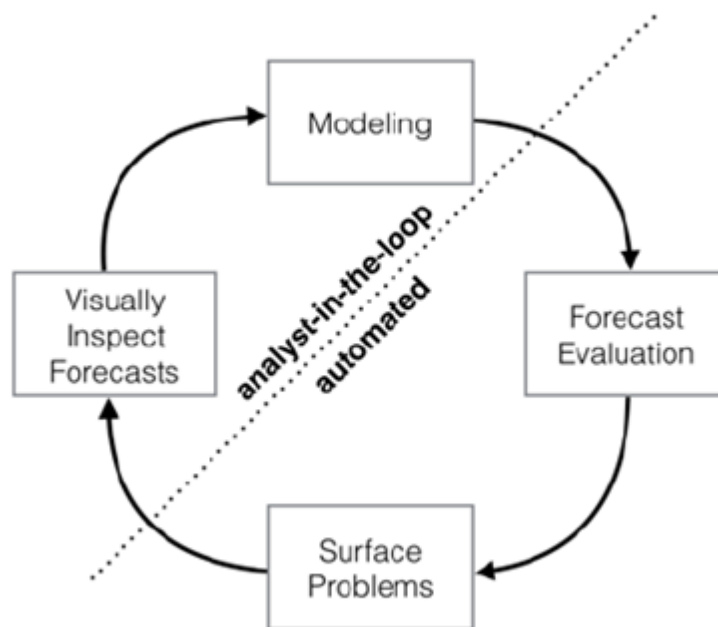


Forecasting at Scale

<https://peerj.com/preprints/3190.pdf>

0. Introduction



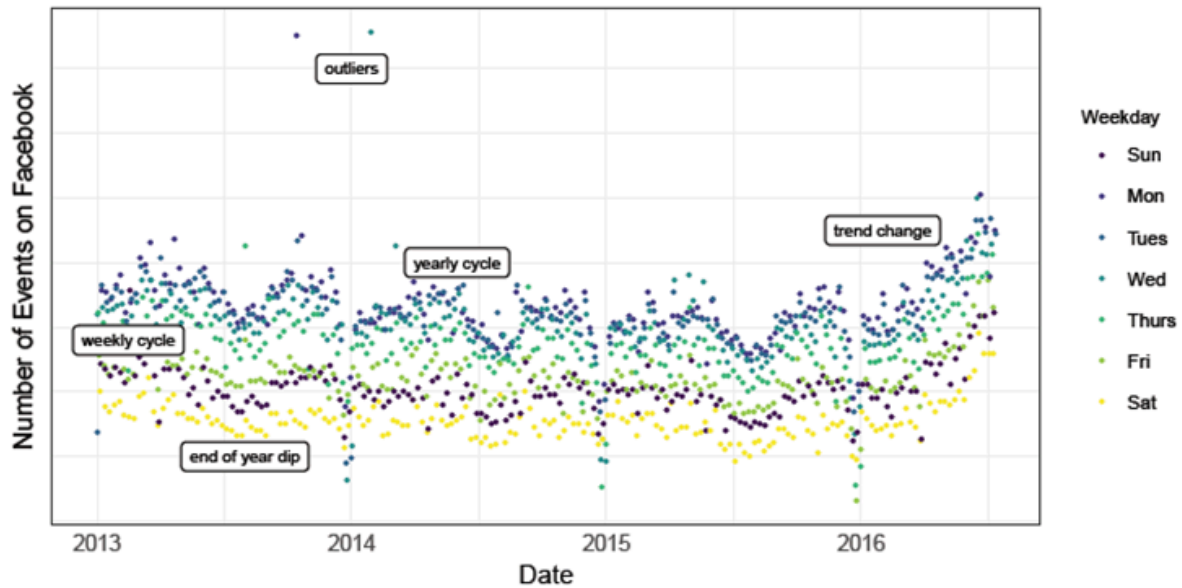
- 시계열 예측의 접근성 문제

- 전통적인 시계열 모델은 복잡한 통계 지식과 파라미터 튜닝이 필요하여 비전문가에게 접근이 어려움.
- Facebook은 이러한 문제를 해결하기 위해 비전문가도 쉽게 사용할 수 있는 시계열 예측 모델 Prophet을 제안.

- Prophet의 특징

- 분석가가 간단한 설정만으로 시계열 모델링이 가능하며, 예측 성능이 낮아지는 시기를 감지하고 피드백을 통해 모델을 개선할 수 있는 구조.

1. Business Time Series의 특성



- **복잡한 패턴**
 - 비즈니스 시계열 데이터는 주간, 연간 등의 계절성과 추세 변화가 명확하게 나타남.
 - 기존 모델들은 이러한 복잡한 패턴을 포착하는 데 한계가 있음.
- **기존 모델의 한계**
 - Auto.ARIMA, ETS, TBATS 등은 특정 패턴을 잘 포착하지 못하거나 추세 변화에 민감하게 반응함.
 - 비전문가는 이러한 모델의 파라미터를 조정하기 어려움.

2. Prophet Forecasting Model

- **모델 구성**

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \leftarrow$$

- Prophet은 시계열 데이터를 다음과 같이 구성:
 - $g(t)$: 추세(trend) 함수
 - $s(t)$: 계절성(seasonality) 함수
 - $h(t)$: 휴일(holidays) 효과
 - $e(t)$: 잔차(error)
- **모델의 장점**

- 유연한 추세 및 계절성 설정이 가능하며, 결측치나 불규칙한 시계열에도 강건함.
- 빠른 피팅 속도와 직관적인 파라미터 조정이 가능하여 비전문가도 쉽게 사용할 수 있음.

3. 주요 기능

$$g(t) = \frac{C(t)}{1 + \exp(-(k + \mathbf{a}(t)^\top \boldsymbol{\delta})(t - (m + \mathbf{a}(t)^\top \boldsymbol{\gamma})))}$$

3.1 비선형, 포화 성장 (Nonlinear, Saturating Growth)

- 성장 예측에서 상한이 존재하는 경우 로지스틱 함수를 사용하여 모델링.
- 상한(C)과 성장률(k)을 시간에 따라 변화하는 함수로 설정하여 현실적인 성장 모델링이 가능

$$g(t) = (k + \mathbf{a}(t)^\top \boldsymbol{\delta})t + (m + \mathbf{a}(t)^\top \boldsymbol{\gamma}),$$

3.2 선형 추세와 변화점 (Linear Trend with Change points)

- 상한이 없는 경우 선형 추세 모델을 사용하며, 추세 변화가 발생하는 지점을 자동으로 감지하여 모델에 반영.

3.3 계절성 (Seasonality)

- 푸리에 급수(Fourier series)를 사용하여 주기적인 패턴을 모델링.
- 주기(P)에 따라 급수의 항 개수(N)를 조정하여 오버피팅과 언더피팅을 방지

$$s(t) = \sum_{n=1}^N \left(a_n \cos \left(\frac{2\pi nt}{P} \right) + b_n \sin \left(\frac{2\pi nt}{P} \right) \right)$$

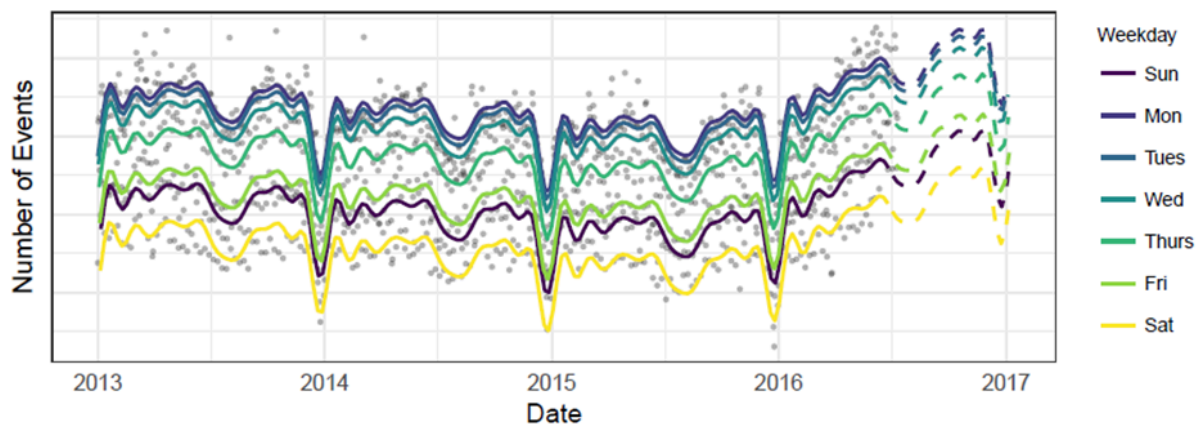
3.4 휴일 및 이벤트 (Holidays and Events)

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)]$$

$$h(t) = Z(t)\kappa.$$

- 불규칙적인 휴일이나 이벤트의 영향을 모델에 반영할 수 있도록 사용자 정의 가능.
- 휴일 전후의 영향도 조정할 수 있어 현실적인 예측이 가능.

4. 결론 및 시사점



- **접근성 향상**
 - Prophet은 비전문가도 쉽게 사용할 수 있는 시계열 예측 도구로, 복잡한 모델링 없이도 정확한 예측이 가능.
- **실용성 강조**
 - 빠른 피팅 속도와 직관적인 파라미터 설정으로 실무에서 유용하게 활용될 수 있음.
- **도메인 지식의 중요성**
 - 휴일 효과나 성장 상한 등의 설정에는 도메인 지식이 필요하므로, 전문가의 판단이 모델 성능에 영향을 미칠 수 있음.