Time Series Analysis Methods and Applications

by Igor Baranov, Michael Parravani, Ariana Biagi, Hui Fang Cai

Abstract The goal of this project is to discover Time Seties analysis modelsand algorithms, show different application.

Introduction

Background

Objective

Plan

Ethical ML Framework

As the goal of this application is only to research the time series methods, many aspects of the ethical ML framework do not directly apply. The data is open source and we can assume was collected in transparent ways. That being said, there is likely a large segment of the populace that is under represented in this ratings dataset - we assume a low income population (limited access to internet, limited time to be spent rating jokes, etc). This will potentially reduce the reccomender's accuracy for that group of the population. If the outcome of this system were to be of more social impact, this would need to be corrected with appropriate data collection methods.

Time Series Data understanding

ql

Dataset 0

Historical stock data for all current S&P 500 companies

Dataset 1

Yahoo Science labeled time series. This set is big, it should be downloaded and used locally: Yahoo Science labeled time series

Dataset 2

NAB Data Corpus: better just load from github directly to R script. NAB Data Corpus

Data preparation

Load the jokes rating data and do the basic checks

Time Series Methods Showcase

Time series decomposition

Time series decomposition is to decompose a time series into trend, seasonal, cyclical and irregular components. Frequency represents data which has been sampled at equispaced points in time: - frequency=7: a weekly series - frequency=12: a monthly series - frequency=4: a quarterly series

```
a <- ts(1:20, frequency = 12, start = c(2011, 3))
print(a)
```

```
#>      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
#> 2011           1   2   3   4   5   6   7   8   9  10
#> 2012  11  12  13  14  15  16  17  18  19  20
```

```
str(a)
```

```
#>  Time-Series [1:20] from 2011 to 2013: 1 2 3 4 5 6 7 8 9 10 ...
```

```
attributes(a)
```

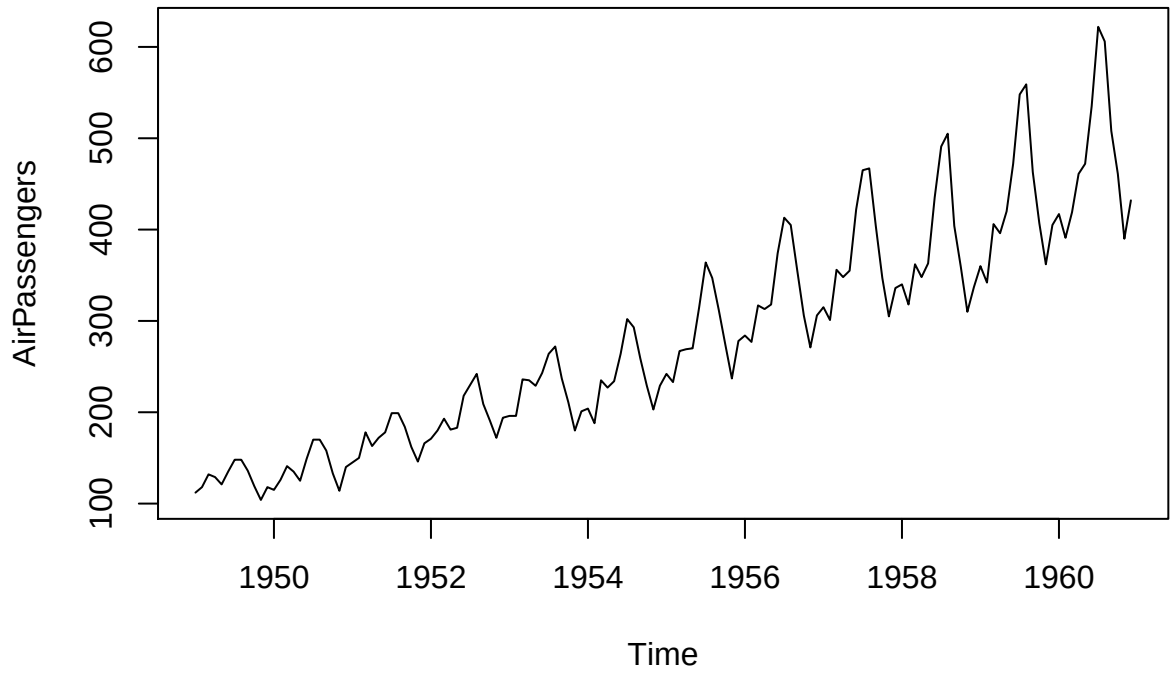
```
#> $tsp
#> [1] 2011.167 2012.750   12.000
#>
#> $class
#> [1] "ts"
```

To decompose a time series into components:

- Trend component: long term trend
- Seasonal component: seasonal variation
- Cyclical component: repeated but non-periodic fluctuations
- Irregular component: the residuals

A time series of AirPassengers is used below as an example to demonstrate time series decomposition. Data AirPassengers: monthly totals of Box Jenkins international airline passengers, 1949 to 1960. It has 144(=12×12) values

```
plot(AirPassengers)
```



Seasonal figures

```
f <- decompose(AirPassengers)
f$figure
```

```
#>  [1] -24.748737 -36.188131  -2.241162  -8.036616  -4.506313  35.402778
#>  [7]  63.830808  62.823232  16.520202 -20.642677 -53.593434 -28.619949
```

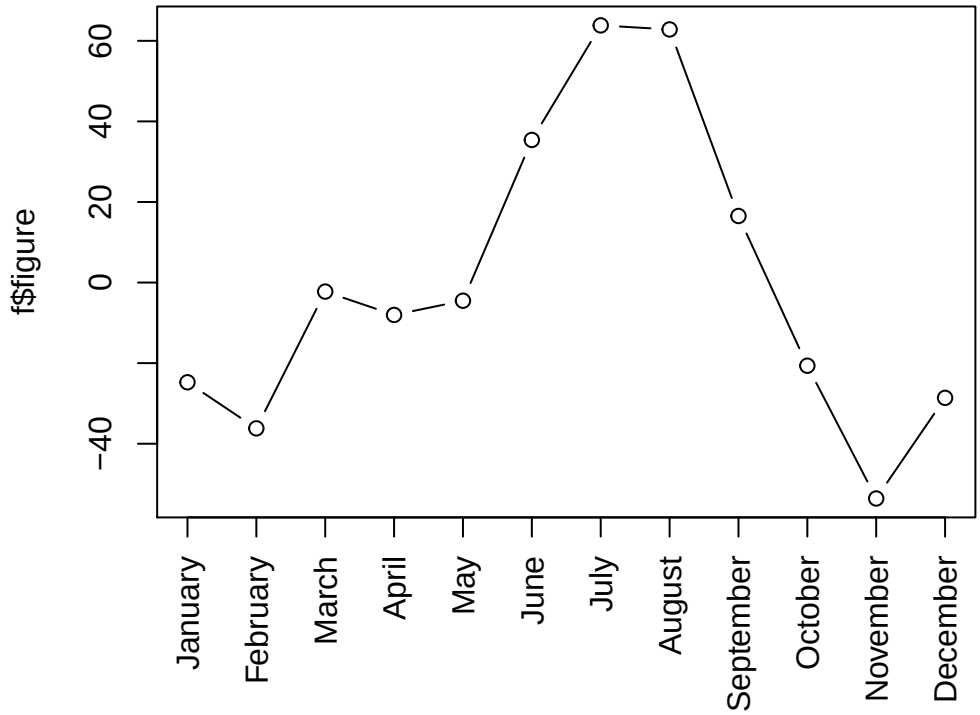


Figure 1: Cyclic Time Series Seasonal Component

Get names of 12 months in English words, label x-axis with month names and then 'las' is set to 2 for vertical label orientation (Fig **??**:

```
plot(f$figure, type="b", xaxt="n", xlab="")
monthNames <- months(ISOdate(2011,1:12,1))
axis(1, at=1:12, labels=monthNames, las=2)
```

In the figure 2, the first chart is the original time series, the second is trend, the third shows seasonal factors, and the last chart is the remaining component.

```
plot(f)
```

Time series forecasting

Time series forecasting is to forecast future events based on known past data. To forecast future events based on known past data For example, to predict the price of a stock based on its past performance. Popular models are:

- Autoregressive moving average (ARMA)
- Autoregressive integrated moving average (ARIMA)

Example below (Fig. 3) shows forecasting using ARIMA model.

```
fit <- arima(AirPassengers, order=c(1,0,0), list(order=c(2,1,0), period=12))
fore <- predict(fit, n.ahead=24)

# error bounds at 95% confidence level
U <- fore$pred + 2*fore$se
L <- fore$pred - 2*fore$se

ts.plot(AirPassengers, fore$pred, U, L, col=c(1,2,4,4), lty = c(1,1,2,2))
legend("topleft", c("Actual", "Forecast", "Error Bounds (95% Confidence)"), col=c(1,2,4), lty=c(1,1,2), cex=0
```

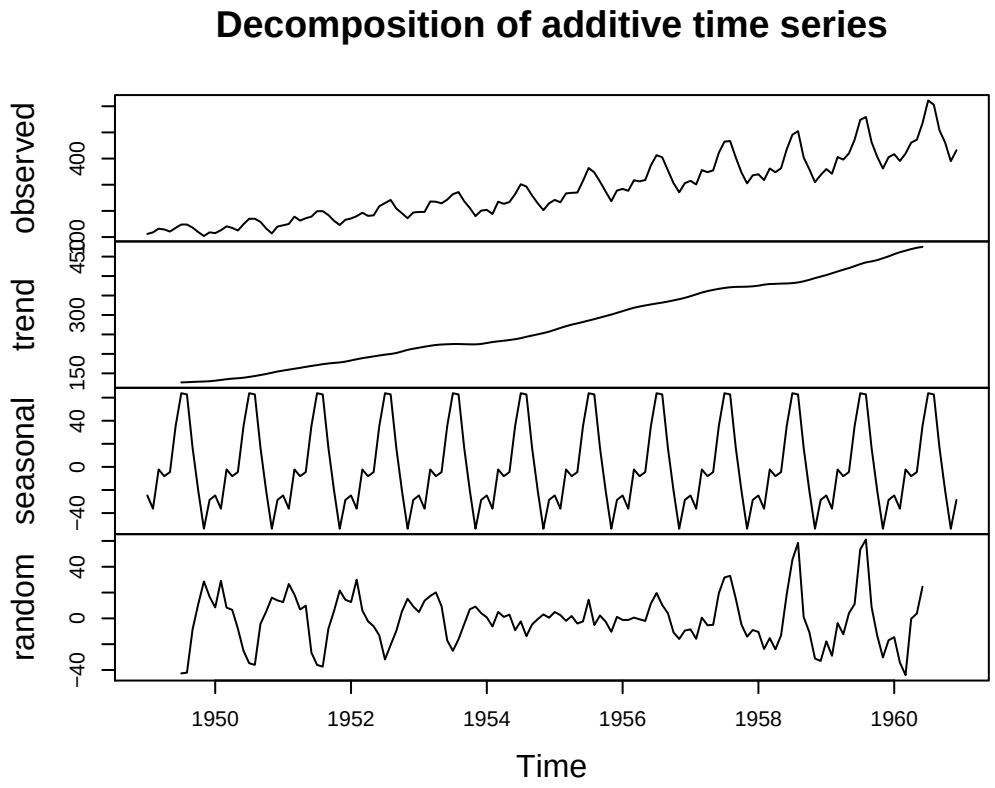


Figure 2: Cyclic Time Series Components

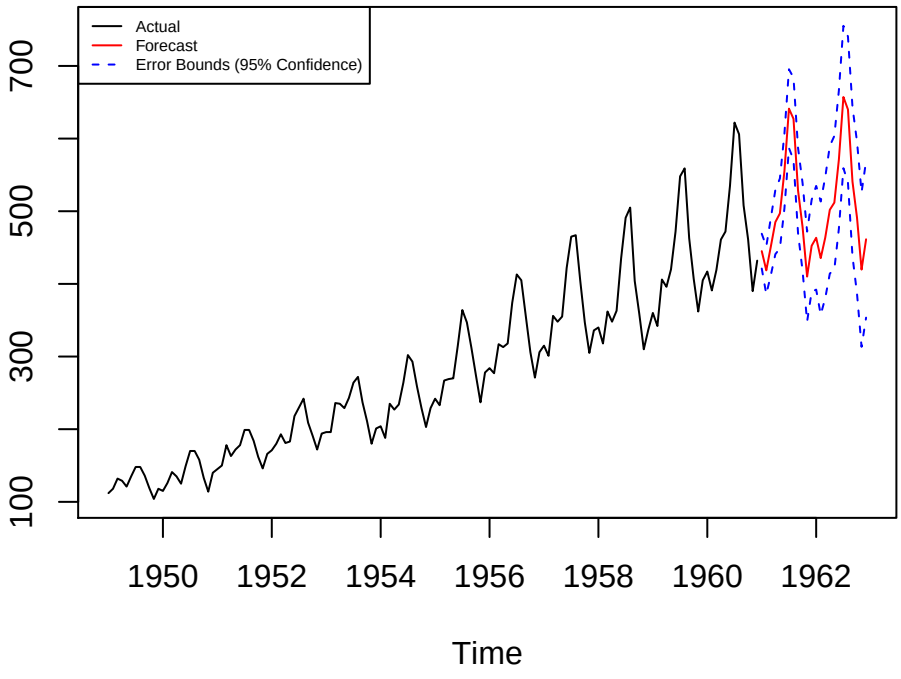


Figure 3: Cyclic Time Series Forecasting with ARIMA model

Example of Time-Series Analysis Practical Application

Prepare Shiny App for Deployment

Saving recommender data objects

```
#saveRDS(rcmnd_ub, file = "jokeRecommender.Rds")
#saveRDS(jokes, file = "jokes.Rds")
```

Deployment Discussion

This model is currently not of much use given its accuracy but it will serve as a proof of concept. This model could be used to help writers of movies/tv shows write jokes appropriate for a specific or large audience.

More data should be collected from this userbase to fill a training dataset. The dataset in its current state is quite sparse. The data would need to be updated every 3-5 years as people's taste changes and people within certian age groups mature. The Shiny app developed would be a deployment method to collect more data.

Further analysis could be done (with the appropriate data) to see how similar taste in humor is related to age.

The model developed in this project was used to create Shiny application currently deployed at ivbsoftware.shinyapps.io/JokeRecommender/. Code of the application could be found in Github.

Note from the Authors

This file was generated using *The R Journal* style article template, additional information on how to prepare articles for submission is here - Instructions for Authors. The article itself is an executable R Markdown file that could be downloaded from Github with all the necessary artifacts.

Igor Baranov
York University School of Continuing Studies

https://learn.continue.yorku.ca/user/profile.php?id=21219

Michael Parravani
York University School of Continuing Studies

Ariana Biagi
York University School of Continuing Studies

Hui Fang Cai
York University School of Continuing Studies