

Modeling Flight Delay and Cancellation

*Note: Sub-titles are not captured in Xplore and should not be used

Zhijie Wang

Abstract— This work uses flight record from Department of Transportation to predict whether a flight will be delayed by more than 30 minutes (either departure delay or arrival delay). The goal is to develop a policy to help a human traveler to make travel decisions under uncertainty, thus features used for modeling are limited to those observable at airport dashboard and some temporal context. Using Random Forest, Naïve Bayesian Network and Hidden Markov Models, we classified the data and achieved 85.63% accuracy with Random Forest Classifier.

Keywords—*Flight Delay, Random Forest, Bayesian Network, Hidden Markov Model.*

I. INTRODUCTION

The idea of working on flight delays came from personal work experience, as I happen to be a IT solution consultant and travel very frequently. For frequent travelers, traveling burnout is a constant struggle and last-minute meetings are bound to happen. There are two objectives in business travel decisions, to arrive on-time and minize travel stress. Arriving on time for meeting and relaxed can lead to better presentation and negotiation outcomes. The actual travel decisions are two step, booking and adjusment. Modeling flight delay is the first stem in solving the problem, and modeling airport congestion status is the second step.

In order to model flight delays, the input is a list of observable features of the flight at the airport information display (airline, departure airport, etc.) and some contextual information (distance, day of the week, etc.). Random Forests and Bayesian Networks are used to model flight delay probability and Hidden Markov model is used to model departure airport conditions. The flight delay probably is treated as a binay classification problem, Delay vs Non-Delay. We define the flight being delayed if one of the condition is met: canceled, departure delay > 30 min, arrival delay > 30 min. This problem has been studied in several research

works. Flight Delay and Cancellation data set is also publicly available on Kaggle with lots of analysis.

We define the airport traffic congestion status as a time series prediction problem with Hidden Markov model of 2 states, Congested and Non-Congested. This is important for making decisions on the fly when a flight is delayed. Each flight has its allocated queue time for runway take off, and due to safety concern, flight can't merge into the queue as easily as highway lane change. If a airport is busy or large amount of flight delayed already, it is hard to reschedule an delayed flight into the queue.

II. RELATED WORK

Several student projects in CS 229 class have tried to model flight delay problem and several research papaers truing to model flight delays. Most of school work leveraged rich feature set, such as terrorism incidents in the past week, weather data and adopt an one model fit all approach[1], [2], [3]. Researchers from NASA and FAA have performed in-depth analysis of flight delay causes and identified seasonality, origin airport and hour of departure contriute to flight delay [4], [5]. Some have adopted a statistical approach to model the the state and transition for a flight to be delayed [6].

III. DATASETS AND FEATURES

The data includes every single domestic flight from 2013 to 2017 September, with scheduled time, actual time, origin airport, depature airport, delay time, flight number, operating airline, taken from Bureau of Transportation Statistics [7].

A. Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

B. Units

- Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
- Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
- Do not mix complete spellings and abbreviations of units: “Wb/m²” or “webers per square meter”, not “webers/m²”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
- Use a zero before decimal points: “0.25”, not “.25”. Use “cm³”, not “cc”. (*bullet list*)

C. Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font).

$$a + b = \gamma \quad (1)$$

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

D. Some Common Mistakes

- The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

IV. USING THE TEMPLATE

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use

the scroll down window on the left of the MS Word Formatting toolbar.

A. Authors and Affiliations

The template is designed for, but not limited to, six authors. A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

1) *For papers with more than six authors:* Add author names horizontally, moving to a third row if needed for more than 8 authors.

B. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord “Format” pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.

basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

C. Figures and Tables

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^a Sample of a Table footnote. (*Table footnote*)Fig. 1. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

ACKNOWLEDGMENT (*Heading 5*)

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was

cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

- [1] Duperier, Sauvestre and Leaf, “CS229 Final Report: Modeling Flight Delays”, Computer Science Department, CS 229, Stanford University, Stanford, CA 2015
- [2] Lawson, D. And Cstillo W., “Predicting flight Delays. Technical report”, Computer Science Department, CS 229, Stanford University, Stanford, CA 2012
- [3] Bandyopadhyay, R. and Guerreror, R. “Predicting airline delays. Technical report”, Computer Science Department, CS 229, Stanford University, Stanford, CA, 2012.
- [4] Wang, Y., Sridhar, N., Jehlen, R., and Klein, A., “Modeling flight delays and cancellations at the national, regional and airport levels in the United States.”, In Eighth USA/Europe Traffic Management Research and Development Seminar (ATM 2009), NASA, Moffet Field, CA, USA and Washington DC, USA and Fairfax, VA, USA, 2009
- [5] E.R. Mueller and F.B. Catterji. “Analysis of aircraft arrival and departure delay characteristics.” In AIAA’s Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical, NASA, Moffet Field, CA, USA, 2002
- [6] Y. Tu, M. O. Ball, and W. Jank, “Estimating Flight Departure Delay Distributions - A Statistical Approach With Long-Term Trend and Short-Term Pattern,” *SSRN Electronic Journal*, 2005.
- [7] “On-Time Performance”, [Transtats.bts.gov](https://www.transtats.bts.gov/), 2017. [Online]. Available: https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=120&Link=0. [Accessed: 07- Dec- 2017].