

Vorhersage touristischer Ankünfte

Verbesserung durch Big Data und neuronale Netze

Prof. Dr. Wolfram Höpken, Tobias Eberle,
Forschungsgruppe Data Science, Institut für Digitalen Wandel (IDW)

Einführung

- Akkurate Vorhersage touristischer Nachfrage aufgrund hoher Volatilität sehr wichtig für touristische Entscheider
- Suchverhalten im Web (repräsentiert durch Google-Trends-Daten) als interessanter Input zur Verbesserung der Vorhersage

Methodik

Datenbeschaffung und -aufbereitung

- Touristische Ankünfte in Schweden von 2008 – 2016 für 5 Sendeländer
- Statistische Aufbereitung der Ankunftsdaten und Herstellung von Stationarität

Konstruktion aggregierter Suchindizes

- Identifikation relevanter Suchbegriffe mittels des Google Keyword Planner
- Extraktion zugehöriger Google-Trends-Daten für die Kategorie Travel and Tourism
- Normalisierung der Suchbegriffe und Konstruktion aggregierter Suchindizes (mit unterschiedlichem Time Lag für unterschiedliche Suchbegriffe)

Vorhersagemodelle für touristische Ankünfte

- ARIMA-Modelle (Autoregressive Integrated Moving Average) (1) autoregressiv d.h. nur mit vergangenen touristischen Ankünften und (2) zusätzlich mit Suchindizes
- Neuronale Netze (MLP = Multilayer Perceptron) (1) autoregressiv und (2) zusätzlich mit Suchindizes

Validierung der Vorhersagemodelle

- Mittelfristiger Vorhersagehorizont von 6 Monaten
- Out-of-Sample-Validierung (83 Monate Trainingsdaten und 6 Monate Testdaten)

Literatur

1. Box, G. E., and G. M. Jenkins. 1970. Time series analysis, forecasting and control. San Francisco: Holden Day.
2. Fuchs, M., W. Höpken, and M. Lexhagen. 2014. „Big data analytics for knowledge generation in tourism destinations – A case from Sweden“. Journal of Destination Marketing and Management, 3 (4):198-209.
3. Höpken, W., D. Ernesti, M. Fuchs, K. Kronenberg, and M. Lexhagen. 2017b. „Big data as input for predicting tourist arrivals.“ In Information and Communication Technologies in Tourism 2017, edited by R. Schegg and B. Stangl, 187-200. New York: Springer.
4. Höpken, W., T. Eberle, M. Fuchs, and M. Lexhagen. 2019. „Google trends data for analysing tourists’ online search behavior and improving demand forecasting: The case of Åre, Sweden. Information Technology & Tourism, 21 (1): 45-62.

Zielsetzung

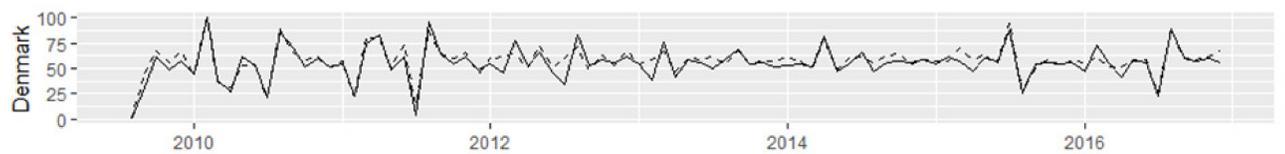
- Erweiterung autoregressiver Vorhersagemodelle (allein auf Basis vergangener Ankünfte) um Suchverhalten im Web (Google-Trends-Daten)
- Verwendung neuronaler Netze (als Verfahren des maschinellen Lernens) anstelle klassischer Zeitreihenmodelle (ARIMA)

Zusammenfassung

- Google-Trends-Daten ermöglichen deutliche Verbesserung der Vorhersage touristischer Ankünfte
- Neuronale Netze liefern genauere Vorhersagen als ARIMA-Modelle bei Verwendung von Google-Trends-Daten als zusätzlicher Input für die Vorhersage

Ergebnis

Monatliche touristische Ankünfte und Suchindex (gestrichelte Linie) für Sendeland Dänemark



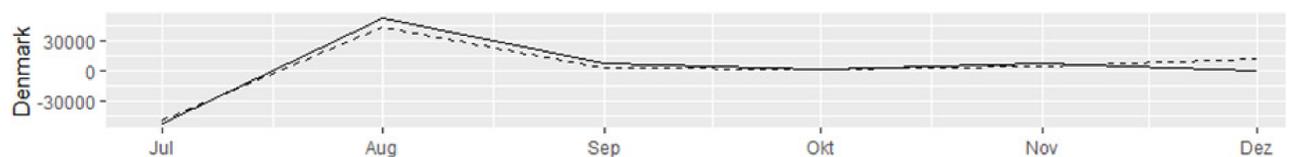
Ergebnisse aller vier Vorhersagemodelle und Sendeländer

Sendeland	Autoregressive Moving Average (ARIMA)				Neuronales Netz (ANN)				Differenz ANN-ARIMA	
	Autoregressiv (RMSE)	Mit Google Trends (RMSE)	Relative Diff. (%)	Shapiro-Wilk	Autoregressiv (RMSE)	Mit Google Trends (RMSE)	Relative Diff. (%)	Shapiro-Wilk	Autoregressiv	Mit Google Trends
Dänemark	24.705,94	6.489,38	-73,73	0,043	30.071,50	6.342,25	-78,90	0,700	+ 21,72 %	- 2,27 %
Finnland	4.076,07	1.564,45	-61,61	0,872	4.929,33	1.441,38	-70,75	0,939	+ 20,93 %	- 7,87 %
Norwegen	22.458,49	15.690,90	-30,13	0,708	14.782,61	12.031,29	-18,61	0,902	- 34,18 %	- 23,32 %
Russland	2.759,68	1.994,37	-27,73	0,633	2.297,83	1.769,57	-22,99	0,974	- 16,74 %	- 11,27 %
UK	4.963,61	2.534,58	-48,93	0,919	5.968,31	3.135,66	-47,46	0,939	+ 20,24 %	+ 23,72 %
Ø	11.792,76	5.654,74	-48,43		11.609,92	4.944,03	-47,74		+ 2,40 %	- 4,20 %

RMSE = Root Mean Squared Error (Vorhersagefehler), Shapiro-Wilk = Test auf Normalverteilung der Residuen

Verringerung des Vorhersagefehlers durch Google Trends Daten um fast 50%
Neuronales Netz liefert bessere Ergebnisse als ARIMA bei Hinzunahme von Google-Trends-Daten

Tatsächliche Ankünfte vs. vorhergesagte Ankünfte (gestrichelte Linie) mit neuronalem Netz und Google-Trends-Daten



Kontakt

Prof. Dr. Wolfram Höpken
Leiter IDW / Forschungsgruppe Data Science
+49 751 501 9764
wolfram.hoepken@rwu.de