Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2006 Proceedings

International Conference on Information Systems (ICIS)

December 2006

Predicting Airline Choices: A Decision Support Perspective and Alternative Approaches

Zan Huang
University of Wisconsin, Milwaukee

Huimin Zhao
University of Wisconsin

Yoshinori Suzuki
Iowa State University

Dan Zhu
Iowa State University

PREDICTING AIRLINE CHOICES: A DECISION SUPPORT PERSPECTIVE AND ALTERNATIVE APPROACHES

Design Science

Zan Huang
Department of Supply Chain and Information Systems
Pennsylvania State University
University Park, PA
zanhuang@psu.edu

Huimin Zhao
Sheldon B. Lubar School of Business
University of Wisconsin-Milwaukee
Milwaukee, WI
hzhao@uwm.edu

Yoshinori Suzuki
Department of Logistics, Operations and MIS
Iowa State University
Ames, IA
ysuzuki@iastate.edu

Dan Zhu
Department of Logistics, Operations and MIS
Iowa State University
Ames, IA
dzhu@iastate.edu

Abstract

The ability to predict the choices of prospective passengers allows airlines to alleviate the need for overbooking flights and subsequently bumping passengers, potentially leading to improved customer satisfaction. Past studies have typically focused on identifying the important factors that influence choice behaviors and applied discrete choice framework models to model passengers' airline choices. Typical discrete choice models rely on two major assumptions: the existence of a utility function that represents the preferences over a choice set and the linearity of the utility function with respect to attributes of alternatives and decision makers. These assumptions allow the discrete choice models to be easily interpreted, as each unit change of an input attribute can be directly translated into change in utility that eventually affects the optimal choice. However, these restrictive assumptions might impede the ability of typical discrete choice models to deliver operational accurate prediction and forecasts. In this paper, we focus on developing operational models that are intended for supporting the actual prediction decisions of airlines. We propose two alternative approaches, pairwise preference learning using classification techniques and ranking function learning using evolutionary computation. We have empirically compared these approaches against the standard discrete choice framework models and report some promising results in this paper.

Keywords: Airline choice prediction, preference learning, ranking function, classification, genetic programming

Introduction

Success in a global economy is linked to a company's ability to offer lower prices, better service, and greater choice than its competitors. In addition, a company must understand the buying behavior of its customers and be able to use this knowledge to predict how consumers make choices (Hardie et al. 1993, Manski 1997). This paper focuses on the airline industry, which has long been noticed for its limited range of choices and poor profitability. We propose and compare new approaches to help airlines predict consumer behavior, thus minimizing the need to overbook flights and subsequently bump passengers (Chatwin 1999).

One of the greatest challenges in aviation management is to understand why travelers select one airline over another (Peles et al. 2001). The optimal number of overbooking reservations for a given flight can be affected by knowing how passengers select an alternative carrier when their first-choice flight is not available (McGill & Van Ryzin 1993). Selection is influenced by a number of factors (e.g. price, convenience, quality of customer service, etc.), all of which must be critically examined in order to implement the appropriate revenue management tools and strategies (Proussaloglou & Koppelman 1995). Traditionally airline choices have been modeled by the discrete choice models (e.g. logit and probit) using the maximum likelihood estimation method (Suzuki 2006; Swait 2001). In these models, the probability of observing a particular alternative as the consumer choice is related to the utility derived from the consumer and alterative attributes through a logit or probit function. The utility is typically derived through linear combination of consumer and alterative attributes. This method maximizes the joint probability of observing actual choice patterns of the consumers included in the dataset. In other words, it minimizes the discrepancies between the actual choices and the predicted choices across all samples of observations. While this method produces models that can be nicely interpreted to explain consumer choice behaviors, its restrictive assumptions prevent it from fully exploiting the data patterns to give highly accurate predictions. The objectives of these studies have been to explain consumers' choice behaviors rather than to support actual prediction and forecasting decisions.

In this paper framed as a design science work (Hevner et al. 2004), we provide a different perspective to the airline choice prediction problem and focus on developing operational models that are intended for supporting the actual prediction decisions of airlines. We propose two alternative approaches: (1) pairwise preference learning using binary classification techniques and (2) ranking function learning using evolutionary computation, for predicting individual choice behavior. The pairwise preference learning approach drops the assumption regarding the existence of the utility function that fully represents the consumers' preferences over the choice set. A pairwise preference function is estimated for each pair of alternatives. The choice prediction is obtained by probabilistically consolidating the pairwise predictions for all pairs of alternatives involved in a choice set. The ranking function maintains the notion that a single function can fully represent the preferences but drops the linearity assumption and searches a much larger space of functional forms. We have conducted empirical evaluation using a real-world airline choice case study. We investigate the prediction of travelers' selections over nine major airlines and evaluate the performance of these alternatives and the traditional approach. Results from this research will shed light on intelligent decision making in airlines. Furthermore, these methods can be readily generalized to other consumer choice problems. The insights gained from this study can help businesses make effective decisions on pricing, yield management, and marketing strategies.

Background and Related Work

The modeling of individual human behavior in choice decisions has played important roles in predicting demand and market shares in transportation and marketing studies (Anderson et al. 1992, Ben-Akiva and Lerman 1985). *Discrete choice models* deal with choice decisions where the choice set contains a finite number of alternatives that can be explicitly listed. Choice of an airline alternative is a typical application of discrete choice models. In discrete choice modeling, two concepts of choice sets are typically considered, the *universal* choice set and the *reduced* choice set. The universal choice set contains all potential alternatives in the context of the application, while the reduced choice set is the subset of the universal choice set considered by a particular individual for a particular choice decision. The reduced choice set could be different across individual choice decisions. This feature makes discrete choice modeling distinct from the multi-class classification problem in the data mining literature. It would be problematic to formulate a discrete choice problem as a multi-class classification problem. For example, consider that a universal choice set of three airlines and the data on a set of travelers' choice decisions are available as the sample observations. One might try to formulate this problem as a ternary classification problem by defining the choice from the three airlines as the class variable and the airline-specific attributes for all three airlines and traveler attributes as the features. However, for a particular observation of choice decision with a reduced choice set of only two airlines as alternatives, the features corresponding to the attributes of the third airline are not well defined.

An important theoretical foundation of discrete choice models is the neoclassical economic theory on preferences and utility, which is built on a set of axiomatic assumptions on pairwise preferences. A decision maker i is assumed to be able to compare two alternatives, a and b, in the (reduced) choice set C_i using a preference operator PR_i. If a PR_i b, the decision maker i either prefers a to b, or is indifferent. We are interested in modeling individual decision maker choice behavior and focus on the individual-specific preference relations. The preference operator is supposed to have the following properties:

1. Reflexivity: $a \; PR_i \; a, \forall \; a \in C_i$.

2. Transitivity: $a \; PR_i \; b$ and $b \; PR_i \; c \Rightarrow a \; PR_i \; c, \forall \; a, b, c \in C_i$. (1)

3. Completeness: $a \; PR_i \; b$ or $b \; PR_i \; a, \forall \; a, b \in C_i$.

Because the choice set C_i is finite, the existence of an alterative that is preferred or indifferent to all others is guaranteed. That is

$$\exists \; a^{i*} \text{ s.t. } a^{i*} \; PR_i \; a, \forall \; a \in C_i. \quad (2)$$

The most important result of the utility theory is the existence of a function because of the three properties of the preference operator (Varian 1992)

$$\mathrm{U}_i : C \rightarrow R: a \rightarrow U(a) \quad (3)$$

such that

$$a \; PR_i \; b \Leftrightarrow U_i(a) \geq U_i(b), \forall \; a, b \in C_i.$$

Therefore, the alternative a^{i*} defined in (2) may be identified as

$$a^{i*} = \arg\max_{a \in C_i} U_i(a) .$$

The choice decision problem is then equivalent to assigning a *utility* value (by certain utility function U_i) to each alternative for each decision maker i, and selecting the alternative a^{i*} with the highest utility.

The concept of utility and the utility theory plays an important role in economics and derivative disciplines. Transforming the choice decision problem to the search of a utility function drastically reduces the complexity of the problem, especially for problems with large discrete choice sets or even continuous choice sets. However, the reflexivity, transitivity, and completeness assumptions present strong limitations for direct application of the preference and utility theory to practical discrete choice problems, mainly due to the failure to capture the uncertainty aspect of human choice decisions. Most discrete choice modeling literature is based on the *random utility model* (Manski 1977), which makes the neoclassical utility theory applicable in practical contexts by modeling utility as a random variable to reflect its uncertainty. The utility for decision maker i regarding alterative a is given by

$$U_{ia} = V_{ia} + \varepsilon_{ia} ,$$

where V_{ia} is the deterministic part of the utility and ε_{ia} is the stochastic part. With this probabilistic formulation, the choice modeling problem is then transformed in most studies to the search of a deterministic utility function that takes alterative attributes and/or decision maker attributes as input and assigns a utility to each alternative a for each decision maker i that best fits the observed choices and alterative and decision maker attributes.

The traditional econometric approach to modeling individuals' choice behaviors is to specify a decision maker's choice probability for each choice alternative by applying a logit-type function on the utility of each alternative to the decision maker and calibrating the model parameters by maximizing the log of likelihood function (typically using the Newton-type numerical optimization algorithms). Let U_{ia} be the decision maker i's utility of choosing alternative a. Then, using the standard multinomial logit model (Anderson et al. 1992, Ben-Akiva & Lerman 1985), the probability that an individual i chooses an alternative a from the choice set C_i is given by the form:

$$P_{ia} = \frac{\exp(U_{ia})}{\sum_{k \in C_i} \exp(U_{ik})} \quad (4)$$

$$U_{ia} = \sum_q \beta_q X_{qia} + \varepsilon_{ia} \ . \quad (5)$$

Here X_{qia} is the q^{th} choice-specific variable (attribute) of alternative a for decision maker i, β_q's are the unknown model parameters to be empirically derived, and ε_{ia} is the stochastic error component. Using (4) and (5), the likelihood function that we seek to maximize can be derived as follows:

$$L = \prod_{i=1}^{n} \prod_{a \in C_i} (P_{ia})^{y_{ia}} \quad (6)$$

where y_{ia} is a binary variable that is coded 1 if individual i chooses alternative a, and coded 0 otherwise. Typically, however, the model parameters are estimated by maximizing the log of (6), i.e., the log-likelihood function, because of the ease of estimation. Thus, parameter estimates are usually obtained by performing the following maximization procedure:

$$\arg\max_{\beta} \left[\sum_{i=1}^{n} \sum_{a \in C_i} y_{ia} \left(U_{ia} - \ln \sum_{k \in C} \exp(U_{ik}) \right) \right]. \quad (7)$$

The nice feature of the standard discrete choice models is that the impact of the factors (choice-specific variables) on the choice decision can be easily interpreted. Coefficient β_q in (7) can be directly interpreted as the amount of utility each unit of factor X_q brings to the decision maker. Most previous studies in marketing and transportation performed discrete choice modeling for the purpose of understanding the major factors that shape the consumer demand and the market structure. Being different from these previous studies, we argue that in addition to the explanatory understanding of the impact of factors that affect consumer choices, the ability to predict consumer choices alone can also be valuable. Given a highly accurate choice prediction model, analysts apply the model to massive data on consumer and choice characteristics to obtain estimates regarding the consumer demand on the alternatives given a choice set, and therefore the market share of each alternative. Detailed information on the impact of individual factors on choice behavior can also be obtained using a simulation approach. The ever-increasing computational capacity has made such analyses practical for real-time operations. We are no longer restricted to only those models that can be easily interpreted by human brain. In this study, we attempt to seek alternative approaches that can give a more accurate prediction of consumer choices than standard discrete choice models at the cost of sacrificing the interpretability of the predictive model. In the rest of the paper, we will focus on the airline choice problem as a case study.

Airline Case Study

A number of earlier studies have empirically modeled the airline choices of travelers using the multinomial logit model or the nested logit model, along with the standard maximum-likelihood estimation procedure (e.g. Morrison & Winston 1989; Nako 1992; Proussaloglou & Koppelman 1995; Yoo & Ashford 1996; Peles et al. 2001; Suzuki 2004, 2006). Upon examining factors that impact passengers' decision-making, these studies have generally concluded that travelers (whether business or leisure) tend to choose the airlines that offer lower airfares, more direct services, and frequent departures to the preferred destination. Some studies also claim that travelers tend to choose airlines in which they are "active" participants of a frequent-flyer or other rewards program. Suzuki (2004) examined the potential impact of bumping, flight delays, and baggage mishandling (lost, damaged, delayed, or pilfered) on future airline-choice decisions. The study concluded that service-failure experiences of this type are unlikely to have significant impact on passengers' future airline choices.

While the aforementioned studies have provided important normative implications about airline managers, as the discrete choice modeling literature in general as discussed in the previous section, they focused only on the *determinants* of airline-choice behaviors but not the *ability* of models to predict the actual choices of passengers. Like most econometrics analyses, these studies typically attempt to find the model that best fits the entire sample observations, and they relied on the interpretation of the resulting model parameters to answer the research questions, such as the impact of service-failure experiences on airline choice. Suzuki (2006) was probably the first study that separated the sample observations into calibration (training) and forecasting (testing), and then tested the accuracy of the standard multinomial-logit model in predicting the actual choices of passengers in the forecasting sample (i.e., the sample that was *not* used to calibrate the models). The results showed that the standard choice models can accurately predict up to 39% of the actual choices of the travelers in the forecasting sample. (Note that the average number of choice alternatives for each decision maker was about 17, so that the expected accuracy of a random selection is about 5.8%.) The results can hardly be generalized, however, since the study performed only one random extraction of the forecasting sample. In other words, the accuracy of logit-type models in forecasting the actual choices of air travelers has not been thoroughly examined using empirical data. Yet, as mentioned earlier, this is a critical airline management issue. Airline performance can be substantially improved by accurately predicting customers' airline-choice behaviors. We would like to systematically test the choice prediction accuracy of the standard discrete choice models and seek alternative approaches that can deliver higher accuracies.

Central to our investigation is data collected from the Des Moines (Iowa) International Airport (DSM) service area, which is defined by the Iowa Department of Transportation (IADOT) as the 14 Iowa counties within a one-hour ride (75 miles) of Des Moines. The total population of the area is approximately 700,000; the total number of households in the service area is about 272,000. The Des Moines International Airport is the only commercial airport with scheduled services in the area. Designed to meet the air travel needs of central Iowa residents, it is located in the southern part of the Des Moines metropolitan area. DSM does not serve as a hub airport for any airline. While most travelers in the immediate service area use DSM as the departure airport (trip initiating point), many travelers residing at the "edge" (near the borders) of the area use other facilities. The IADOT (1999) estimates that about 31% of the travelers residing in the service area use "out-of-region" airports on a regular basis.

According to the IADOT and DSM airport management, the out-of-region airports most often used by regional travelers are Kansas City International, Minneapolis-St. Paul International, and Omaha International. Kansas City and Omaha are served by Southwest Airlines, a major, low-cost carrier, and Minneapolis is a Northwest Airlines hub. The distances from Des Moines to Kansas City, Minneapolis, and Omaha are approximately 200 miles, 230 miles, and 150 miles, respectively. There are other airports within a 250-mile radius of Des Moines, but they offer limited services (limited airlines and service frequencies) and, therefore, are not widely used by travelers in the service area.

Since most travelers in our study initiated trips from the aforementioned airports (DSM, Kansas City, Minneapolis, and Omaha), we used these facilities to model travelers' choice behavior. The major airlines serving these four airports are American, America West, Continental, Delta, Northwest, Southwest, TWA (which is now American), United, and U.S. Airways. The summed passenger traffic of these nine airlines accounts for approximately 95%, 90%, 86%, and 90% of the overall airport traffic in DSM, Kansas City, Minneapolis, and Omaha, respectively. Given this condition, we treat these nine major airlines as the *universal choice set* of the travelers in our study area. The actual choice set of a traveler, however, is a subset of these nine airlines, as there is no route in which all the nine airlines provide scheduled services.

To collect the necessary data, passengers in the study area were surveyed in June 2001. The survey was pilot tested by using about 20 people, who are mainly businesspeople and graduate students. The final survey was created based on their comments and suggestions. Respondents were asked to provide information on the most recent domestic trip originating from DSM as well as any trip that originated from any of the aforementioned non-DSM (out-of-region) airports. Included in the survey were questions relative to date of trip, destination, and carrier chosen for each trip. A total of 529 responses were obtained. The total number of collected trip data (complete data) was 635, indicating that an average of 1.2 trip data was provided by each respondent (some respondents provided data on both the trip from DSM *and* that from other airport(s)). Of the collected 635 trip data, 104 were deleted from our sample for the following reasons: (1) inadequate data (chose airlines other than the nine major airlines); (2) represented trip occasions in which travelers had only one choice of airline (travelers did not "choose" airlines); or (3) trip dates were not recent enough for travelers to have accurate memories (more than two years old). After eliminating these "unusable" trip data, the total number of sample data reduced to 531. The demographics of the survey respondents are reported in Table 1. Table 2 describes the airline attributes. These variables were chosen based on the previous studies on airline choice modeling. We have included major factors that have been claimed to significantly affect airline choices.

Table 1. Demographics of Survey Respondents		
	Average of all travelers	Total count of travelers
Travel frequency (per year)	9.3	-
Traveler age	43.2	-
Participating freq. flyer programs	1.4	-
Business travelers (%)	36.8	-
Leisure travelers (%)	63.2	-
Freq. flyer program membership		
American	-	151
America West	-	31
Continental	-	20
Delta	-	70
Northwest	-	105
Southwest	-	10
TWA	-	132
United	-	214
US Air	-	27

Table 2. Airline Attributes	
Attribute	Description
Frequent flyer program (FFP_{ijt})	This is a binary variable indicating the FFP membership of traveler *i* for airline *j* at time *t*. To calculate this variable, we identify, for each traveler *i*, a set of airlines that satisfies the following conditions at trip occasion *t*: (1) traveler *i* has been a FFP member of the airline for two or more years, and (2) has used the airline at least once during the last two years. The FFP variable of traveler *i* for airline *j* at time *t* is coded 1 if *j* satisfies both of these conditions simultaneously, and is coded 0 otherwise. Notice that this procedure counts only the airlines for which a traveler is an "active" FFP member.
Airfare ($FARE_{ijt}$)	Since respondents had poor memories of actual airfares (especially for the airlines that were not chosen), we use the perceived airfares (a 0/1 variable) to capture the price effect. For each of the nine major airlines, respondents were asked to indicate whether they consider the airline to be a "low-fare carrier" (fares are lower than the industry average). If a traveler considers a particular airline to be a low-fare carrier, the airfare variable of this airline for this traveler is coded 1, and is coded 0 otherwise.
Service frequency ($FREQ_{ijt}$)	Service frequency measures the number of flight services offered by a carrier on a given route, and reflects the relative convenience of the carrier's flight schedules (Proussaloglou &Koppelman 1995). We use the Official Airline Guide (OAG Flight Guide North America, Dec. 2000 issue) to obtain this variable. The OAG Flight Guide lists scheduled airline services (including both direct and "legal" connection flights) of virtually all the origin-destination

	routes in the U.S. (see OAG Flight Guide for the definition of legal connection flights). By using this publication, we count, for each airline *j*, the number of scheduled flights offered per week in the origin-destination route flown by traveler *i* at time *t*.
Flight miles ($MILE_{ijt}$)	Flight miles reflect the on-flight trip length (or time) of travelers. For many trip occasions, travelers cannot use non-stop flights because of the unavailability of such services (this is particularly true in non-major airports such as DSM). On these occasions, travelers may fly into a hub airport that is actually in a different direction than their final destination. This type of inefficient routing (more trip miles) decreases traveler utility. We use the Data Bank 1A of U.S. Department of Transportation, a 10% sample data of all the U.S. domestic airline tickets, to obtain this variable. Using data from January 2000 to December 2000, we extract all of the sample-ticket data in the origin-destination route flown by traveler *i* at time *t*, and calculate the average flight miles (itinerary miles) for each airline.
Direct flight ($DIRECT_{ijt}$)	This variable reflects the availability of a carrier's direct flights in a given route. We use the average number of flight legs (average coupons) to measure this variable (the smaller the average flight legs the more direct the services). Using the Data Bank 1A for the time periods between January 2000 and December 2000, we extract all the sample-ticket data in the route flown by traveler *i* at time *t*, and calculate the average number of flight legs by airline.

Proposed Alternative Approaches

In this study, we offer a different perspective to the airline choice prediction problem, that is, to support the actual prediction decisions of airlines. In order to achieve the highest possible prediction accuracy based on the observed data, we drop the two major assumptions of the discrete choice model framework: the existence of the utility function that fully represents consumers' preference in (3) and the linear form of the utility function in (5). Not relying on these assumptions, the resulting models cannot be nicely interpreted as the discrete choice models, where each unit change of a particular input variable is associated with a constant change in utility of the alternative for the decision maker. On the other hand, our proposed approaches are able to better capture the data patterns present in the observed data and can be used to derive models that better fit the observed data and provide more accurate predictions on previously unseen cases.

Pairwise Preference Learning

The random utility models still rely on the three axiomatic assumptions regarding preferences, which justify the existence of the deterministic utility function. In practical choice decisions, one or more axiomatic assumptions in (1) might not hold (especially the transitivity assumption). Even if all assumptions hold for a particular domain, identifying the utility function might be an unnecessarily difficult formulation of the problem. Ultimately, the decision maker cares about the optimal choice, not necessarily about how much more utility is associated with this choice than alternatives. This detailed utility information is not necessarily relevant for choice decision, and the choice data are not suited for inferring such information. Instead of relying on these fundamental assumptions and searching for the utility function for the entire choice set, we propose to focus on the modeling of pairwise preferences regarding specific pairs of alternatives. For a particular pair of alternatives, a binary choice model is estimated, which may have a simpler functional form than a utility function. Furthermore, the unobserved heterogeneity among the alternatives can be fully captured by allowing different parameterization or even different functional forms for different pairs of alternatives. Especially with choice decision problems that have relatively small choice sets, such as the airline choice modeling problem in our context, estimating binary choice models for each pair of alternatives is feasible and practical.

Specifically, in our pairwise preference learning approach, a binary choice function

$$f(i,a,b)=\begin{cases}1, & \text{if } a\ PR_i\ b\\ 0, & \text{otherwise}\end{cases} \tag{8}$$

is estimated from the observed data. The binary choice function takes the attributes of decision maker i and attributes of the alternatives a and b as input. Domain knowledge can be leveraged to construct meaningful derived attributes for specific application contexts. For example, when alternatives have a common attribute X (e.g. the ticket price of the itinerary), one may decide that including a derived attribute $X_a - X_b$ (price difference) might be preferred to including X_a and X_b as two separate attributes into the model. Any binary classification algorithms can be directly applied to estimate this binary choice function. The choice of the specific algorithm (i.e., logistic regression, decision trees, neural network, and support vector machines) will depend on the trade-off between predictive performance and understandability of the resulting models.

Because no restrictive assumptions are imposed on the binary preference operators, the binary choice models learned from the data are capable of capturing more accurate patterns regarding preferences on the pair of alternatives involved. However, the set of binary choice models might not be compatible with each other, therefore producing preference relations that do not satisfy transitivity assumption. Take a three-alternative set $C = \{a, b, c\}$ as an example, for decision maker i three binary choice functions need to be estimated, $f(i,a,b)$, $f(i,b,c)$, and $f(i,a,c)$. As these binary choice functions are estimated separately, situations such as the following are likely to occur: $f(i,a,b) = 1$, $f(i,b,c) = 1$, and $f(i,a,c) = 0$, corresponding to preference relations $a\ PR_i\ b$, $b\ PR_i\ c$, *and* $c\ PR_i\ a$. Such incompatible preference relations are inconsistent with the transitivity assumption and will lead to failure in identifying the optimal choice from a choice set. A reconciliation mechanism that is able to incorporate uncertainty is needed. Unlike the random utility theory, where a stochastic term is added to the utility associated with each alterative, we assume the binary preference relation itself to be stochastic due to unobserved attributes of the decision maker and the alternatives as well as the general uncertainty of human preference. Thus, in our framework, we would prefer probabilistic binary classification algorithms such as logistic regression and naïve Bayes classifier that are able to output quantities relating to the probability of predicted preference relations:

$$g(i, a, b) = P(f(i, a, b) = 1) = P(a\ PR_i\ b). \tag{9}$$

For a given choice set $C = (a_1, \dots a_N)$, the probability for a_t to be the optimal choice (i.e., $a_t\ PR\ a,\ \forall\ a \in C$) for decision maker i would be determined by

$$\pi(a_t, i, C) = \prod_{a \in C, a \neq a_t} g(i, a_t, a)$$

assuming that the uncertainty of preference for different pairs of alternatives are independent from each other. The optimal choice from the choice set C for decision maker i is then identified as

$$a^{i*} \in \arg\max_{a \in C} \pi(a, i, C) \tag{10}$$

Note that when the binary classification algorithm does not produce probability interpretation of the choice outcomes, (9) reduces to (8) and (10) becomes a simple majority rule that treats the alternative with the largest number of dominated alternatives as the optimal choice.

Ranking Function Learning Using Evolutionary Computation

Another approach is to learn a function for ranking the airline choices. This ranking function is a (nonlinear) function of the attributes of an airline option and the decision maker, such that the airline option with the highest ranking is predicted as the one that will be chosen by the passenger. Observing that evolutionary computation techniques, such as genetic programming (GP), have been successfully used to learn ranking functions in other domains, such as personalized Web search (Fan et al. 2005), we apply GP in learning a ranking function for the airline choice prediction problem.

Evolutionary computation is a heuristic search approach that simulates natural evolution. Genetic algorithms (GA) and GP are two major evolutionary computation techniques and have been widely used in difficult optimization problems (Banzhaf et al. 1998). The basic idea is to evolve a population of individuals (candidate solutions) from generation to generation toward convergence to the best possible individuals. GA and GP follow essentially the same procedure but differ in their representations of individuals. While GA usually represents an individual with a fixed-length binary string, GP is much more flexible and is capable of representing an individual with any possible (variable-size) computer program. Thus, GP can learn nonlinear functions without fixed pre-defined forms. This allows the linearity assumption for the utility function (5) made by the discrete choice model framework to be relaxed such that a much larger space of function forms can be explored.

We adopt a tree structure in the genetic encoding of a ranking function for airline choices. There are two kinds of tree nodes, terminals and functions. Terminals are either attributes or constants. Functions include arithmetic operations, + (weighted addition), – (weighted subtraction), × (multiplication), and ⁄ (division). + and – take three parameters. × and ⁄ take two parameters. Figure 1 illustrates the GP representation of a sample ranking function.

GP usually employs three major genetic operations: reproduction, crossover, and mutation. Individuals are assessed by a goodness measure, referred to as fitness. For airline choice prediction, the fitness value of an individual is the accuracy of the corresponding ranking function. A selection mechanism is then used such that fitter individuals get better chances to survive and to produce descendants. Crossover exchanges some characteristics across multiple (usually two) selected individuals (parents) to generate new ones (children.) It combines the characteristics of parents by swapping a selected sub-tree of one parent with a selected sub-tree of the other. Mutation brings in innovation by changing some characteristics of a selected individual. It randomly selects a point in a tree and replaces the sub-tree starting at that point with a new randomly generated sub-tree.

We adopt the tournament selection method. When a tournament is held to select a parent, a small number of participants are randomly drawn from the current population and the winner, the fittest individual in the tournament, is selected. The selection mechanism also takes the size of a candidate solution into account. If two solutions have identical fitness values, the smaller solution is preferred. This preference of smaller solutions helps alleviate potential *overfitting* problems (i.e., a ranking function has unnecessarily high complexity and relatively low predictive power.) As previous studies have found that incorporating elitist selection tends to increase the convergence speed, we also incorporate this mechanism. It can be viewed as a special selection and reproduction mechanism, where the best individual reproduces itself.

(a) Ranking Function

FFP / (FARE + 0.75 (DIRECT + 0.71 MILE) (DIRECT − 0.27 FARE))

(b) GP Representation

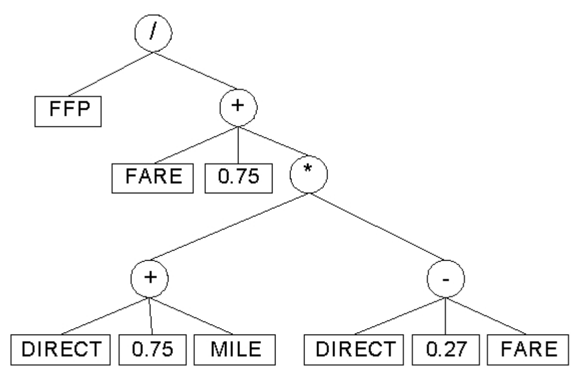

Figure 1. GP Representation of a Sample Ranking Function

Empirical Evaluation

We have empirically evaluated the proposed alternative methods using the Iowa dataset. For the pairwise preference learning approach, the independent variables we used for estimating the binary choice function $f(i, a, b)$ regarding airlines a and b for traveler i include the following: FFP_{ia}, FFP_{ib}, $FARE_{ia}$, $FARE_{ib}$, $MILE_{ia} - MILE_{ib}$, $DIRECT_{ia} - DIRECT_{ib}$, $BUSINESS_i$. $BUSINESS_i$ indicates whether the trip is for business or leisure purpose. Four commonly-used classification algorithms were used to construct the binary choice models, including the binary logistic regression, decision trees (Quinlan 1986), back-propagation neural networks (Lippmann 1987), and support vector machines (Cristianini & Shawe-Taylor 2000). For these classification algorithms, we leveraged the implementation in the WEKA package (Witten & Frank 2005).

The GP program was implemented in Java by extending the GPsys system (developed by Adil Qureshi). The parameter values used in the evaluation are listed in Table 3. The selection of parameter values was based on previous general guidelines. For example, small mutation rates and large crossover rates have been found to be generally effective.

Table 3. Genetic Programming Parameter Values

Parameter	Value
Population Size	200
Number of Generations	500
Crossover Rate	0.8
Mutation Rate	0.1
Tournament Size	7
Max Depth of Tree	20
Max Depth of Tree at Creation	10
Max Depth of Sub-tree at Mutation	5

We used 10-fold cross-validation (Kohavi 1995) to estimate the accuracy of each learned prediction model. Cross-validation randomly divides a training dataset into approximately equal-sized, stratified sub-sets called folds, and repeatedly uses each fold for performance testing while the other folds are used for training a model. The average of the testing performance measured over the runs is then used as an overall performance estimate. Empirical evaluation has shown that 10-fold cross-validation usually results in reasonably accurate estimates (Kohavi 1995). In addition, we repeated 10-fold cross-validation 20 times for each method to get more reliable estimates.

Table 4. Cross-validation Performance Comparison

Method	Average Accuracy	Standard Deviation
Multinomial Logit	0.4120	0.0369
Pairwise Preference Learning		
Binary Logit	0.5160	0.0647
Decision Tree	0.4651	0.0704
Neural Network	0.5117	0.0613
Support Vector Machines	0.5003	0.0657
Ranking Function Learning	0.5477	0.0633

Table 4 presents the average accuracy and standard deviation of the cross-validation testing of the traditional multinomial logit model and the proposed pairwise preference learning and ranking function methods. Table 5 presents the improvements of the proposed approaches over the traditional approach. Both proposed approaches significantly outperformed the traditional multinomial logit model for our data. Pairwise preference learning using decision tree achieved the modest improvement of 5.31%. Pairwise learning with binary logit model, neural network, and support vector machines increased prediction accuracy by 8.83%, 9.96%, and 10.39%, respectively.

The ranking function learning using genetic programming achieved the best prediction accuracy of 54.77%, a 13.57% improvement over the multinomial logit model.

Table 5. Improvement of Proposed Methods on Cross-validation Performance

Method	Improvement Over Multinomial Logit		
	Mean	95% Interval Lower Bound	95% Interval Upper Bound
Pairwise Preference Learning			
Binary Logit	0.1039	0.0865	0.1214
Decision Tree	0.0531	0.0356	0.0706
Neural Network	0.0996	0.0821	0.1171
Support Vector Machines	0.0883	0.0708	0.1058
Ranking Function Learning	0.1357	0.1182	0.1531

We also present the accuracy measures obtained by building model and testing the model using the entire dataset in Table 6. We refer to this measure as the within-sample accuracy, which corresponds to the typical goodness-of-fit measures in discrete choice model framework and other explanatory-drive data analysis. We observed that the multinomial logit model and ranking function method had similar within-sample accuracy and prediction accuracy. The pairwise preference learning methods, however, had significantly higher within-sample accuracy than the prediction accuracy. Both proposed approached achieved similar fit to the data that is significantly better than that of the traditional discrete choice model.

Table 6. Accuracy of the Complete Dataset

Method	Accuracy
Multinomial Logit	0.4181
Pairwise Preference Learning	
Binary Logit	0.5989
Decision Tree	0.6328
Neural Network	0.6704
Support Vector Machines	0.5612
Ranking Function Learning	0.5725

The performance comparison results confirmed our intuition that by relaxing the restrictive assumptions of the traditional discrete choice model framework, our proposed choice prediction approaches were able to provide significantly improved choice prediction accuracy. Our proposed methods might not necessarily provide concise explanation for the predictions. Nevertheless, the substantial improvement in prediction accuracy could still justify their use in practical airline choice decision support systems.

Conclusions and Future Research

In this paper, we have presented two new approaches to predicting airline choices of passengers. While past research has focused on identifying factors influencing customers' choice behaviors, we contribute a different perspective, developing useful decision support models. In order to achieve high prediction accuracy, we relaxed two important assumptions in the standard discrete choice modeling framework. Relaxing the assumption regarding the existence of a utility function that can fully represent the preferences over the choice set, we proposed a pairwise preference learning approach. Any binary classification algorithm can be applied under this approach to learn binary preference functions over pairs of alternatives. A preference reconciliation procedure is then used to combine preference predictions on all pairs of alternatives from a choice set to predict choice. Relaxing the linearity assumption of the utility function, we proposed the ranking function learning approach, which employs genetic programming to find a nonlinear ranking function that best fit the choice observations. To the best of our knowledge, this represents the first effort in designing convincing scientific models in supporting the actual prediction choice of airlines. Our empirical evaluation results show that the proposed approaches lead to significantly improved prediction performance compared to the standard discrete choice model using multinomial logistic regression. With our airline choice dataset, both proposed approaches successfully predicted the actual choices of more than 10% of the travelers in the testing set compared to the multinomial logic model. These results show that our proposed approaches have tremendous impact in providing design guidelines for the companies in the airline and other industries to implement effective marketing strategies and improve customer relationship management. Despite being less intuitive, these more accurate choice prediction models can be practically deployed in today's high-performance marketing information systems to support personalized marketing actions on the individual consumer level and market-level analysis of market share and demand structure.

This study opens up several avenues for future research. First, more comprehensive evaluation experiments using larger datasets are needed to further validate the findings from this study. Second, while we have focused on airline choice prediction, the proposed methods are general and can be applied to customer-product choice prediction problems in other industries. Third, future research may also study how multiple approaches can be effectively combined to make even better predictions.

References

Anderson, S. P., Palma, A. D., and Thisse, J.-F. *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, MA, 1992.

Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D. *Genetic Programming: An Introduction*, Morgan Kaufman, San Francisco, CA, 1998.

Ben-Akiva, M. and Lerman, S. R. *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA, 1985.

Chatwin, R. E. "Continuous-time airline overbooking with time dependent fares and refunds," *Transportation Science* (33), 1999, pp. 182-191.

Cristianini, N. and Shawe-Taylor, J. "An introduction to support vector machines," *Cambridge University Press*, 2000.

Fan, W., Gordon, M. D., and Pathak, P. "Genetic programming-based discovery of ranking functions for effective Web search," *Journal of Management Information Systems* (21:4), 2005, pp. 37–56.

Hardie, B. G., Johnson, E. J., and Fader, P. S. "Modeling loss aversion and reference dependence effects on brand choice," *Marketing Science* (12), 1993, pp. 378-394.

Hevner, A. R., March, S. T., Park, J., and Ram, S. "Design science in information systems research," *MIS Quarterly* (28:1), 2004, pp. 75-105.

Iowa Department of Transportation. *Intercity Passenger Service Study*, Ames, IA., 1999.

Kohavi, R. "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1143.

Lippmann, R. P., "An introduction to computing with neural networks," *IEEE Acoustics Speech and Signal Processing Magazine* (4:2), 1987, pp. 4-22.

Manski, C. "The structure of random utility models," *Theory and Decision* (8), 1977, pp. 229-254.

McGill, J. I. and Van Ryzin, G. J. "Revenue management: Research overview and prospects," *Transportation Science* (33), 1999, pp. 233-256.

Morrison, S. A. and Winston, C. "Enhancing the performance of the deregulated air transportation system," in *Brooking Papers on Economic Activity*: *Microeconomics,* 1989, pp. 61-112.

Nako, S. M. "Frequent flyer programs and business travelers: An empirical investigation," *The Logistics and Transportation Review* (28), 1992, pp. 395-414.

OAG Worldwide, *Official Airline Guide*, North American Ed., Oak Brook, IL., Dec. 2000.

Peles, E., Nikamp, P., and Rietveld, P. "Airport and airline choice in a multiple airport region: An empirical analysis for the San Francisco Bay area," *Regional Studies* (35), 2001, pp. 1-9.

Proussaloglou, K. and Koppelman, F. “Air carrier demand: An analysis of market share determinants,” *Transportation* (22), 1995, pp. 371-388.

Quinlan, J. R. “Induction of decision trees,” *Machine Learning* (1:1), 1986, pp. 81-106.

Suzuki, Y. “Modeling and testing the two-step decision process of travelers in airport and airline Choices,” *Transportation Research Part E*, 2006, forthcoming.

Suzuki, Y. “The impact of airline service failures on travelers’ carrier choice: A case study of central Iowa”, *Transportation Journal* (43:2), 2004, pp. 26-36.

U.S. Department of Transportation (various editions). *Data Bank 1A*, Washington, DC.

Varian, H. *Microeconomic Analysis*, Norton & Company, New York, NY., 1992.

Witten, I. H. and Frank, E. *Data mining: Practical Machine Learning Tools and Techniques,* Morgan Kaufmann, San Francisco, 2005.

Yoo, K. E. and Ashford, N. “Carrier choices of air passengers in pacific rim: Using comparative analysis and complementary interpretation of revealed preference and stated preference data,” *Transportation Research Record* (1562), 1996, pp. 1-7.