Ho Chi Minh City University Of Technology

Data Mining
Assignment Presentation

Merlion - AutoML

Lecturer: Mr. Le Hong Trang

Student:

Vo Cong Thanh	1814038
Phan Khanh Thinh	1814182
Nguyen Duy Thin	1814149

Outline

1. Introduction to the Problem
2. What is Merlion?
3. Merlion Features
4. Time Series Forecasting Process
5. AutoML for Forecasting
6. Data Description
7. Visualization
8. Quantitative Evaluation Metrics
9. Conclusions

MERLION

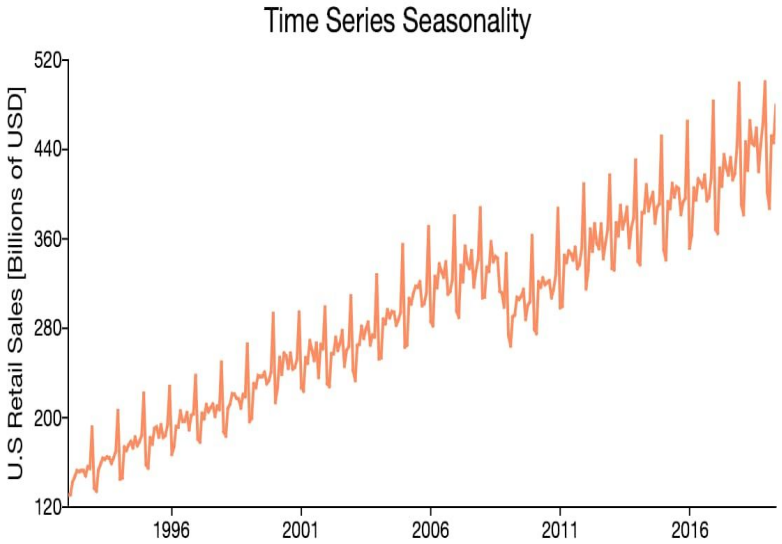
1.
Introduce to problem

Introduce to problem

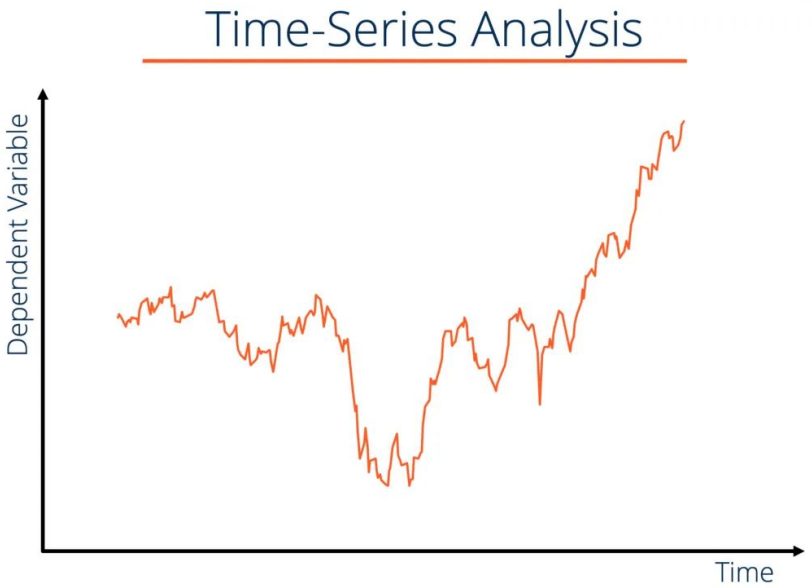
- IT operation management
- Manufacturing industry
- Cyber security
- Forecast the trends and values of key metrics accurately
- Detect anomalies rapidly and accurately
- Consistent

Introduce to problem

U.S Retail Sales



Stock Market



Time Series

- **Trend:** a long-term increase or decrease in the data which might not be linear. Sometimes the trend might change direction as time increases. I
- **Cycle:** exists when data exhibit rises and falls that are not of fixed period. The average length of cycles is longer than the length of a seasonal pattern. In practice, the trend component is assumed to include also the cyclical component. Sometimes the trend and cyclical components togethe are called as trend-cycle. I
- **Seasonality**: exists when a series exhibits regular fluctuations based on th season (e.g. every month/quarter/year). Seasonality is always of a fixed and known period.
- **Irregular Remainder:** a stationary process

2. Merlion

What is Merlion ?

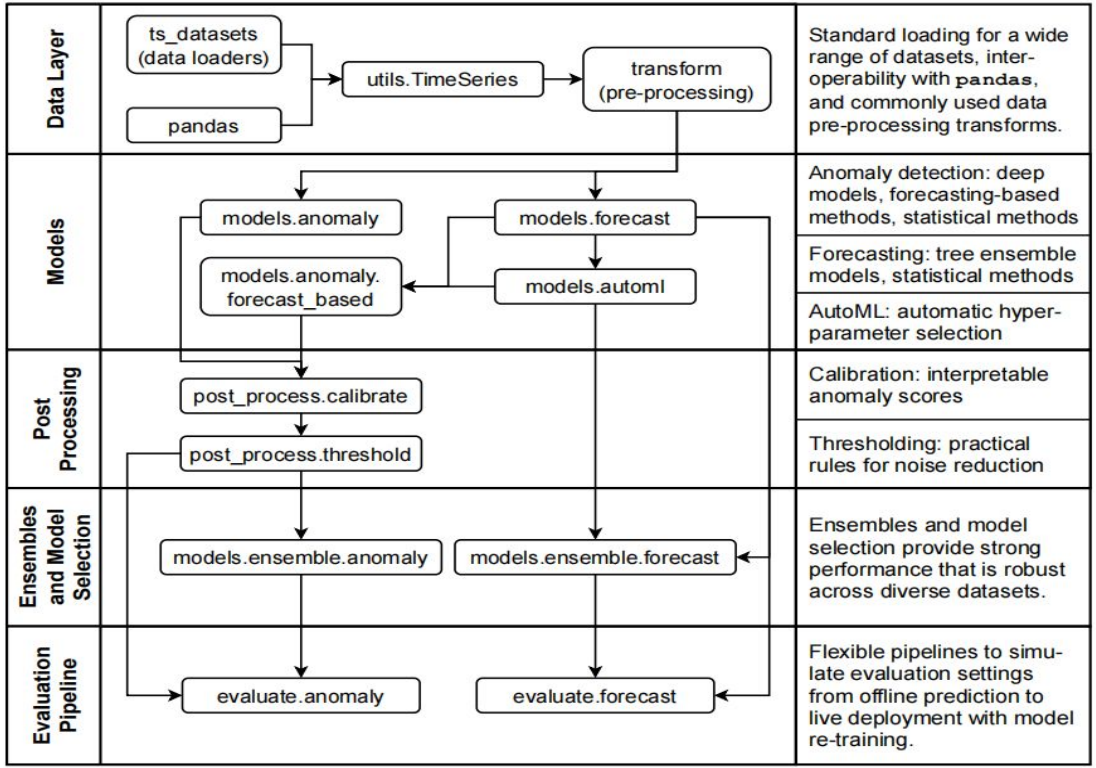
- Unifield interface
- Detect anomaly
- Forecasting in time series
- Ease-of-use
- Visualization and anomaly score calibration
- AutoML for hyperparameter tuning and model selection and model ensembling
- Framework simulates the deployment and re-training of a model

3.
Merlion Features

Features

	Forecast		Anomaly		AutoML	Ensembles	Benchmarks	Visualization
	Uni	Multi	Uni	Multi				
alibi-detect	–	–	✓	✓	–	–	–	–
Kats	✓	✓	✓	✓	✓	–	–	✓
statsmodels	✓	✓	–	–	–	–	–	–
gluon-ts	✓	✓	–	–	–	–	✓	–
RRCF	–	–	✓	✓	–	✓	–	–
STUMPY	–	–	✓	✓	–	–	–	–
Greykite	✓	–	✓	–	✓	–	–	✓
Prophet	✓	–	✓	–	–	–	–	✓
pmdarima	✓	–	–	–	✓	–	–	–
Merlion	✓	✓	✓	✓	✓	✓	✓	✓

Architecture and Principles



4.
Time Series Forecasting Process

Time Series Forecasting Process

1. Initializing a forecasting model (including ensembles and automatic model selectors)
2. Training the model
3. Producing a forecast with the model
4. Visualizing the model's predictions
5. Quantitatively evaluating the model
6. Saving and loading a trained model
7. Simulating the live deployment of a model using a ForecastEvaluator

5.
AutoML for Forecasting Process

AutoML for Forecasting

Traditional AutoML:

- Conventional hyper-parameter Optimization
- It automates the selection, composition and parameterization of machine learning models

In Merlion:

- Detection of some characteristics of time series
- Speed up model using approximation strategy: initial list of candidate model that achieve good performance
- Re-train each of these candidates until mode convergence, and finally select the best model by AIC

AutoML for Forecasting

Example in AutoSarima:

- Parameter of Sarima
  - Order: (20,1,5)
  - Seasonal Order: (1,12,1,20)
- HyperParameter Tuning of Sarima Model
  - Order: (auto, auto, auto)
  - Seasonal Order : (auto,auto,auto,auto)

6.
Data Description

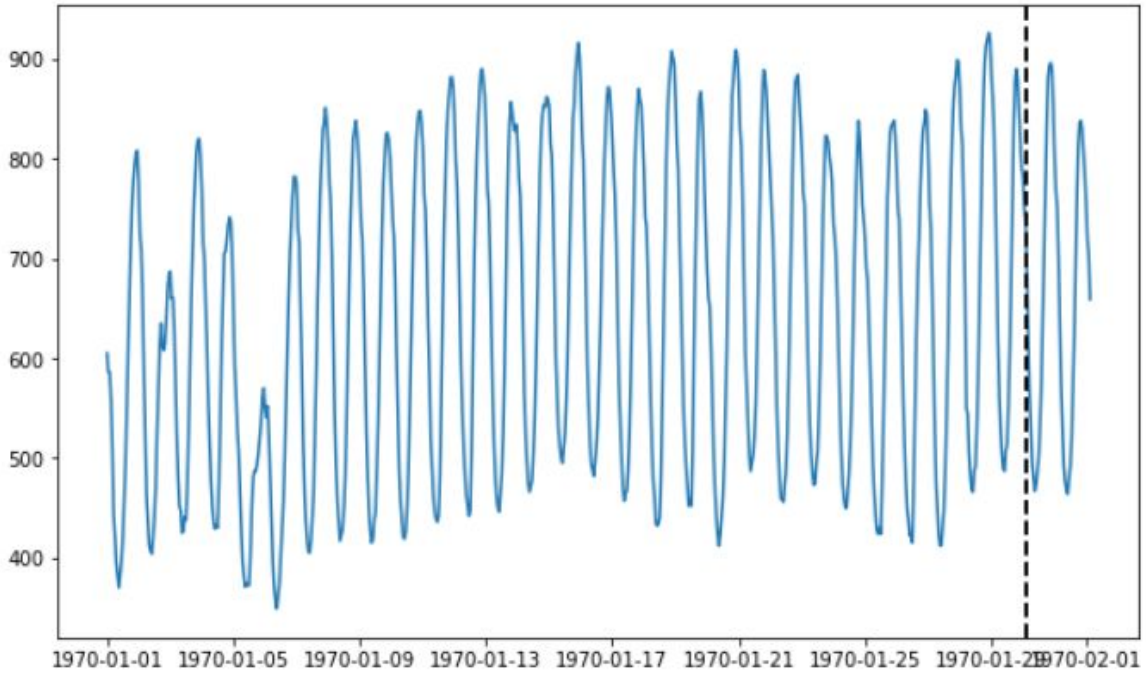
Data Description

M4's Hourly Dataset

	H1
1970-01-01 00:00:00	605.0
1970-01-01 01:00:00	586.0
1970-01-01 02:00:00	586.0
1970-01-01 03:00:00	559.0
1970-01-01 04:00:00	511.0
...	...
1970-01-31 23:00:00	785.0
1970-02-01 00:00:00	756.0
1970-02-01 01:00:00	719.0
1970-02-01 02:00:00	703.0
1970-02-01 03:00:00	659.0

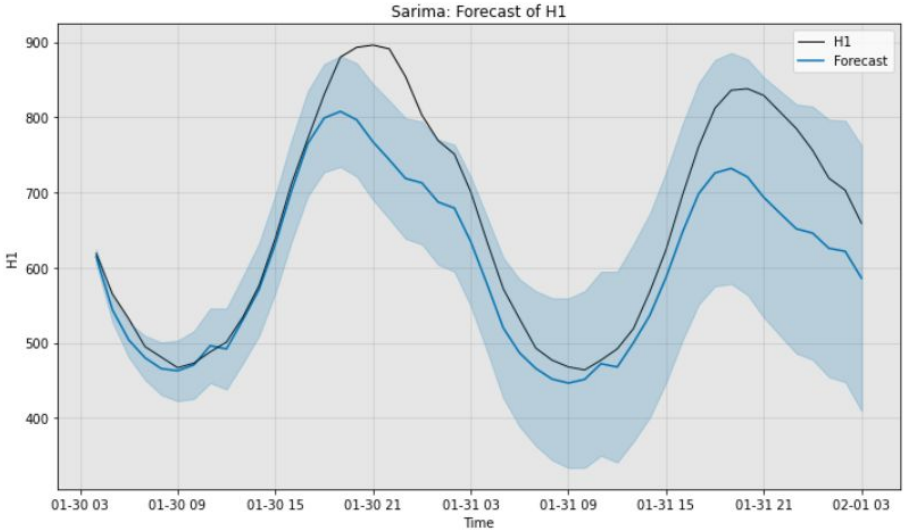
748 rows × 1 columns

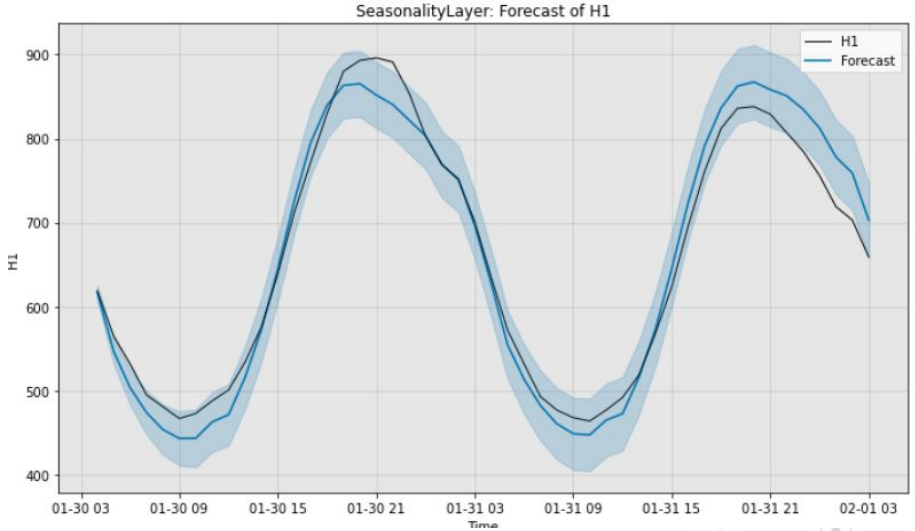
Line Graph of Dataset



7.
Data Visualization

Visualization





Quantitative Evaluation Metrics

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}},$$

$$\mathrm{sMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{|y_t| + |\hat{y}_t|} * 200(\%),$$

$$e_t = y_t - \hat{y}_t$$

	sMAPE	RMSE
Sarima	7.81	70.29
Arima	5.00	36.67
Prophet	3.72	32.06
MSES	35.03	191.49
ForecasterEnsemble	7.76	56.30
Selector	3.72	32.06
AutoSarima	3.50	27.61

Conclusion

- Extensible interfaces and implementations
- Improves the performance of multiple forecasting models
- Good performance on benchmark dataset
- Visualization module for more qualitative analysis

Thanks for listening!