Predicting Hotel Demand Using Destination Marketing Organizations’ Web Traffic Data

Bing Pan, Ph.D.
College of Charleston and Hong Kong Polytechnic University

Haiyan Song, Professor and Ph.D.
Hong Kong Polytechnic University

Yang, Yang
University of Florida
Content of the Talk

1. Inspiration for the study
2. Studies on forecasting hotel demand
3. Studies on forecasting economic activities using online data
4. Research question and sub-questions
5. Research procedure and results
6. Conclusions and Implications
The Inspiration

Center for Disease Control uses case reports from doctors’ offices and hospitals to monitor influenza outbreak.

Every time you search on Google, you leave a trace of your query along with the time of access and your IP address/cookies.

Google and CDC Scientists use search volumes for certain queries from Google to predict flu outbreak in the U.S. and around the world.

The predictors include 45 queries such as “influenza complication”, “cold/flu remedy”, “General influenza symptoms”, etc.

Influenza percentages estimated by the model (black) and provided by the CDC (red) in the mid-Atlantic region.

Can we use online behavioral data to predict tourism economic activities?
Forecasting methods for Tourism and Hotel Demand

- Expert Judgment
- Moving Average
- Autoregressive Models
- Monte Carlo
- Neural Network
- Exponential Smoothing
- Regression
- Pickup Methods
Forecasting methods for Hotel Demand

Andrew, Cranage and Lee (1990) used Box-Jenkins method and exponential smoothing to forecast monthly occupancy of an actual hotel.

Schwartz and Hiemstra (1997) used past booking curves with a similar shape to forecast daily occupancy rates.

Weatherford and Kimes (2003) demonstrated the superiority of exponential smoothing and pickup methods.
Forecasting methods for Hotel Demand

Law (1998) adopted neural network approach to forecast Hong Kong's hotel occupancy.

Choi (2003) identified economic indicators for the hotel industry in the United States and used composites of those variables as early indicators to predict recession for the hotel industry.
Forecasting with Online Data

Choi and Varian (2009) used search engine data on “jobs” and “welfare/unemployment” in an ARIMA model to predict unemployment claims;

Askitas and Zimmermann (2009) demonstrated the strong correlation between keyword searches and unemployment rates in Germany using monthly data in a simple EC model.

Zhang, Jansen and Spink (2009) estimated a number of ARIMA models using raw search engine keyword data to detect changes in user behaviour across different time periods.
Forecasting with Online Data

Gruhl et al. (2005) used automated data mining on blogs to predict the volume of book sales.

Zhang et al. (2010) used six months of Twitter feeds and the volumes of tweets containing the expression of hope and fear each day to correlate them with stock market.

Bollen, Mao, and Zeng (2011) later confirmed their findings. A mutual fund was once set up using their models to invest in stock market.
Forecasting with Online Data

Lampos and Cristianini (2010) also used the volume of Twitter messages on certain keywords to track and predict flu epidemics.

Asur and Huberman (2010) used the rate of chatter from three million tweets to construct a linear model to predict box-office revenues of movies in advance of their release.
Advantages of Using Online Data

Past methods are based on historic performance, which rely on a consistent pattern of tourist activities and a stable economic structure.

Online data are real-time, high-frequency (daily and weekly instead of quarterly or annually), and sensitive to user behaviour's small changes.

Researchers in other fields have proven that the data are very valuable in generating accurate forecasts.
Predicting Tourism Demand with Online Data

Choi, Hyunyong, and Varian (2009) incorporated Google search volume data to predict visitor arrival in Hong Kong from nine different countries.

Pan, Wu & Song (2012) used five travel-related queries to a tourist destination, combined with Autoregressive Moving Average model, to predict hotel room demand.

These two studies are that they are ex-post in nature: both studies used the online data in the current time to predict the current tourist activities (Song, 2008).
Main Research Question

Can we use DMO’s website traffic to predict future hotel demand in one destination?
Data Description

We picked Charleston, South Carolina, USA

(1) STR Inc. provides hotel demand data; samples 110/190 hotels in Charleston; provide weekly occupancy (occupancy) and weekly roomnights sold estimated from occupancy (demand)

(2) Charleston Area Convention Visitors Bureau provides web log data through Google Analytics: visits and visitors

Visitors: identified by unique tracking cookies

Visits: number of individual sessions by visitors. Sessions are continuous web page access within a site with any adjacent access less than 30 minutes.
Sub-Research Questions

(1). Which data is more helpful to forecast hotel demand/occupancy, the volume of web visits, or the number of web visitors?

(2). Which type of forecasting model should be used to incorporate web log data: the model incorporating it as a direct predictor, or the model incorporating it as an indirect threshold variable?

(3). Does the model have the same short-run and long-run forecasting accuracy?
Sub-Research Questions

(4). What is the best model in the maximum amount of reduction in error rates?

(5). How much can the web hit data contribute to the reduction of error rates?
Research Step 1

Explorative analysis, correlation analysis, and stationarity test.
Plots of Time-Series in Logarithm

- Invisitors
- Invisits
- Indem
- Inocc
Cross-Correlogram of Web Traffic and Hotel Demand
Step 1 Results

(a) The modified Dickey-Fuller Test: stationary and no unit root

(b) Cross-Correlogram: lag of 4 represents largest correlation

(c) Impulse-Response Analysis: web traffic has the largest shock on demand about four weeks
Research Step 2

Tests with autoregressive moving average (ARMAX) Model and threshold autoregressive (TAR) model.

Week 21 of 2007 to week 46 of 2010 (182 weeks) was chosen as the estimation sample and the period of week 47 of 2010 to week 2 of 2011 (8 weeks) as the validation sample.
ARIMA Model

\[ y_t = \alpha + \sum_{i=0}^{j} \beta_i x_{t-i} + \mu_t \]

\[ \mu_t = \sum_{i=1}^{m} \rho_i \mu_{t-i} + \sum_{j=1}^{n} \theta_j \varepsilon_{t-j} + \varepsilon_t \]
TAR Model

\[ y_t = \begin{cases} 
\phi_1 + \sum_{i=1}^{m} \phi_{1i} y_{t-i} + \varepsilon_t & \text{for } x_{t-1} \leq \psi \\
\phi_2 + \sum_{i=1}^{m} \phi_{2i} y_{t-i} + \varepsilon_t & \text{for } x_{t-1} > \psi 
\end{cases} \]
Step 2 Results

(a) ARIMA Model: the lag length of 1 and 52 for AR terms, and 2, 3, 4, and 6 for MA terms.

(b) TAR Model: lag lengths up to 6.

(c) *Ex ante* forecasting: TAR Models are worst performing; ARMAX is better than ARMA for predicting 4 weeks ahead, but not for 8 and 20 weeks ahead forecasting.
Research Step 3

Forecasting improvement due to ARMA -> ARMAX model.
Rates of Error Reduction

4 weeks ahead forecasting:

MAPE: improvement by 14.5%
RMSPE: improvement by 0.81%

RMSPE penalizes extreme prediction errors
Research Step 4

Testing different splits in the estimation and validation subsets.
Step 4 Results

30 different splits:

Web traffic data are useful in forecasting hotel roomnights sold.
4 weeks ahead: Improvement is 7.4% by MAPE and 5.9% by RMSPE
8 weeks ahead: 10.6% by MAPE and 6.3% by RMSPE

ARMAX models are most significant when predicting large values
Conclusions

(1) *visits* and *visitors* are almost equally effective in predicting demand for hotel rooms, but not occupancy rates.

(2) The ARMAX model that incorporates the web traffic data as the predictive variable is more effective than the TAR model;

(3) the web traffic data are most useful in predicting demand for hotel rooms four- or eight-weeks-ahead.
Conclusions

(4) The ARIMA model incorporating the web traffic data is most effective in improving the accuracy of forecast demand for hotel rooms with a specific configuration of the lagged variables.

(5) The ARMAX model is most useful during peak seasons when hotel demand is high.
Discussions and Implications

This study validated the value of web traffic data of a DMO in help to predict local hotel demand.

Furthermore, other online data, in the forms of tweets, blogs, or likes on social media, could contribute to larger amount of increase in forecasting accuracy. This study points to the future direction of using a variety type of online data to predict tourist activities and hotel demand.
Discussions and Implications

Local DMOs and convention and visitors bureaus could post their web traffic data to their local hotel constituencies and help them with accurate forecasting.

Specific models need to be tested for unique destination.

Planning horizon is expected to get shorter in the future.
The Future of Real-Time Tourism Monitoring

Website traffics;
Search engine query volumes;
Log data from mobile devices;
Facebook, Twitter, Weibo, and blog content and volume;
Pictures and videos posted.
Forecasting Tourism with Online Pulse Data

- Website Log
- Search Engine Query Volume
- Facebook Updates
- Tweets
- Flickr
- Mobile Access
- Various Sensors
- Cameras

A Smart Tourism Monitoring System
Thank You and Questions