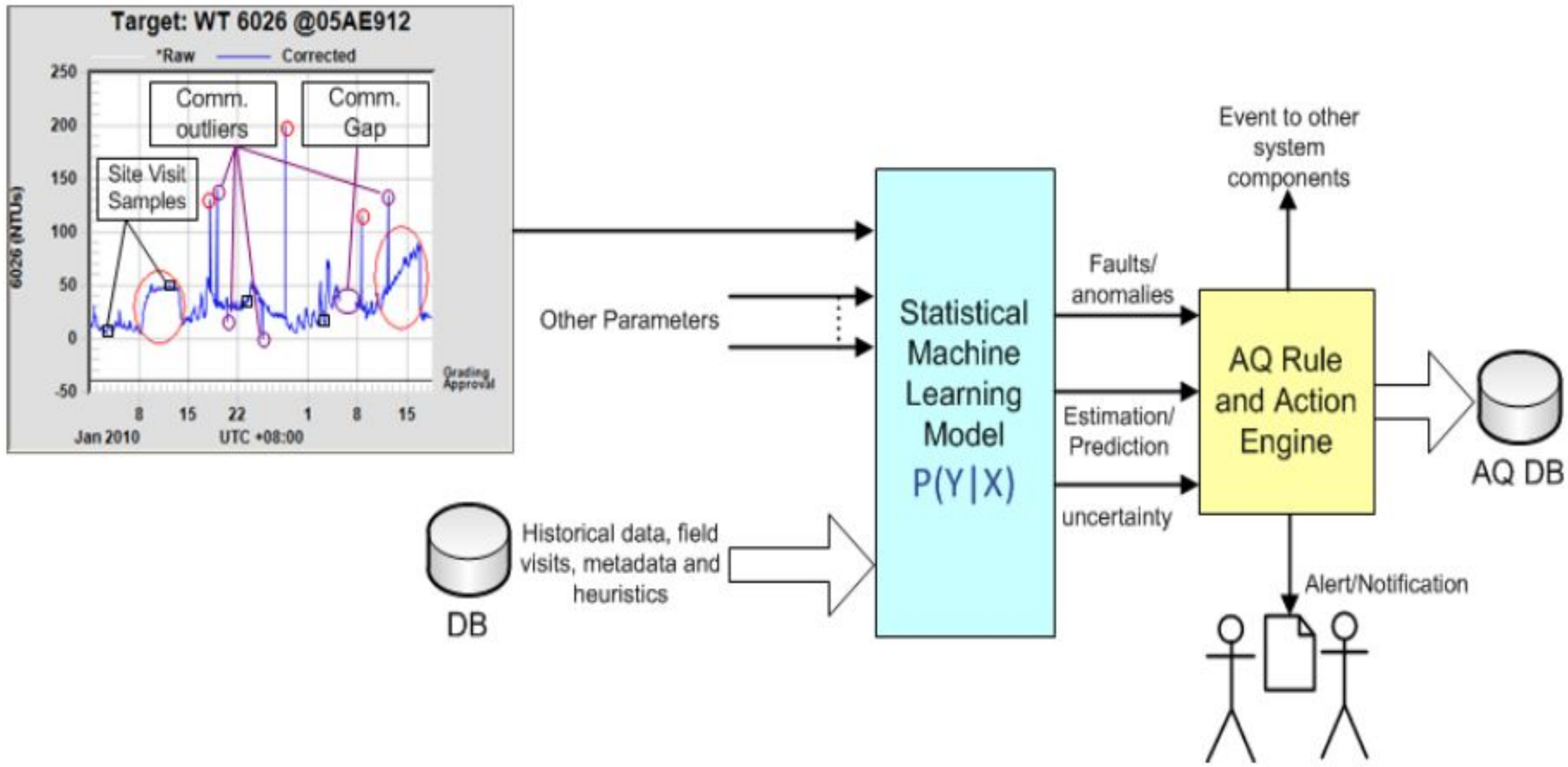
- ***Data is needed for***

Allocation, engineering design, prediction and forecasting, environmental impact assessments, transportation, fisheries and ecosystems management, resource extraction, industrial use, recreation

Data Acquisition and Management



Data Processing Pipeline



How Can Machine Learning Help

Machine learning can automate, simplify and improve many aspects of water monitoring including:

1) Improving modeling and analysis
2) Detecting and correcting equipment malfunctions
3) Detecting environmental anomalies
4) Predicting the effects of policy decisions
5) Automating and controlling allocation and distribution

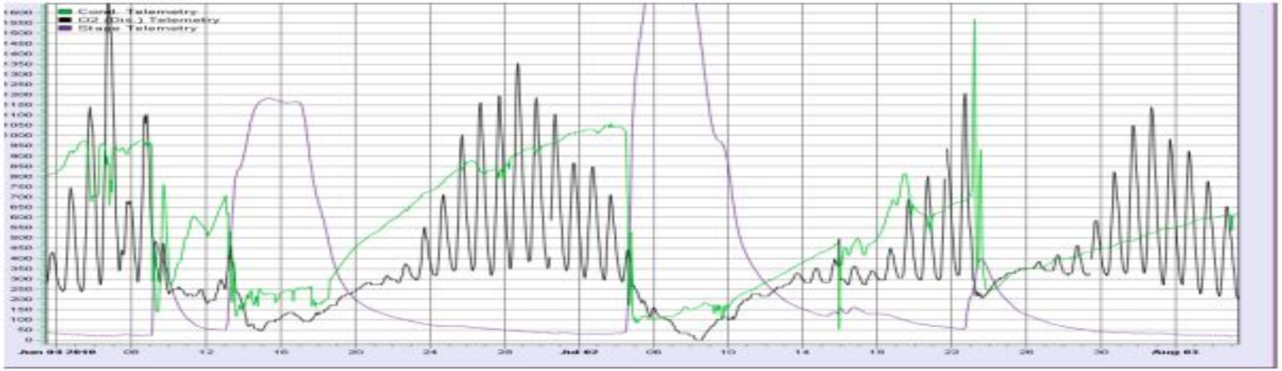
- Environmental time series in general are complex and hard to model Problems:
    1. *Highly non-stationary*
    2. *Highly non-linear*
    3. *Many changes in dynamics*
    4. *Can contain outliers, anomalies, gaps, etc.*
- Our models need to be:
    1. *General*
    2. *Flexible*
    3. *Robust*
    4. *Interpretable*
    5. *Fast and efficient for real-time applications*
    6. *Easy to setup and use*

- Our first approach is develop good probabilistic models for several basic problems
    1. *Gap filling/forecasting*
    2. *Fault detection*
    3. *Anomaly/outlier detection*
- Probabilistic models provide many beneficial properties that are important in an industrial setting
    1. *Consistent, unified framework*
    2. *Provides uncertainty in results*
    3. *Suggests natural extensions to deal with many kinds of issues*

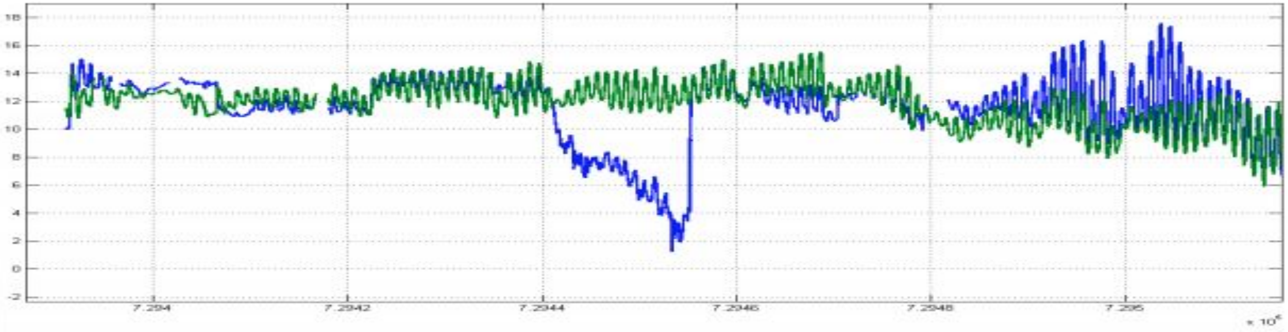
The Power of Redundancy

We can exploit correlated signals to build more robust models. Even simple linear methods work well under this regime.

Nonlinearly correlated signals from same sensor



Linearly correlated signals from different sensors



Topic:

ANN on Ultrasonic Flowmeter

ANN on Ultrasonic Flowmeter

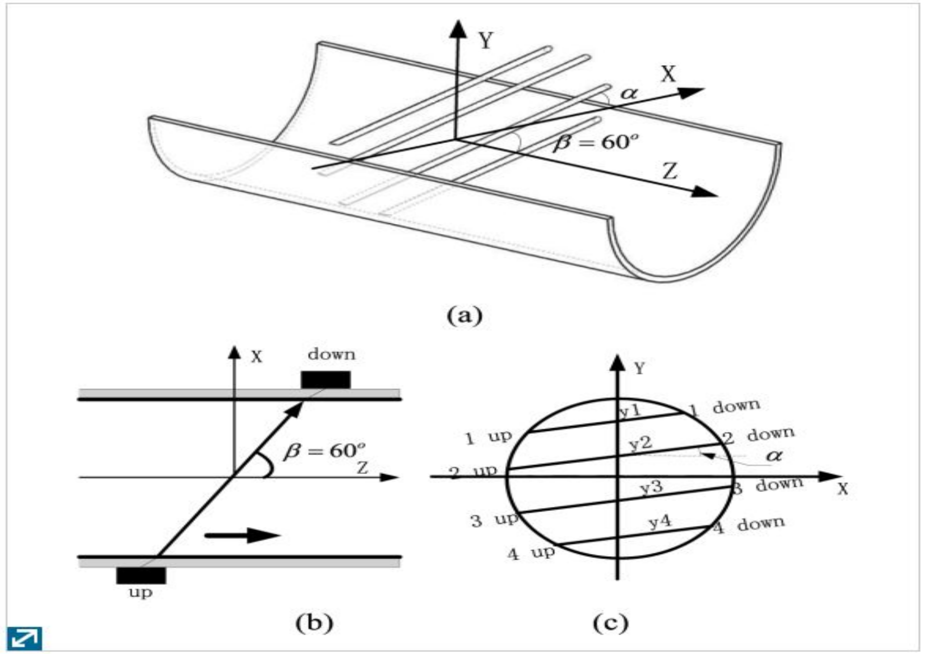
- Multipath ultrasonic flowmeters are considered to be able to provide a more accurate measurement of mean flow velocity on a cross-section of a pipe by taking into account the effect of flow profile and integrating the flow velocities measured from different sound paths.
- However, the practical limitations of industry sites, space constrains and various upstream pipe configurations may not allow for a sufficiently long straight pipe, which will result in the flow field on the measurement section not fully developed or even seriously distorted and the meter's readings erroneous.
- For these reasons, the investigation of proper weighted integration of the flow velocities measured from different sound paths is desired to obtain a robust and accurate estimation of the mean flow velocity on a cross-section of the pipe.

ANN Based Weighted Integration of Flow Velocities on Individual Sound Paths

To obtain an accurate estimation of the mean flow velocity over the cross-section of the pipe, the weighted integration of the flow velocities on the individual sound paths is carried out, which can be described as

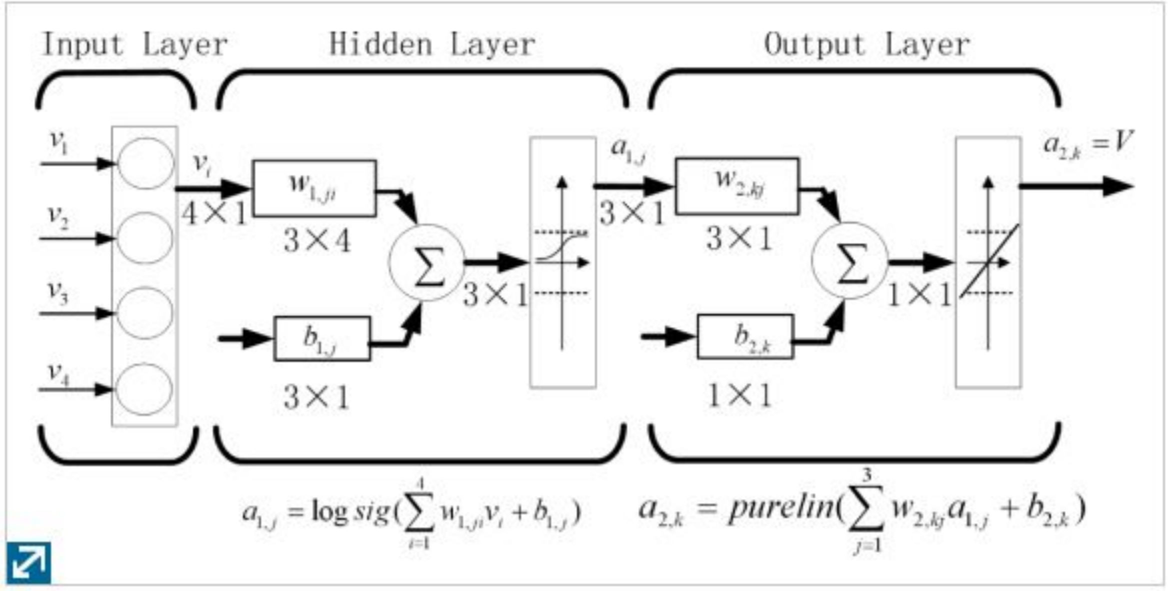
$$V=\sum w(i)v(i).$$

where v(i) is the measured flow velocity on the ith acoustic path, and w(i) is the integration weight related to the ith acoustic path.



ANN for Weighted Data Integration of Multipath Ultrasonic Flowmeters

A neural network with linear neurons was designed, in which the weights of the neurons directly correspond to the weights of individual ultrasonic paths. However, the performance of this method becomes deteriorated because of the non-linear relationship between the flow velocities on the individual sound paths and the mean flow velocity over the cross-section in the case of a strong asymmetric flow and swirling profile. To improve this, we designed a 3-layer feed-forward ANN with a strong non-linear mapping and learning ability.



Training Dataset

- Before the ANN can be used, it must be trained and tested by using a data set. The data set should include the flow velocities on the individual sound paths and the mean flow velocity over the cross-section of the pipe for different Reynolds numbers corresponding to the conditions under which the flowmeter is supposed to be used. We constructed the data set by means of CFD simulations.

- To obtain reasonable results consistent with the practical flow inside the pipe with single and double elbows, the Reynolds stress model (RSM) for turbulences and the enhanced wall treatment (EWT) were used in our simulations

Thank You