

Ανάλυση Χρονοσειρών με το Prophet

Δρ. Σωτήριος Δ. Νικολόπουλος

Big Data & Analytics

Πανεπιστήμιο Πελοποννήσου

Τμήμα Λογιστικής & Χρηματοοικονομικών

s.nikolopoulos@go.uop.gr

1 Ανάλυση Χρονοσειρών με το Prophet

1 Ανάλυση Χρονοσειρών με το Prophet

Ανάλυση Χρονοσειρών με το Prophet

Η ανάλυση χρονοσειρών είναι ένας τρόπος ανάλυσης και εκμάθησης της συμπεριφοράς συνόλων δεδομένων σε μια χρονική περίοδο.

Επιπλέον, βοηθά στην κατανόηση της συμπεριφοράς των δεδομένων με τη δημιουργία γραφημάτων.

Στον προγραμματισμό R, η ανάλυση χρονοσειρών μπορεί να πραγματοποιηθεί εύκολα χρησιμοποιώντας τη συνάρτηση `ts()` ή/και `xts()` και `zoo()`, η οποία λαμβάνει το διάνυσμα δεδομένων και το μετατρέπει σε αντικείμενο χρονοσειράς σύμφωνα με τις παραμέτρους που ορίζονται στη συνάρτηση.

Ανάλυση Χρονοσειρών με το Prophet

Το Facebook Prophet είναι ένα εργαλείο που αναπτύχθηκε από τη Facebook για την πρόβλεψη αντικειμένων ή δεδομένων χρονοσειρών.

Βοηθά τις επιχειρήσεις να κατανοήσουν τη συμπεριφορά των προϊόντων τους με την πρόβλεψη τιμών, πωλήσεων ή καιρού.

Το Facebook Prophet βασίζεται σε ένα διασπώμενο μοντέλο, σε τάση, εποχικότητα και αργίες, το οποίο βοηθά στη δημιουργία πιο ακριβών προβλεπτικών μοντέλων με αυτούς τους περιορισμούς.

Θεωρείται πιο αποτελεσματικό από το μοντέλο ARIMA καθώς βοηθά στη ρύθμιση και προσαρμογή των παραμέτρων εισόδου.

Το μοντέλο Prophet μπορεί να αναπαρασταθεί μαθηματικά ως:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (1)$$

Ανάλυση της εξίσωσης:

- $y(t)$: Αντιπροσωπεύει την προβλεπόμενη τιμή σε ένα συγκεκριμένο χρονικό σημείο t .
- $y(t): g(t)$: Αυτός ο όρος αποτυπώνει το στοιχείο της τάσης της χρονοσειράς. Το Prophet επιτρέπει την μοντελοποίηση διαφορετικών τύπων τάσεων, συμπεριλαμβανομένης της γραμμικής και της κορεσμένης ανάπτυξης.
- $y(t): s(t)$: Αυτός ο όρος αποτυπώνει το στοιχείο της εποχικότητας. Το Prophet μπορεί να μοντελοποιήσει την εποχικότητα διαφορετικών περιόδων, όπως ημερήσια, εβδομαδιαία ή ετήσια.

- $y(t): h(t)$: Αυτός ο όρος αποτυπώνει τις επιπτώσεις των εορτών στη χρονοσειρά.
Μπορείτε να καθορίσετε τις αργίες και την αναμενόμενη επίδρασή τους στην πρόβλεψη.
- $y(t): \varepsilon(t)$: Αυτός ο όρος αντιπροσωπεύει το στοιχείο του σφάλματος.
Αποτυπώνει την υπόλοιπη μεταβλητότητα στα δεδομένα που δεν εξηγείται από τους άλλους όρους.

Πρόσθετες Σημειώσεις:

Κάθε στοιχείο $(g(t), s(t), h(t))$ μπορεί να μοντελοποιηθεί περαιτέρω χρησιμοποιώντας συγκεκριμένες μαθηματικές συναρτήσεις.

Το Prophet χρησιμοποιεί διαφορετικές τεχνικές για κάθε στοιχείο, όπως σειρές Fourier για την εποχικότητα και τοπικές γραμμικές συναρτήσεις για την τάση.

Η συγκεκριμένη μορφή των συναρτήσεων που χρησιμοποιούνται για κάθε στοιχείο εξαρτάται από τις επιλεγμένες παραμέτρους του μοντέλου και τα χαρακτηριστικά των δεδομένων.

Είναι σημαντικό να σημειωθεί ότι ενώ αυτή η εξίσωση παρέχει μια γενική επισκόπηση του μοντέλου Prophet, η πραγματική υλοποίηση περιλαμβάνει πιο σύνθετους μαθηματικούς υπολογισμούς.

Εάν ενδιαφέρεστε για τις λεπτομέρειες των συγκεκριμένων συναρτήσεων που χρησιμοποιεί το Prophet, μπορείτε να ανατρέξετε στην τεκμηρίωση του Facebook Prophet:

Πρόσθετες Σημειώσεις:

Κάθε στοιχείο $(g(t), s(t), h(t))$ μπορεί να μοντελοποιηθεί περαιτέρω χρησιμοποιώντας συγκεκριμένες μαθηματικές συναρτήσεις.

Το Prophet χρησιμοποιεί διαφορετικές τεχνικές για κάθε στοιχείο, όπως σειρές Fourier για την εποχικότητα και τοπικές γραμμικές συναρτήσεις για την τάση.

Η συγκεκριμένη μορφή των συναρτήσεων που χρησιμοποιούνται για κάθε στοιχείο εξαρτάται από τις επιλεγμένες παραμέτρους του μοντέλου και τα χαρακτηριστικά των δεδομένων.

Είναι σημαντικό να σημειωθεί ότι ενώ αυτή η εξίσωση παρέχει μια γενική επισκόπηση του μοντέλου Prophet, η πραγματική υλοποίηση περιλαμβάνει πιο σύνθετους μαθηματικούς υπολογισμούς.

Προβλέψεις Χρονοσειρών χρησιμοποιώντας το Prophet του Facebook

Εάν ενδιαφέρεστε για τις λεπτομέρειες των συγκεκριμένων συναρτήσεων που χρησιμοποιεί το Prophet, μπορείτε να ανατρέξετε στην τεκμηρίωση του Facebook Prophet: <http://facebook.github.io/prophet/>

Προβλέψεις Χρονοσειρών χρησιμοποιώντας το Prophet του Facebook

Σημαντικοί Όροι στο Μοντέλο Facebook Prophet

1. Τάση (Tási)

Η τάση είναι μια μεταβολή στην εξέλιξη μιας χρονοσειράς, είτε αυξητική είτε μειωτική. Μαθηματικά, μπορεί να εκφραστεί με την εξίσωση:

$$y = C / (1 + \exp(-k(t - m)))$$

Όπου:

- C αντιπροσωπεύει τη μεταφορική ικανότητα (carry capacity).
- k αντιπροσωπεύει τον συντελεστή ανάπτυξης (growth).
- m αντιπροσωπεύει τη παράμετρο μετατόπισης (offset parameter).

Σημείωση: Αυτή είναι μια πιθανή μαθηματική εκδοχή για την τάση στο Prophet. Το συγκεκριμένο μοντέλο ενδέχεται να χρησιμοποιεί διαφορετική εξίσωση για την απεικόνιση της τάσης.

2. Εποχικότητα

Η εποχικότητα είναι ένα χαρακτηριστικό των χρονοσειρών που εμφανίζεται σε συγκεκριμένες χρονικές περιόδους/εποχές και επηρεάζει την τάση.

Για παράδειγμα, οι πωλήσεις παγωτού μπορεί να παρουσιάζουν αυξητική τάση γενικά, αλλά να έχουν και εποχικές διακυμάνσεις με μεγαλύτερες πωλήσεις τους καλοκαιρινούς μήνες.

Στο παρακάτω παράδειγμα, θα κατεβάσουμε το σύνολο δεδομένων AirPassengers και θα προβλέψουμε τις τιμές του χρησιμοποιώντας το μοντέλο Facebook Prophet.

Αφού εκτελέσετε ολόκληρο τον κώδικα, το μοντέλο θα εμφανίσει τις προβλεπόμενες τιμές μαζί με την τάση και την εποχικότητα των δεδομένων.

Βήμα 1: Εγκατάσταση της απαραίτητης βιβλιοθήκης

```
# Install the library  
install.packages("prophet")
```

Βήμα 2: Φόρτωση της απαραίτητης βιβλιοθήκης

```
# Load the library  
library(prophet)
```

Βήμα 3: Λήψη του συνόλου δεδομένων AirPassengers

```
# Dataset  
ap <- read.csv("example_air_passengers.csv")
```

Βήμα 4: Κλήση της συνάρτησης prophet για την προσαρμογή του μοντέλου

```
m <- prophet(ap)
```

Βήμα 5: Δημιουργία data.frame για να αποθηκευτούν οι προβλέψεις

```
# Predictions
```

```
future <- make_future_dataframe(m,  
periods = 365)
```

```
# Print predictions
```

```
cat("\nPredictions:\n")
```

```
> tail(future)
```

```
ds
```

```
504 1961-11-26
```

```
505 1961-11-27
```

```
506 1961-11-28
```

```
507 1961-11-29
```

```
508 1961-11-30
```

```
509 1961-12-01
```

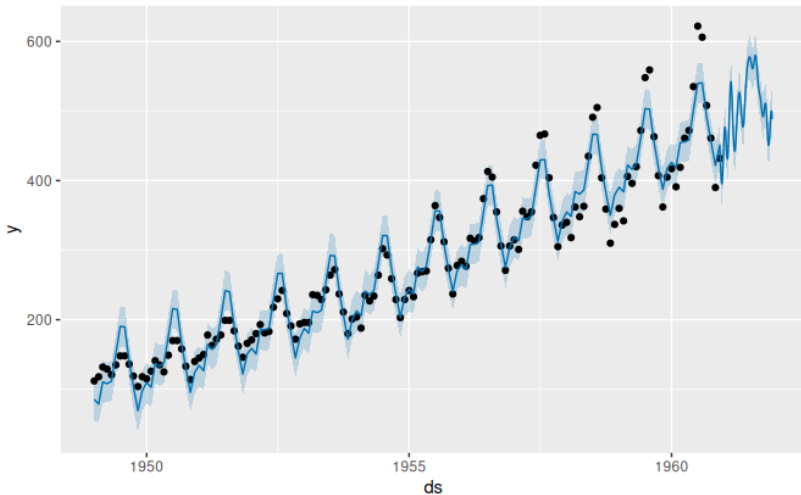
Βήμα 6: Πρόβλεψη δεδομένων

```
# Forecast
forecast <- predict(m, future)
tail(forecast[c('ds', 'yhat', 'yhat_lower', 'yhat_upper')])
ds          yhat    yhat_lower yhat_upper
504 1961-11-26 499.9036  471.6592  526.1495
505 1961-11-27 499.0477  468.5078  528.4497
506 1961-11-28 497.4603  467.5715  528.9536
507 1961-11-29 495.1649  468.0936  525.2362
508 1961-11-30 492.2037  462.0221  522.0661
509 1961-12-01 488.6356  459.6447  517.4455
```

Βήμα 7: Απεικόνιση της πρόβλεψης με γράφημα

```
# Plot
plot(m, forecast)
```

Υλοποίηση στον R



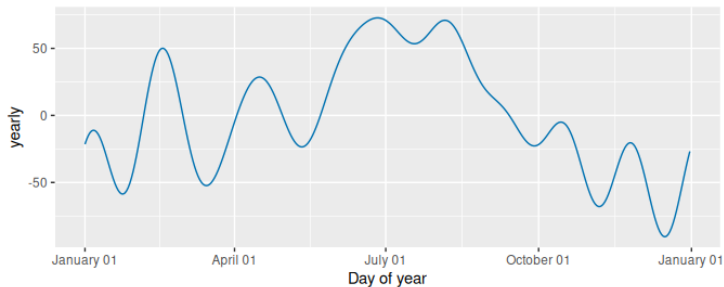
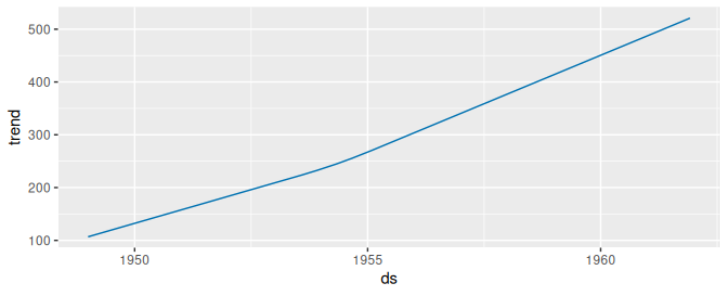
Το παραπάνω γράφημα απεικονίζει τις προβλεπόμενες τιμές των επιβατών (AirPassengers). Συγκεκριμένα:

- Οι μαύροι κύκλοι αναπαριστούν τα αρχικά δεδομένα.
- Η σκούρα μπλε γραμμή αναφέρεται στην προβλεπόμενη τιμή (\hat{y}).
- Η ανοιχτόχρωμη μπλε περιοχή υποδηλώνει τις τιμές \hat{y}_{upper} και \hat{y}_{lower} .

Βήμα 8: Απεικόνιση της τάσης, της εβδομαδιαίας και της ετήσιας εποχικότητας

```
# Plot  
prophet_plot_components(m, forecast)
```

Υλοποίηση στον R



Γράφημα 1: Τάση

Το παραπάνω γράφημα απεικονίζει την τάση του συνόλου δεδομένων, η οποία δείχνει ότι ο αριθμός των επιβατών αυξάνεται σε μια δεδομένη χρονική περίοδο.

Γράφημα 2: Εποχικότητα

Το δεύτερο γράφημα απεικονίζει την εποχικότητα του συνόλου δεδομένων σε μια δεδομένη χρονική περίοδο, συγκεκριμένα την ετήσια εποχικότητα.

Το γράφημα δείχνει ότι ο αριθμός των επιβατών είναι υψηλότερος μεταξύ των μηνών Ιουνίου έως Αυγούστου.