

마코프 체인을 이용한 항공운항의 정시성 예측

하만석, 남궁정일, 박수현
국민대학교 비즈니스IT전문대학원
e-mail:{msha, greenji, shpark21}@kookmin.ac.kr

Prediction of “Air-moving on Schedule” by Markov chain

Man-Seok Ha, Jung-II Namgung and Soo-Hyun Park
Graduate School of Business IT, Kookmin University

요 약

항공서비스에 있어서 여러 중요한 요소가 있지만 그 가운데 가장 중요한 요소로 운항의 정시성을 들 수 있다. 본 논문에서는 마코프 체인을 이용하여 항공사 및 공항의 정시성을 평가하고 예측하고자 하였다. 실험용 데이터로 선택한 미국 항공운항 데이터를 기상 악화에 따른 도착지연시간을 기준으로 정시성이 우수한 항공사를 제안하는 기능을 테스트하였다.

본 논문에서는 과거의 항공운항 데이터를 활용하여 미래의 운항 지연을 예측할 수 있도록 마코프 체인을 기반으로 한 예측 방법을 제시하였다.

1. 서론

고객들이 항공서비스를 선택하는데 있어서 중요한 요소로는 편리한 서비스, 안전, 정시도착과 같은 여러 요소가 있다. 그 중에서도 항공운항의 정시성은 매우 중요한 요소라고 할 수 있다.

공항과 항공사에서는 항공운항의 정시성을 향상시키기 위해 다각적으로 노력하고 있다.

본 논문에서는 항공 운항의 정시성을 사전에 정확히 예측할 수 있는 확률 모형을 제안하기 위하여 마코프 체인 프로세스를 사용하였다.

2. 관련연구

2.1. 항공운항의 정시성

항공운항의 정시성을 측정하고 분석하기 위하여 다양한 연구들이 시도되어 왔다.

Mayer 등은 항공운항시간을 최적화하기 위하여 다양한 통계 기법들을 활용하여 연구를 수행했다.[1]

Coy 등은 2001.9.11. 이전과 이후의 항공운항정시성에 대한 통계적인 분석을 시도하였다.[2]

조재희 등은 항공운항에 관련된 다양한 데이터를 수집하고 다양한 요소들에 대해 시각화를 시도하였다.[3]

백의영 등은 비행 도착/출발 시간을 최적화하기 위한 경로 분석을 위한 연구를 수행하였다.[4]

다양한 관련 연구들을 통해 통계적으로 정시도착 기준

으로 15분 이상 도착 시간이 지연될 경우 해당 항공서비스의 만족도는 급격히 하락하는 것으로 나타났다.[5,6]

항공운항 정시성은 날씨, 계절, 지역 등의 다양한 변수들로 인하여 정확한 예측이 어렵다는 점이 존재한다.

항공운항의 정시성을 측정하기 위한 다양한 연구들이 있지만 공항 및 항공사의 정시성을 평가하기 위하여 확률 모형을 적용시킨 연구는 찾아보기 어려웠다.

2.2. 마코프 체인

마코프 체인 프로세스는 어떠한 현상의 과거와 현재 상태를 근거로 미래의 상태를 추측하는 확률 모형이다. 과거에 발생한 데이터에 기반하여 가까운 미래를 매우 정확하게 예측할 수 있는 기법으로서 현재 많은 분야에 다양하게 응용되고 있다.[7]

본 논문에서 다루고자 하는 항공 운항의 정시성은 기상 상황 등의 복잡한 요인에 따라 매우 불특정한 패턴을 보이는 경우가 많으므로 미래를 예측하기가 매우 어려운 분야이다. 예를 들어 어제까지 정시에 도착한 비행기가 내일 갑작스런 기상의 변화로 인해 결항되거나 지연될 수 있다. 따라서 이러한 불특정한 추세를 가진 항공운항의 정시성 예측을 위해서는 마코프 체인 모델을 적용하는 것이 적절한 것으로 생각된다.

3. 정시성 예측 모델링

본 논문에서 사용한 데이터는 1987년부터 2008년까지 22년간 수집된 미국의 항공운항데이터로서 데이터의 사이즈는 11GB이며 총 123,534,969건의 자료가 저장되어 있다.

본 자료를 ASA 웹사이트로부터 csv 파일로 다운로드 받은 후 전체 자료를 HDFS에 업로드하였고 Hadoop과 Hive를 사용하여 통계 자료를 분석하였다.[8,9]

데이터 분석을 위해 자체적으로 개발한 자바 프로그램을 이용하여 기초적인 통계 자료를 수집하였다.

3.1 운항 지연 발생건수

대부분의 항공 운항 지연은 기상 상태의 악화로 인해 발생하고 있으므로 마코프 체인 모델링을 위해 기상 악화로 인한 운항지연 시간의 편차가 가장 큰 OH 항공사의 2007-2008년의 운항데이터를 활용하였다.

2007년 1월부터 2008년 12월까지 24개월간의 기상 악화로 인한 도착지연시간 통계는 <표 1>과 같다.

기상악화로 인한 도착지연시간 통계를 살펴보면 대체로 강수량이 많은 동절기와 하절기에 도착지연이 빈번하게 발생함을 볼 수 있다.

<표 1> 기상악화로 인한 도착지연시간 통계(분)

월	2007년	2008년
1	56,150	44,386
2	87,057	68,357
3	59,731	65,986
4	67,453	40,286
5	50,726	40,015
6	92,083	90,214
7	100,330	81,841
8	84,319	74,267
9	37,753	38,964
10	53,453	14,287
11	38,400	31,500
12	88,120	82,744

3.2 상태집합

평균 운항 지연 시간 예측을 위하여 <표 2>와 같이 임계값의 범위를 설정하였다.

<표 2> 임계값의 범위

S_1 : 0 - 50,000분
S_2 : 50,001 - 80,000분
S_3 : 80,001분 이상
$S = \{ S_1, S_2, S_3 \}$

2007년 1월부터 2008년 12월까지 평균 도착 지연시간을 정의한 임계값의 범위를 <표 3>에 매칭하여 상태를 나열하였다.

<표 3> 도착지연시간에 따른 상태집합 설정

월	2007년	상태	2008년	상태
1	56,150	S_2	44,386	S_1
2	87,057	S_3	68,357	S_2
3	59,731	S_2	65,986	S_2
4	67,453	S_2	40,286	S_1
5	50,726	S_2	40,015	S_1
6	92,083	S_3	90,214	S_3
7	100,330	S_3	81,841	S_3
8	84,319	S_3	74,267	S_2
9	37,753	S_1	38,964	S_1
10	53,453	S_2	14,287	S_1
11	38,400	S_1	31,500	S_1
12	88,120	S_3	82,744	S_3

나열된 상태로부터 각각의 상태 { S_1, S_2, S_3 }에서 다른 상태로의 전이 횟수를 구하고 이를 바탕으로 <표 4>와 같이 전이확률을 구했다.

<표 4> 상태 전이 행렬

다음 이번	S_1	S_2	S_3
S_1	0.38	0.25	0.38
S_2	0.38	0.38	0.25
S_3	0.22	0.22	0.56

3.3 초기 확률 산출 및 결과

2008년 12월의 운항 정시성 예측을 위한 초기 확률을 산출하기 위해 2008년 6월부터 11월까지의 6개월간의 운항 데이터를 사용하였다. 이에 따라서 산출된 초기 확률 값은 다음과 같다.

$$P = (0.5, 0.17, 0.33)$$

초기 확률과 상태전이행렬을 이용하여 다음과 같은 결과를 얻었다.

$$P = (0.34, 0.28, 0.37)$$

다음달의 평균 운항 지연 확률은 S_3 일 때 0.37로 가장 높게 나타났다.

4. 결론

본 논문에서는 마코프 체인을 이용하여 항공운항의 정시성 예측 모델링을 제안하였다. 미국 항공운항데이터를 제안한 모델에 적용하여 기상 악화로 인한 도착 지연에 대한 예측 확률을 산출하였다.

마코프 체인을 이용할 경우 가까운 미래는 예측할 수 있지만 먼 미래는 예측하기 어려운 단점이 있다. 향후 먼 미래의 예측과 좀더 다양하고 세부적인 임계값과 데이터를 이용하여 예측률을 향상시킬 필요가 있다. 또한 편리한 사용자 인터페이스 제공을 위해 JSP로 웹사이트를 개발할 예정이다.

ACKNOWLEDGMENT

이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임.(NRF-2013R1A1A2012461)

참고문헌

- [1] Christopher Mayer, Todd Sinai. 2002. Network Effects, Congestion externalities, and air traffic delays: or why all delays are not evil, Nber Working paper series.
- [2] Steven Coy. 2005. Management of airline arrival performance before and after September 11, 2001 in US domestic markets. Journal of air transport management. 11, pp. 209-230.
- [3] Jae-Hee Cho, De-Kui Li. 2011. Airport Punctuality Analysis Using Multi-Dimensional Visual Analysis Method. Journal of Korea IT Service.
- [4] Euiyoung Baik, Jaehee Cho. 2003. Analysis of Airline Network using Incheon and Narita Passenger Flight Origin-Destination Data. Journal of Information Technology Applications and Management.
- [5] Michael J. Mazzeo. 2003. Competition and Service Quality in the U.S. Airline Industry, Review of Industrial Organization. 22, pp. 275-296.
- [6] Jeffrey T. Prince, Daniel H. Simon. 2009. Multimarket Contact and Service Quality: Evidence From On-Time Performance in the U.S. Airline Industry,

Academy of Management Journal. Vol.52, No.2, pp. 336-354.

[7] 마코프 체인을 이용한 모바일 악성코드 예측 모델링 기법 연구, 김종민, 김민수, 김귀남, 융합보안 논문지 제14권 제4호, 2014.6.

[8] <http://stat-computing.org/dataexpo/2009>.

[9] Analysis of "Air-Moving on Schedule" Big Data Based On CRISP-DM Methodology, Man-Seok Ha, Jung-II Namgung, Soo-Hyun Park, ARPN Journal of Engineering and Applied Sciences, 2015.3