Expert Systems with Applications 41 (2014) 1176–1188

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Ensemble methods for advanced skier days prediction

Michael A. King *, Alan S. Abrahams, Cliff T. Ragsdale

Pamplin College of Business, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

ARTICLE INFO

Keywords:
Ensemble learning
Data mining
Forecasting
Skier days

ABSTRACT

The tourism industry has long utilized statistical and time series analysis, as well as machine learning techniques to forecast leisure activity demand. However, there has been limited research and application of ensemble methods with respect to leisure demand prediction. The research presented in this paper appears to be the first to compare the predictive power of ensemble models developed from multiple linear regression (MLR), classification and regression trees (CART) and artificial neural networks (ANN), utilizing local, regional, and national data to model skier days. This research also concentrates on skier days prediction at a micro as opposed to a macro level where most of the tourism applications of machine learning techniques have occurred. While the ANN model accuracy improvements over the MLR and CART models were expected, the significant accuracy improvements attained by the ensemble models are notable. This research extends and generalizes previous ensemble methods research by developing new models for skier days prediction using data from a ski resort in the state of Utah, United States.

1. Introduction

Over the past two decades, consumer travel behavior and patterns have changed. The length of both the traditional family vacation and the associated planning horizon has significantly decreased (Luzadder, 2005; Montgomery, 2012; Zalatan, 1996). This trend is specifically evident with respect to snow skiing leisure activities at North American ski resorts. According to John Montgomery, managing director with Horwath HTL, a leading consulting firm in the hospitality industry, "if you booked a family ski trip 10 years ago, it was for a Saturday to Saturday block. Come hell or high water you were going." However, extended family ski vacations are now the rarity while shorter trips planned several days before departure have become quite common (Montgomery, 2012). This change is at least partially due to the Internet providing potential travelers with immediate travel decision information about snow conditions and last minute travel promotions.

Management at ski resorts must continue to adapt to these changing travel patterns by employing accurate demand forecasting techniques which, in turn, influence resort capacity planning operations. The tourism industry has long utilized statistical and time series analysis, as well as machine learning techniques to forecast leisure activity demand. However, there has been limited research and application of ensemble methods with respect to leisure demand prediction. This research uses local, regional, and national data to construct a skier days prediction model for a Utah-based ski resort. A skier day is the skiing industry standard metric for a single skier or snowboarder visit at one resort for any amount of time during one day (www.nsaa.org). We illustrate the predictive accuracy of forecasting models developed from multiple linear regression, classification and regression trees, and artificial neural networks techniques and demonstrate how prediction accuracies from these models may be increased by utilizing ensemble learning methods.

The 2009/2010 North American ski industry (NAICS 713920l) season counted nearly 60 million skier days, representing an approximate $16.305B industry. As illustrated in Fig. 1, this mature industry, is characterized by limited skier day growth, with only a 1.374% compounded annual growth rate over the last thirty years. As the 2007/2010 economic recession eroded consumer discretionary income (www.nsaa.org), competition within the skiing industry became even more aggressive. To sustain a long-term competitive advantage, individual ski resorts must provide superior experiences, high quality ancillary services (e.g., food services, lodging and sleigh rides) and year round outdoor activities all of which are predicated on accurate skier days and ancillary services estimates (Clifford, 2002).

The remainder of this paper is organized as follows. Section 2 provides a literature review and overview of ensemble methods. Section 3 discusses the unique contributions of this work. The method and research design implementations are described in Section 4. Section 5 provides a detailed discussion of the research results while Section 6 presents managerial implications and future directions.

* Corresponding author. Address: Department of Business Information Technology, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. Tel.: +1 540 231 6596.

E-mail address: michael.king@vt.edu (M.A. King).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.08.002

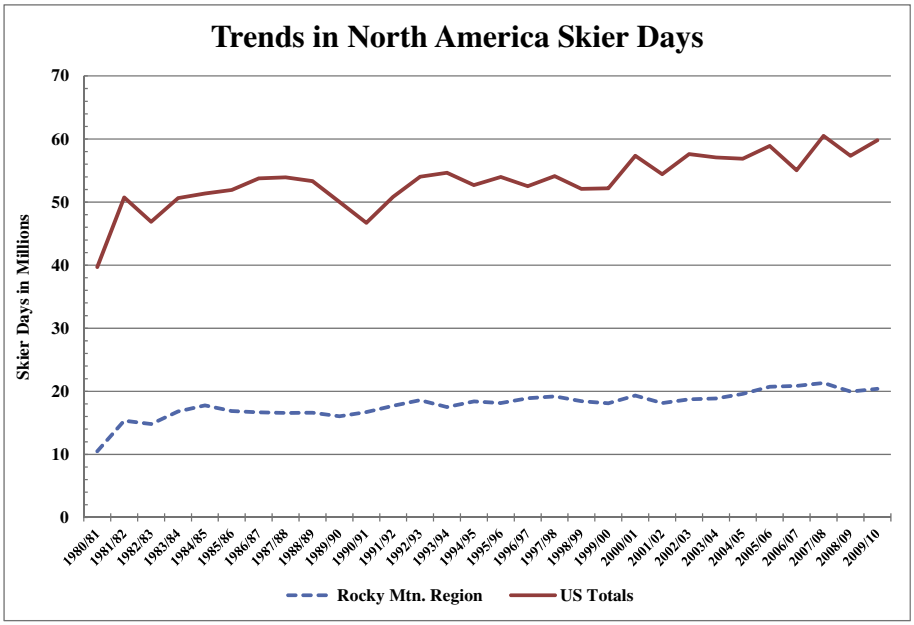

Fig. 1. Trends in North American skier days.

2. Literature review

The following section provides an overview of tourism forecasting research and the subsequent section presents background information on ensemble learning methods.

2.1. Related research

A significant theme of leisure or hospitality research published over the last two decades is the application of a wide array of forecasting techniques, such as time series analysis, econometric modeling, machine learning methods, and qualitative approaches for modeling tourism demand (Song & Li, 2008). Several comprehensive survey articles concentrating on tourism demand modeling have been published, each providing coverage of the forecasting method(s) utilized by the cited authors (Li, Song, et al., 2005; Lim, 1999; Song & Li, 2008). While there is limited research on tourism demand forecast combination or ensemble learning methods contained in these survey articles, Song, Witt, et al. (2009) as well as Oh and Morzuch (2005) make excellent cases for combining tourism demand forecasts. Song, et al. demonstrated that a single forecast formed by averaging a set of forecasts for inbound Hong Kong tourism will, by definition, be more accurate than the least accurate forecast, thus mitigating some forecasting risk. Oh and Morzuch (2005) provided a similar argument by illustrating how a forecast created by combining several time series forecasts for Singapore tourism outperformed the least accurate forecast and, in some situations, was more accurate than the most accurate individual forecast.

Table 1 is a concise list of highly cited tourism forecasting articles that apply MLR, CART or ANN modeling techniques and is indicative of the limited nature of current academic literature with a micro economic research focus. Also note that Table 1 contains only five prior articles related to skier days forecast, with two articles utilizing MLR and none applying ANN, CART, or ensemble techniques. This is also indicative of the limited availability of ski

Table 1
Related research.

Author	Forecasting method	Forecast target
Uysal and Roubi (1999)	Multiple regression, ANN	Tourist arrivals, Canadian inbound to US, aggregate
Law (2000)	ANN	Tourist arrivals, inbound to Taiwan, aggregate
Burger, Dohlan, et al. (2001)	ANN, moving average, multiple regression, ARIMA	Tourist arrivals, inbound to Durban South Africa, aggregate
Tan, McCahon, et al. (2002)	Multiple regression, economic models	Tourist arrivals, inbound to Indonesia, Malaysia, aggregate
Cho (2003)	Exponential smoothing, ARIMA, ANN	Tourist arrivals, inbound to Hong Kong, aggregate
Hu, Chen, et al. (2004)	Moving average, multiple regression, exponential smoothing	Restaurant customer arrivals, Las Vegas, US, local
Kon and Turner (2005)	ANN, exponential Smoothing, basic structural method	Tourist arrivals, inbound to Singapore, aggregate
Naude and Saayman (2005)	Multiple regression	Tourist arrivals, inbound to South Africa, aggregate
Pai and Hong (2005)	ANN, ARIMA, SVM	Tourist arrivals, inbound to Barbados, aggregate
Patsouratis, Frangouli, et al. (2005)	Multiple regression, economic models	Tourist arrivals, inbound to Greece, aggregate
Palmer, José Montaño, et al. (2006)	ANN	Travel tourism, inbound to Balearic Islands, aggregate
Chen (2011)	Linear, nonlinear statistical models	Tourist arrivals, outbound from Taiwan, aggregate
Shih, Nicholls, et al. (2009)	Multiple regression	Skier days, inbound to Michigan, US, local
Hamilton, Brown, et al. (2007)	Multiple regression, ARMAX	Skiers days for New England ski resorts
Riddington (2002)	Learning curve, time varying parameter	Skier days, outbound to Europe from UK, aggregate
Perdue (2002)	ANOVA, economic models	Skier days, inbound to Colorado, US, local
Pullman and Thompson (2002)	Multiple regression	Skier days, inbound to Utah, US, local
This research	Multiple regression, ANN, CART, ensembles	Skier days, inbound to Utah, US, local