Sérgio Moro, Paulo Rita, (2016) "Forecasting tomorrow's tourist", Worldwide Hospitality and Tourism Themes, Vol. 8 Issue: 6, pp.643-653, doi: 10.1108/WHATT-09-2016-0046

Anticipating Tomorrow's Tourist

Abstract

Purpose – This study presents a very recent literature review on tourism demand forecasting, based on fifty relevant articles published between 2013 and June 2016.

Design/methodology/approach – For searching the literature, the fifty most relevant articles according to Google Scholar ranking were selected and collected. Then, each of the articles was scrutinized according to three main dimensions: the method or technique used for analyzing data; the location of the study, and the covered timeframe.

Findings – The most widely used modeling technique continues to be time series, confirming a trend identified prior to 2011. Nevertheless, artificial intelligence techniques, and most notably neural networks are clearly becoming more used in recent years for tourism forecasting. This is a relevant subject for journals related to other social sciences, such as Economics, and also tourism data constitutes an excellent source for developing novel modeling techniques.

Originality/value – The present literature review offers recent insights on tourism forecasting scientific literature, providing evidences on current trends and revealing interesting research gaps.

Keywords Tourism forecasting, tourism demand, tourists' behavior, modeling, tourism prediction

Paper type Literature review

1. Introduction

Forecasting tourism demand is a top rated subject for both researchers and practitioners, holding a profound impact on the tourism and hospitality industry. The importance of such theme has been recognized for a long time (e.g., Witt & Witt, 1995), remaining one of the most relevant trends in tourism research in the present days (Frechtling, 2012).

Seasonality has been intuitively considered one of the most influencing factors on tourism (Butler, 1998). Such intuitive knowledge has been solidly confirmed by numerous research studies (e.g., Alexandrova & Vladimirov, 2016). Another influencing factor is the knowledge possessed on today's tourists': besides the fact that a satisfied and pleased tourist is more likely to repeat its destination visit, the characteristics of the tourists of a specific destination may configure common profiles (Emel et al., 2007); therefore, it is necessary to know better today's tourists for anticipating tomorrow's tourists' profiles and flows. Other much lesser common factors that may influence tourism demand include unexpected events, whether in the form of natural disasters or human-driven crises (e.g., economic, political or military) (e.g., Veiga, 2014).

The development of information systems and technologies has created the possibility of harvesting large amounts of data for a vast number of businesses, including tourism and hospitality (O'Mahony et al., 2009). More recently, the advent of Big Data has pushed the quantity of data collected to a previously unforeseen level, resulting in Terabytes of information available for knowledge extraction and decision making (e.g., Jayawardena et al., 2013).

Data-based driven solutions for supporting decision making have arose at the same time data was being collected, since data itself is of no use unless it is translated into information and knowledge for making judged decisions (Turban et al., 2011). Statistical methods have always been used for analyzing data (e.g., Silvermann, 1986). Nevertheless, in the most recent decades, the rise of artificial intelligence and machine learning led to research in exploring data for decision support. Data mining involves concepts and methods with origin in several sciences, with a special emphasis on machine learning and statistics, for extracting predictive knowledge from raw data, aiming at supporting managerial decisions

(Moro et al., 2016b). Therefore, it is with no surprise that tourism research has also embarked on the exploration of hospitality and tourism data adopting data mining approaches, with tourists' forecasting being one of the most interesting problems studied, given its impact on this industry (Song et al., 2013).

The present research aims at analyzing academic literature since 2013 up to June 2016, for unfolding the most recent developments in forecasting tourism demand using data-driven approaches. A set of relevant journal articles is scrutinized under three critical dimensions: the method or technique used for analyzing data; the location of the study, and the timeframe considered. The results are then discussed in the light of previous findings, with the goal of unveiling research gaps for future studies.

The next section conducts a background literature review on tourism forecasting. Section 3 presents the methodology adopted, with Section 4 displaying the results. Finally, in the last section, the findings are discussed and conclusions are drawn.

2. Literature review

To understand tourists' behavior toward destinations constitutes a key asset for the tourism industry (Van Vuuren & Slabbert, 2012). Such advantage may lead to an anticipation of tourism demand, allowing managers to cope with the forecasted demand. Therefore, research in tourism forecasting is vast over the years. Modeling forecast requires past data for inferring future occurrences (Leeflang & Wittink, 2000). Information systems and technologies have provided the necessary means for building predictive models that may lead to a better understanding of tourism flows, helping managers to build expectations on tourists' arrivals.

Published studies and books on tourism forecasting can be traced back to the nineteen seventies and even in previous decades (e.g., Armstrong, 1972). However, a search on Google Scholar for results including both keywords "tourism" and "forecasting" in the decades after the 1970s reveals that such trend has observed a huge increase in the number of publications, especially in the three decades between 1980 and 2010, with each decade resulting in more than three times the hits from the previous one (Table 1). Moreover, although the 2010-2016 period encompasses only seven years, it holds already almost the number of hits from the 2000-2009 timeframe, revealing that such growing trend is likely to continue in the years to come. A similar analysis from Table 2 on “data mining” reveals an exponential growth in the same period (2000-2009). A subsequent search querying for both “data mining” and “tourism” & “forecasting” (Table 3) confirms the increase in the usage of data mining for tourism forecasting after 2000, with a total of 3,510 hits between 2000 and 2009. However, in the following period, 2010-2016, while encompassing only seven years, the number of hits is currently of 5,700, two thousand more than in the previous decade, unveiling that using data mining for tourism forecasting is a research yet far from being exhausted. In fact, the application of data mining to forecasting tourism is a real trend of this millennium.

A simple technique for forecasting is the no-change naïve model, which tends to outperform more sophisticated econometric and statistical techniques, according to Witt et al. (2003). The reasoning of the naïve approach is in assuming that the conditions that led to the current observed value will prevail in the future (Coons, 2015). Two widely adopted statistical time series methods for tourism forecasting are exponential smoothing and the seasonable autoregressive integrated moving average (ARIMA), as stated by Cho (2003). The former considers both the estimation for the present period and a random generated error in the present for a new estimation, while the latter is typically a procedure with three parameters referring to the terms autoregressive, differencing, and moving average for the seasonal part of the model. The same study also adopted artificial neural networks (NN), a technique bred from artificial intelligence tentative of mimicking human brain and neural processing (Asensio et al., 2014). The results from Cho (2003) show that NN outperform the statistically rooted techniques of exponential smoothing and ARIMA, with a special emphasis when the patterns for tourists’ arrival are less obvious. In fact, a characteristic of NN is precisely the ability to learn non-linear relations between the input features and the outcome to predict (Moro et al., 2016a). This happens because NN may include several hidden nodes with different weights contributing to compute the final outcome (Moro et al., 2014). The most recent machine learning techniques that have been applied to tourism forecasting include support vector machines (SVM), genetic algorithms, fuzzy systems and hybrid models. SVM were adopted by Pai et al. (2010) for implementing a seasonal regression model (hence, support vector regression) with good prediction results. In an SVM approach, the input belonging to the real space ($x \in \Re^M$) is transformed into a high m-dimensional feature space by using a nonlinear mapping that depends on a kernel (Smola & Schölkopf, 2004). Genetic algorithms are search heuristics rooted on the process of natural selection, while fuzzy systems are based on an analog logic considering continuous values between 0 and 1. Hybrid models use a combination of two or more techniques for trying to benefit from the best of each of them (Cortez, 2014). The study by Hadavandi et al. (2011) is an example of an hybrid solution based on both genetic algorithms and fuzzy systems for tourism forecasting, achieving interesting results when compared to other approaches (e.g., Markov based model).

Several literature reviews on tourism forecasting have been published from time to time. Two of them are particularly relevant to the present analysis, considering these identified also the methods used: the studies by Song & Li (2008), encompassing the years between 2000 and 2006; and by Peng et al. (2014), which analyzed the 1980-2011 period. The former included also qualitative (i.e., non-data-based) approaches, and only 9.1% of the studies included adopted approaches linked to artificial intelligence techniques. The latter study's sample of articles holds a slight increase in the usage of these techniques, to 10.2% of the articles. However, both studies conclude that while artificial intelligence models outperform the remaining in some specific cases, there is not a unanimous outperformance for these types of models. The present study aims at analyzing the most recent period of 2013 up to June 2016 for identifying if the novel developments in tourism forecasting have produced changes in previously identified trends.

3. Research methodology

This study conducts a literature review on tourism forecasting. Therefore, the first task is to gather relevant literature on the domain being analyzed for building a comprehensive body of knowledge (Moro et al., 2015). Google Scholar (GS) is one of the three most widely used search engines to query for academic publications (Harzing, 2013). It was chosen for querying its database for articles using the following search query:

"forecasting tourism" OR "forecast tourism" OR "forecasting tourists" OR "forecast tourists"

The filters applied included setting the timeframe period for publications from 2013 up to the present, and excluding patents. The number of hits (as queried in 21/June/2016) is 1,070. By taking advantage of Google Scholar ranking system, with roots on the renowned Google's search engine, the fifty most relevant articles published in journals were collected for a deeper analysis. Only articles from experiments using data-driven approaches for tourism forecasting i.e., including empirical analyses based on real data were considered.

Each of the collected articles was scrutinized for understanding which method for data analysis was adopted, what was the timeframe considered, and from which country were the data originated. Such three dimensions constituted the three vectors for the critical analysis of the literature collected. Nevertheless, other features from the articles were also analyzed, such as the authors' names and affiliations, the journals' titles, and the publishers.

4. Results

The fifty articles collected were published in a total of forty different journals (Table 4 shows those from which more than one article was selected), corresponding to eighteen different publishers (i.e., Table 5 for those publishers with more than one article selected). Such numbers prove that tourism forecasting is not entirely restricted to specific tourism literature, even though tourism gets the largest share, with 44% of articles; instead, the empirical studies found a vaster range of sciences, with a special emphasis on Management, Economics and Technology literature (Table 6). In fact, tourism forecasting is an important issue for management in general (e.g., Mamula, 2015) and for both origin and destination countries' economies (e.g., Chatziantoniou et al., 2016). Also, the fact that loads of data in tourism are currently available for exploring using cutting edge information technologies makes it an attractive subject for empirical studies to evaluate novel data modeling methods and applications (e.g., Liang, 2014). Nevertheless, it is a leading tourism journal such as Tourism Management that accommodates the highest number of publications focusing on tourism forecasting. From a publishers' point of view, Elsevier and Taylor & Francis are currently the two publishers clearly ahead in tourism forecasting journal article publications.

Set on the fifty articles analyzed, three main dimensions were analyzed: the timeframe which encompassed the data used within each study; the location from where the data was extracted; and the modeling techniques adopted.

Since all of the articles describe empirical data-driven experiments, and considering tourism has a strong seasonality influence but also other variables should be accounted for, it is interesting to understand from which years are the data collected for the experiments in order to assess if the periods are recent enough to include other factors such as the financial crisis emerged in 2008 and the effects risen afterwards. Figure 1 shows that most of the articles perform experiments based on data from the yearly 2000's, with few articles after 2013: just sixteen of them included 2013, nine 2014 and two 2015, despite the studies being very recent. One of the key aspects of data-driven knowledge discovery is the recency of data, especially considering that consumer behavior and, in particular, tourists' behavior changes over the years (Han et al., 2011). Therefore, while accounting for seasonality requires to include data from several years, using outdated data conceals the risk of negatively influencing models built on these data for forecasting tourism demand.

The continents that have received the most attention from researchers in these three and a half recent period are Asia, with 23 articles, and Europe, with 20 (Table 7). The dominance of Asian and European focused studies may be due to the appealing high growth rates shown by the former, i.e. materializing its market potential, and to the high market share of the latter. These reasons should also be coupled with higher levels of uncertainty occurring nowadays derived from globalization and thus asking for the application of rigorous scientific methods to predict the future and allow the industry to be more proactive.

The countries that contributed most for such figure are China, with eleven, and Spain, with nine. China and Spain are good examples of researched countries since they have high market shares in their continents but also portrait a high grow and a leading tourist destination, respectively. However, it should be stressed that most of the studies based on Spain used Catalonia data and were authored or co-authored by Claveria (e.g., Claveria et al., 2015), in a total of six of them published in such a narrow timeframe (from 2013 up to June 2016). In fact, this author has followed different approaches apparently based on the same data, proving the inherent potential of data-driven knowledge discovery experiments, which are hardly exhausted with the first discoveries. Although the sample of articles is not large (fifty articles), it is awkward to observe such a few number of studies using data from North America, just two. Also, none of the articles collected focused on the South American continent. These constitute interesting gaps for future research to fill. Also worth of noticing are the three papers found based on Kenyan data, all from different authors. Such result contrasts with the literature review by Peng et al. (2014), where only fourteen papers from the total of 2,584 were found for Africa.

Table 9 exhibits the modeling techniques adopted for the fifty articles considered. Exactly half of the studies adopted the seasonal autoregressive integrated moving average model (ARIMA). Also, two more adopted another time series method, the structural time series. Such a result is aligned with the findings by Peng et al. (2014), in which also more than half of the literature adopted time series' methods. This finding reveals that the strong influence of seasonality in tourism aligned with the adoption of large timeframes (Figure 1) continues to make of time series the most popular method for tourism forecasting. However, artificial intelligence techniques such as neural networks (used for thirteen times) and support vector machines (adopted six times) appear now as the second dominant method, as opposed to the review from Peng et al. (2014); such finding reveals a shift in the most recent years (i.e., after 2013) as opposed to the period before 2011, analyzed the paper of Peng et al. (2014), in which only around 17% adopted these advanced artificial intelligence techniques. Also, the percentage weight of the econometrics-based methods (e.g., Markov regime-switching model) has decreased, when compared to the cited study. It would be interesting to observe what future reserves for artificial intelligence applications to tourism forecasting.

5. Conclusions

Forecasting tourism demand is a rather old issue in which numerous researchers have focused on. Nevertheless, it is one of the most important problems, as the hospitality and tourism industry itself is under pressure for predicting future demand.

The present study aimed at offering a very recent literature review on data-based empirical studies for forecasting tourism demand, in an attempt to anticipate tomorrow’s tourist. By offering a review of the literature covering fifty relevant publications after 2013 up to June 2016, thus a very recent timeframe, the present article provides a summary on the most recent trends in this domain, identifying research gaps with eyes set on what the future holds regarding tourism forecasting.

The findings show that time series are still the most widely adopted method, confirming previous literature reviews on an earlier period (i.e., before 2011). In fact, the seasonality associated with tourism continues to justify such result. However, artificial intelligence techniques are already showing a substantial use in what concerns to modeling tourists’ behavior. In particular, artificial neural networks are remarkably recognized as an effective prediction method. Also, the literature found is not restricted to tourism journals, proving that tourism themes are also of interest for a broader range of social sciences (e.g., Economics), and that tourism data constitutes a valuable asset for assessing novel technologies for modeling. Most of the data adopted for the experiments is from destinations located in Asia or Europe, with a particular emphasis on China and Spain. It would be interesting to observe in the future if some of the trends found will prevail or which other trends will emerge on tourism forecasting.

References

Alexandrova, A., & Vladimirov, Y. L. (2016). Tourism clusters in Russia: what are their key features? The case of Vologda region. Worldwide Hospitality and Tourism Themes, 8(3).

Armstrong, G. W. G. (1972). International tourism: coming or going: the methodological problems of forecasting. Futures, 4(2), 115-125.

Asensio, J. M. L., Peralta, J., Arrabales, R., Bedia, M. G., Cortez, P., & Peña, A. L. (2014). Artificial Intelligence approaches for the generation and assessment of believable human-like behaviour in virtual characters. Expert Systems with Applications, 41(16), 7281-7290.

Butler, R. (1998). Seasonality in tourism: Issues and implications. The Tourist Review, 53(3), 18-24.

Chatziantoniou, I., Degiannakis, S., Eeckels, B., & Filis, G. (2016). Forecasting tourist arrivals using origin country macroeconomics. Applied Economics, 48(27), 2571-2585.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. Tourism Management, 24(3), 323-330.

Claveria, O., Monte, E., & Torra, S. (2015). A new forecasting approach for the hospitality industry. International Journal of Contemporary Hospitality Management, 27(7), 1520-1538.

Coons, J. W. (2015). Predicting Turning Points in the Interest Rate Cycle (RLE: Business Cycles) (Vol. 2). Routledge.

Cortez, P. (2014). Modern optimization with R. New York: Springer.

Emel, G. G., Taskin, Ç., & Akat, Ö. (2007). Profiling a domestic tourism market by means of association rule mining. Anatolia, 18(2), 334-342.

Frechtling, D. (2012). Forecasting tourism demand. Routledge.

Hadavandi, E., Ghanbari, A., Shahanaghi, K., & Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. Tourism Management, 32(5), 1196-1203.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Harzing, A. W. (2013). A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. Scientometrics, 94(3), 1057–1075.

Jayawardena, C., McMillan, D., Pantin, D., Taller, M., & Willie, P. (2013). Trends in the international hotel industry. Worldwide Hospitality and Tourism Themes, 5(2), 151-163.

Leeflang, P. S., & Wittink, D. R. (2000). Building models for marketing decisions: Past, present and future. International Journal of Research in Marketing, 17(2), 105-126.

Liang, Y. H. (2014). Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan. Computers & Industrial Engineering, 74, 111-119.

Mamula, M. (2015). Modelling and Forecasting International Tourism Demand-Evaluation of Forecasting Performance. International Journal of Business Administration, 6(3), 102.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, 62, 22-31.

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. Expert Systems with Applications, 42(3), 1314-1324.

Moro, S., Cortez, P., & Rita, P. (2016a). A framework for increasing the value of predictive data-driven models by enriching problem domain characterization with novel features. Neural Computing and Applications, In press.

Moro, S., Rita, P., & Vala, B. (2016b). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. Journal of Business Research, 69(9), 3341-3351.

O'Mahony, C., Ferreira, M., Fernandez-Palacios, Y., Cummins, V., & Haroun, R. (2009). Data availability and accessibility for sustainable tourism: An assessment involving different European coastal tourism destinations. Journal of Coastal Research, 1135-1139.

Pai, P. F., Lin, K. P., Lin, C. S., & Chang, P. T. (2010). Time series forecasting by a seasonal support vector regression model. Expert Systems with Applications, 37(6), 4261-4265.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. Tourism Management, 45, 181-193.

Silvermann, B. W. (1986). Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, 26.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—A review of recent research. Tourism Management, 29(2), 203-220.

Song, H., Gao, B. Z., & Lin, V. S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. International Journal of Forecasting, 29(2), 295-310.

Turban, E., Sharda, R., Delen, D., & Efraim, T. (2011). Decision support and business intelligence systems, 9th Edition. Pearson.

Van Vuuren, C., & Slabbert, E. (2012). Travel motivations and behaviour of tourists to a south african resort. Tourism & Management Studies, 295-304.

Veiga, L. (2014). Economic crisis and the image of Portugal as a tourist destination: the hospitality perspective. Worldwide Hospitality and Tourism Themes, 6(5), 475-479.

Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection. Journal of Travel Research, 42(2), 151-158.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. International Journal of Forecasting, 11(3), 447-475.

Tables

Table 1 - Results of querying Google Scholar with “tourism” & “forecasting”.

Timeframe	**<1980**	**1980-1989**	**1990-1999**	**2000-2009**	**2010-2016**
Nr. of hits	1,300	2,150	7,150	23,100	21,100

Table 2 - Results of querying Google Scholar with “data mining”.

Timeframe	**<1980**	**1980-1989**	**1990-1999**	**2000-2009**	**2010-2016**
Nr. of hits	1,400	2,300	19,200	864,000	373,000

Table 3 - Results of querying Google Scholar with “data mining” & “tourism” & “forecasting”.

Timeframe	**<1980**	**1980-1989**	**1990-1999**	**2000-2009**	**2010-2016**
Nr. of hits	9	12	134	3,510	5,700

Table 4 - Journals from which more than one article was selected.

Journal	**Nr. of articles**
Tourism Management	8
Applied Economics	4
Tourism Management Perspectives	3
International Journal of Tourism Research	2
Asia Pacific Journal of Tourism Research	2
Tourism Economics	2

Table 5 - Publishers from which more than one article was selected.

Publisher	**Nr. of articles**
Elsevier	16
Taylor & Francis	12
Wiley	3
Emerald	2
Scientific Research Publishing	2
SCIEDU	2
Ingenta connect	2

Table 6 – Research domain from journals from which articles were selected.

Research domain	**Nr. of articles**
Tourism	22
Management	7
Economics	7
Technology	7
Statistics and Applied Mathematics	3
Social Sciences	2
Generic	2

Table 7 – Continents from the destinations studied.

Continent	**Nr. of articles**
Asia	23
Europe	20
Africa	3
Oceania	3
Central America	2
North America	2

Table 8 – Countries from the destinations studied.

Countries	**Nr. of articles**
China	11
Spain	9
Taiwan	6
Croatia	4
Australia	3
France	2
Kenya	2
Turkey	2
UK	2
USA	2
Bahamas	1
Bhutan	1
Cayman Islands	1
Dominican Republic	1
El Salvador	1
Greece	1
Italy	1
Jamaica	1
Japan	1
Malaysia	1
Philippines	1
South Korea	1
St. Lucia	1
Tunisia	1

Table 9 – Modeling techniques adopted.

Modeling technique	**Nr. of articles**
ARIMA (seasonal autoregressive integrated moving average model)	25
Neural networks	13
Support vector machines	6
No-change naïve	5
Exponential smoothing	4
Genetic algorithms	4
Fuzzy systems	3
Multiple regression	3
Radial basis function	3
Time-varying parameter structural vector autoregression	3
Error correction model	2
Gaussian process	2
Hybrid models	2
Linear regression	2
Structural time series	2
Bai–Perron sequential search model	1
Complex networks	1
Delphi	1
Hierarchical cluster	1
Holt–Winters	1
K-means	1
Logistic regression	1
Markov regime-switching model	1
Monte Carlo simulation	1
Multivariate adaptive regression splines	1
Nearest neighbor	1
Singular spectrum analysis	1

Figures

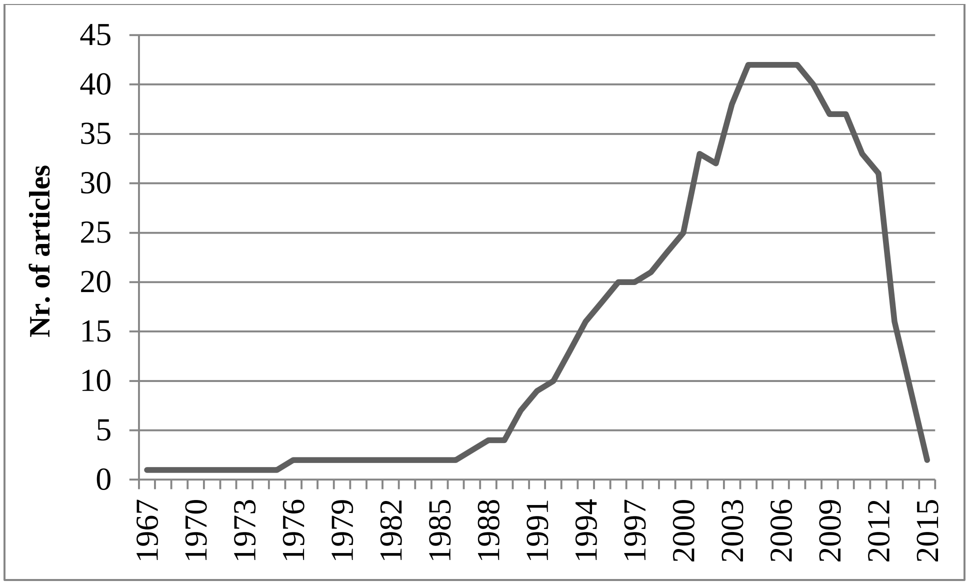

Figure 1 - Years included for the experiments.