



**HAL**  
open science

## Towards a more complete view of air transportation performance combining on-time performance and passenger sentiment

Philippe Monmousseau, Stéphane Puechmorel, Daniel Delahaye, Aude Marzuoli, Eric Féron

### ► To cite this version:

Philippe Monmousseau, Stéphane Puechmorel, Daniel Delahaye, Aude Marzuoli, Eric Féron. Towards a more complete view of air transportation performance combining on-time performance and passenger sentiment. ICRAAT 2020, 9th International Conference for Research in Air Transportation, Jun 2020, Tampa, United States. hal-02873440

**HAL Id: hal-02873440**

<https://enac.hal.science/hal-02873440v1>

Submitted on 19 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Towards a more complete view of air transportation performance combining on-time performance and passenger sentiment

Philippe Monmousseau, Stephane Puechmorel, Daniel Delahaye  
Optimization and Machine Learning Group  
ENAC, Université de Toulouse  
Toulouse, 31055, France  
philippe.monmousseau@enac.fr

Aude Marzuoli, Eric Feron  
School of Aerospace Engineering  
Georgia Institute of Technology  
Atlanta, Georgia 30332-0250, USA  
amarzuoli3, feron@gatech.edu

**Abstract**—This paper aims at presenting a novel approach to airline sentiment analysis processing using Twitter data. By transforming trained sentiment classifiers into regressors, the daily sentiment distribution obtained can be represented as a trimodal Gaussian Mixture leading to a simple but efficient classification algorithm. These classes can be considered as daily sentiment scores. This classification applied to passenger generated tweets and airline generated tweets for five major US airlines highlights major difference in experience between passengers and airlines. This methodology also confirms the existing gap between flight performance and passenger experience and the necessity of considering and implementing passenger-centric metrics.

**Keywords**—Air transportation system, Big Data, Sentiment analysis, Passenger-centric metric, Classification

## I. INTRODUCTION

The Air Transportation System is a complex interconnected system that carried more than 631 million passengers on domestic flights in the United States in 2010 according to the Bureau of Transportation Statistics (BTS) [1]. Passengers are at the core of this system, yet its performance is still essentially measured using flight-centric metrics.

Over the past few years, a shift from flight-centric metrics to passenger-centric metrics to evaluate the performance of the Air Transportation System both in the United States by the Federal Aviation Administration (FAA) with NextGen [2] and in Europe by the European Commission within its 2011 White Paper [3]. In the US, the Joint Planning and Development Office has proposed and tested metrics regarding NextGen's goals, but there are still metrics missing from the passenger's viewpoint, especially regarding door-to-door travel times [4]. Cook et al. [5] designed propagation-centric and passenger-centric performance metrics, and compare

them with existing flight-centric metrics. Precursor work was made by Marzuoli et al. in [6] and [7] using mobile phone data in order to analyze the performances of airports from the passengers' perspective. These studies validated the use of passenger-centric data to better assess the overall health of the Air Transportation System. However mobile phone data is proprietary data and is not often publicly available.

With more than 68 millions active users in the United States [8], Twitter is an important pool of public user-created data, where passengers can directly express their feelings with respect to a specific airline. Passenger sentiment analysis on Twitter is a promising approach to the creation of a passenger-centric metric, and many studies have focused on improving sentiment analysis since Pang et al. [9] thanks to the increase of available online reviews. Most works on sentiment analysis however focus on analyzing and improving the performance of classifiers such as Pak and Paroubek in [10] or Da Silva et al. in [11] and lack an application of the classifiers output. A thorough survey and classification of sentiment analysis methods was undertaken by Pang and Lee in [12].

Very few works actually propose an application of the classifiers output. Wang et al. [13] presented a framework to visualize real-time sentiment during political events in the United States using a crowd-sourced labeling method. Samonte et al. [14] proposed a sentiment analysis pipeline with some simple post analysis of the classification results and applied it to local airlines in the Philippines.

The contribution of this paper is to propose a method to extract the daily sentiment distributions of passengers in such a form that it can then be analyzed to evaluate the airlines performance with respect to passengers, paving the

way to a sentiment-based passenger-centric metric for the Air Transportation System.

The rest of the paper is structured as follows: Section II describes the methodology used to extract and process the daily sentiment distributions from the Twitter data. The analysis of the classification results is presented in Section III. Section IV concludes this study and discusses possible future steps.

## II. METHODOLOGY

### A. Data extraction

The Twitter dataset available for this study consists of all the tweets found using a basic search for each handle of 5 major US airlines, namely @united, @Delta, @AmericanAir, @SouthwestAir, @SpiritAirlines. Each entry consists of a timestamp, a user id, the content of the tweet and the handle used to retrieve the tweet. This dataset spans the entire period from January 1st 2018 to September 30th 2019.

It was then filtered to keep only tweets written in English using a two step process. The language of each tweet is initially taken as the own indicated by Twitter's API. The tweets labelled as "unknown" are then processed through the following language recognition algorithm and their language label are updated accordingly. Using the Natural Language Toolkit NLTK [15] and based on the work of Truica et al. [16], the number of common stop-words contained in a tweet is extracted for each available language in NLTK and the language with the highest count is selected. Due to the limited length of each tweet, a bias towards English has been introduced as well in the count ordering, i.e. if English and another language have the same count of common stop-words, English will have precedence.

### B. Sentiment analysis

A first step in sentiment analysis is to clean the documents analyzed, here the tweets. This cleaning process was already performed in [7] and [17] and consists of the following steps: Any reference to websites or pictures was replaced by a corresponding keyword. Every mention to another Twitter user within a tweet (@someone) as well as most emojis were similarly replaced. Since this database contains many replies from airlines to their customers, individual signatures of each agent were also replaced by a keyword. Dates and times were also generically replaced by keywords (e.g. "3rd Jan 2017" becomes "DATE" and "4pm" becomes "TIME"). The resulting text was then filtered from common stop-words and from the generic keywords used during the cleaning process.

Two different datasets were used to train three different classifiers each. The first dataset used was the labelled dataset used in a Kaggle competition [18]. The associated dictionary was created after removing words appearing in less than 20 tweets or in more than 75% of the full dataset. A second dataset and final cleaning process was generated based on the work of Read [19], also known as a distant supervised set used in many sentiment analysis models, with Go et al. [20] creating an impressive training set of 1,600,000 tweets. These tweets are from 2009 and are not specific to airline communication therefore this dataset was not considered here. Emoji filters were used to extract tweets from the initial dataset and automatically label them with a positive or negative sentiment according to Table I. The text cleaning process is also improved by merging negation words ("no", "not" and "never") with the word that follows it. The tokens used for the creation of the dictionary are the resulting bigrams, i.e. combinations of two words that follow each other in a tweet, with the same frequency filter as for the Kaggle dataset.

Table I: Emoji sentiment association

Category	Emojis
Positive	":)", "=:)", ":-)", ":-)", ":-)", ":-D", ":-D", "=:D"
Negative	":(", ":-(", "=-(", ":-@", ":-@", ":- "

For both methods, the scikit-learn library [21] was used to train the three classifiers considered, i.e. a random forest classifier, a naive Bayesian classifier and a logistic regressor. Once trained, the sentiment score used is the probability score of a tweet to be classified as positive, transforming in a way the classifiers into regressors. The final sentiment score is then the average of all six regressors and goes from 0 to 1, 0 indicating a negative tweet and 1 indicating a positive tweet.

### C. Classifying using a Gaussian Mixture representation

Once the sentiment score is calculated for each English tweet, it is possible to extract the underlying distribution per day and per airline, assuming a Gaussian Mixture model. Sentiment analysis usually classifies texts as *positive*, *negative* or *neutral*, therefore a trimodal Gaussian Mixture model was assumed for each day of tweets and for each considered airline. Using a Bayesian Gaussian Mixture model [22] enabled to consider uni- and bimodal cases if relevant. A day of tweets can therefore be represented in a 9 dimension vector  $(\mu_i, \sigma_i, \omega_i)_{i=1..3}$  such that its sentiment distribution

can be approximated as following the following probability function:

$$P = \sum_{i=1}^3 \omega_i \cdot \mathcal{N}(\mu_i, \sigma_i) \quad (1)$$

where  $\mathcal{N}(\mu, \sigma)$  is normal gaussian probability function of mean  $\mu$  and standard deviation  $\sigma$ .

A straight-forward classification method can then be derived based on these gaussian mixtures using the following algorithm. First the distributions are cleaned from their modes with a weight  $\omega_i$  smaller than 10% in order to make sure to capture all the uni- and bimodal distributions. Then the unimodal distributions are split into two classes whether their mean is greater or lower than 0.5. The bimodal distributions are split into three classes depending on the location of their means: both lower than 0.5, both higher than 0.5 or one on each side of 0.5. Trimodal distributions are simply split into two classes depending on the location of its most weighted peak with respect to 0.5. The classes are summarized in Table II.

Table II: Class description

Class	Distribution type	Categorization
0	Trimodal	$\mu_0 \leq 0.5$
1	Trimodal	$\mu_0 > 0.5$
2	Bimodal	$\mu_i \leq 0.5$ and $\mu_j \geq 0.5$
3	Bimodal	$\mu_i > 0.5$ and $\mu_j > 0.5$
4	Bimodal	$\mu_i < 0.5$ and $\mu_j < 0.5$
5	Unimodal	$\mu \leq 0.5$
6	Unimodal	$\mu > 0.5$

By construction, classes 3 and 6 can be clearly described as representing days with an overall *positive* mood, while classes 4 and 5 clearly represent days when a *negative* mood dominated. Class 2 can be seen as days where sentiments were polarized between *positive* and *negative*. Classes 0 and 1 would represent the normal situation where there are *positive*, *negative* and *neutral* tweets in various proportions without necessarily any one or two sentiments taking over.

#### D. Visualizing the sentiment space

Each vector  $(\mu_i, \sigma_i, \omega_i)_{i=1..3}$  represents a point in the space of trimodal Gaussian Mixture probability functions, space in which the Euclidian distance is not relevant. A useful distance in this space is the Wasserstein distance [], which can be understood as a transportation problem: The distance between two points  $P_1 (\mu_{1i}, \sigma_{1i}, \alpha_i)_{i=1..3}$  and  $P_2 (\mu_{2j}, \sigma_{2j}, \beta_j)_{j=1..3}$  in this space is equivalent to the minimal cost of moving the 'pile of earth'  $P_1$  (represented by

its probability density function) into the pile  $P_2$ . It amounts to solving the following Linear Programming problem:

$$\begin{aligned} & \min \sum_{i,j} x_{ij} \cdot d_{ij} \\ \text{s.t. } & \forall j, \sum_i x_{ij} = \beta_j \\ & \forall i, \sum_j x_{ij} = \alpha_i \\ & \forall (i,j), x_{ij} \geq 0 \end{aligned} \quad (2)$$

where  $d_{ij}$  represents the Fisher information distance between the two normal distributions  $\mathcal{N}(\mu_{1i}, \sigma_{1i})$  and  $\mathcal{N}(\mu_{2j}, \sigma_{2j})$ . The Fisher information distance  $d_F$  between two normal distributions  $\nu_1 \sim \mathcal{N}(\mu_1, \sigma_1)$  and  $\nu_2 \sim \mathcal{N}(\mu_2, \sigma_2)$  is calculated as follows:

$$\mathcal{F} = \sqrt{((\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2)((\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2)} \quad (3)$$

$$d_F(\nu_1, \nu_2) = \sqrt{2} \ln \left( \frac{\mathcal{F} + (\mu_1 - \mu_2)^2 + 2(\sigma_1^2 + \sigma_2^2)}{4\sigma_1\sigma_2} \right) \quad (4)$$

Once this Wasserstein distance is defined, it can be used along with the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique [23] in order to obtain a 2D representation of the space of trimodal Gaussian Mixture probability functions that preserves its implicit structure.

### III. RESULTS

The methodology presented in Section II-C was applied to two different sets of tweets extracted from the initial database. These sets were created based on the writer of the tweets, separating tweets coming from passengers versus tweets coming from the airline account.

#### A. Classification results

Counting the number of days related to each airline for every class yields some interesting insights regarding the composition of each class and the difference between passenger tweets and airline tweets. These airline distributions are plotted in Figure 1 & 2.

A first takeaway from the passenger perspective in Figure 1 is that none of the *positive* classes (i.e. classes 3 and 6) are represented during the considered period. One class gathers a total of 76.0% of airline-days: class 0. This indicates that passenger sentiment is usually split between the three modes, although with a bias towards a *negative* mood. The second largest class is class 4, the class with two *negative* modes, with 19.7%. The split between these two classes is similar for four of the five considered airlines with around 500 days in class 0 and 100 days or less in class 4, whereas American Airlines has an rather even split of 300 days for each class. This indicates that American Airlines passengers have the highest ratio of displeasing days, close

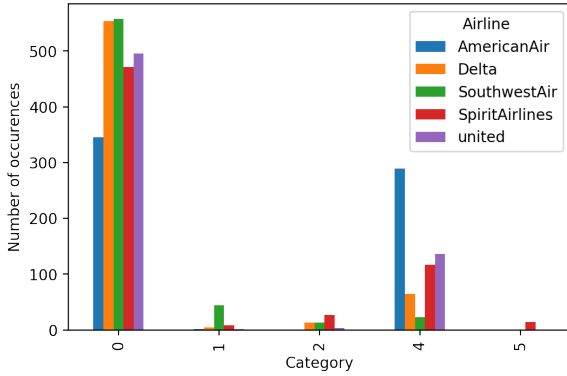


Figure 1: Airline distribution per class for passenger tweets

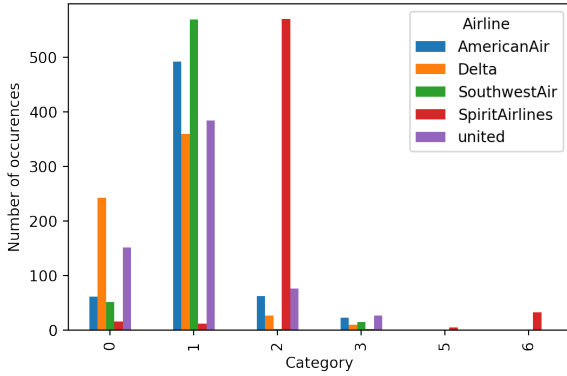


Figure 2: Airline distribution per cluster for company tweets

to 1/2. Spirit Airlines is the only airline with days in class 5, representing days where passengers are overall in a similar *negative* mood.

From an airline perspective, Figure 2 tells a different story: In clear contrast with the passenger class distribution, in the case of airline tweets, the *negative* classes (i.e. classes 4 and 5) are not or barely represented, with only five days in class 5 for Spirit Airlines. This indicates the opposition between how situations are experienced and expressed by passengers and how they are mitigated by the airline communications.

Regarding the main classes for airlines, class 1 concentrates 57.0% of airline-days, followed by class 2 with 23.1% and class 0 with 16.4%. Classes 1 and 2 have however opposite compositions: Spirit Airlines holds for around 75% of class 2 while being almost absent from class 1. This indicates that Spirit’s communication contains more tweets conveying a *negative* mood than the other airlines. As for the

passenger perspective, Spirit Airlines is also the only airline with days in class 5, which would indicate days when the airline twitter feed were essentially conveying a *negative mood*. Spirit Airlines is however the only airline with days in class 6, indicating that it is able to convey a *positive* mood on certain days.

The total number of airline-days per class is resumed in Table III. This representation highlights the quasi-orthogonality of the two perspectives: Classes with high representation for airlines are comparatively empty from a passenger perspective and *vice versa*.

Table III: Total number of airline-days per class

Class	0	1	2	3	4	5	6
Airline	521	1816	736	75	0	5	32
Passengers	2422	61	58	0	630	14	0

It is also possible to compare the daily class of these two perspectives day by day, in order to better visualize the opposition between passenger expressed experience and airline customer communication. Table IV shows the correspondances between airline classes and passenger classes. It is worth noting that for five days where the airlines are in class 6 (i.e. a unimodal *positive* mood), the passenger daily sentiment is in class 5 (i.e. a unimodal *negative* mood), another example of the opposite perception between airlines and passengers. Similarly, days when airlines are in class 3 (i.e. a bimodal *positive* mood) are perceived and expressed by passengers as belonging to mood classes with a *negative* bias (classes 4 and 0). On the opposite, days when airlines express a more *negative* mood in class 0 are also perceived as mainly negative by passengers with 80.8% in class 0 and 17.3% in class 4.

Table IV: Class correspondances between passenger and airline perspectives

		Passengers						
		0	1	2	3	4	5	6
Airlines	0	421	4	6	0	90	0	0
	1	1384	46	24	0	361	1	0
	2	536	8	26	0	159	7	0
	3	56	3	1	0	15	0	0
	5	3	0	0	0	1	1	0
	6	22	0	1	0	4	5	0

### B. Class visualization

The 2D representation of the daily sentiment distributions using the distance introduced in Section II-D along with a color code for their associated classes are shown in Figure 3 for passengers and in Figure 4 for airlines. In these

figures each point represents a day of tweets for one of the considered airlines.

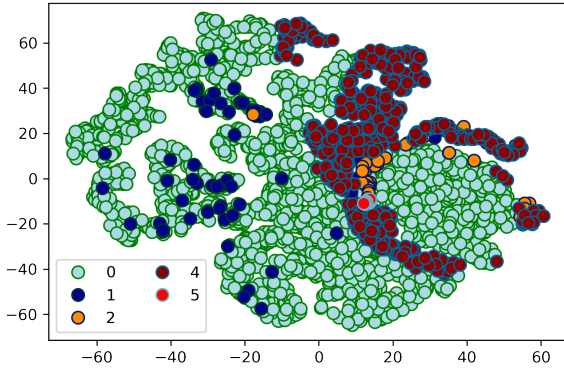


Figure 3: A 2D clustered representation of daily sentiment distribution of passenger tweets in a reduced dimension

In Figure 3, as expected from the previous results, the dominant class 0 spans the full space and encircles the other classes. Though the classes were not constructed by clustering, all classes are clearly separated from the others, with the exception of class 1 and a few outlying points of the other classes. The fact that class 1 is scattered within the class 0 cluster advocates toward a sensitive frontier between these two classes from a passenger perspective. Class 5 is concentrated in a small area in this representation space, whereas class 4 is more spread out. This indicates that the days with a distribution mood unimodal and *negative* (class 5), this distribution did not vary much from one day to another. In other words, it is sufficient to look at the mean of one of these days to have a good estimation of the other class 5 day means. Class 4 being more spread out, the most representative day of the class has to be found by another mean.

From an airline perspective, shown in Figure 4, the frontier between classes 0 and 1 is clearly defined. Further investigations should look into this frontier to know which tweet formulations should be avoided by airlines in order to stay in the better of the two classes, class 1. Class 2 is also clearly separated from classes 0 and 1, but is overlapped by the most positive classes, classes 3 and 6. Recalling that class 2 was dominated by Spirit Airlines, this overlapping suggests that the airline is aiming for a *positive* messaging but fall shorts of achieving it.

In order to find the day best representing each class, the Wasserstein distance can be used again to compute the central distribution of each class, i.e. the distribution that has

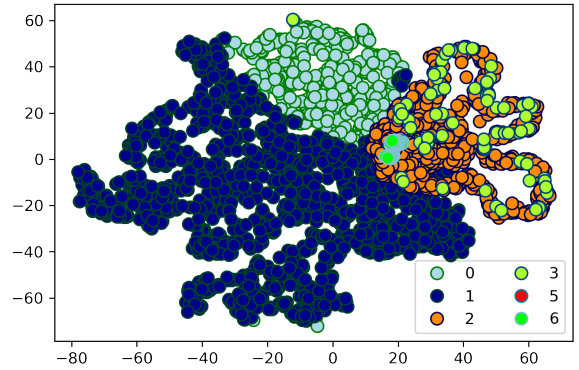


Figure 4: A 2D clustered representation of daily sentiment distribution of airline tweets in a reduced dimension

the smallest average distance to all the other points. These distributions are plotted in Figure 5 for the passenger dataset and in Figure 6 for the airline dataset. The distribution equation, with each parameter rounded at  $10^{-3}$ , is indicated on top of each subfigure for information.

Comparing the central distribution of a same class but from the two available perspectives draws the conclusion that though the class definition does not change, its representation varies drastically from one perspective to another. For example, the centroid of class 0 for passenger tweets has two modes on the *negative* side, whereas the the centroid for the airline tweets has two modes on the *positive* side, though the mode with the highest weight is the negative one by construction. Regarding the unimodal and *negative* class 5, it's mean is closer to the *positive* side for airlines than it is for passengers. Similarly, for class 1 the main mode mean is closer to the *negative* side for the passenger class centroid than for the airline one. The same can be said for class 2 and it's main positive mode.

### C. Passenger experience versus flight performance

Currently the air transportation system is essentially evaluated using flight-centric metrics such as flight delay, and lacks passenger-centric metrics. The class defined in this paper can help put in perspective the difference between these two approaches. Flight departure information over the considered period were extracted from the Bureau of Transportation Statistics (BTS) website. After analyzing and testing different distributions, the Student's T continuous distribution was kept as best fitting the daily delay distributions. Here a delay can be negative, meaning that the flight left earlier than the scheduled departure time. It is

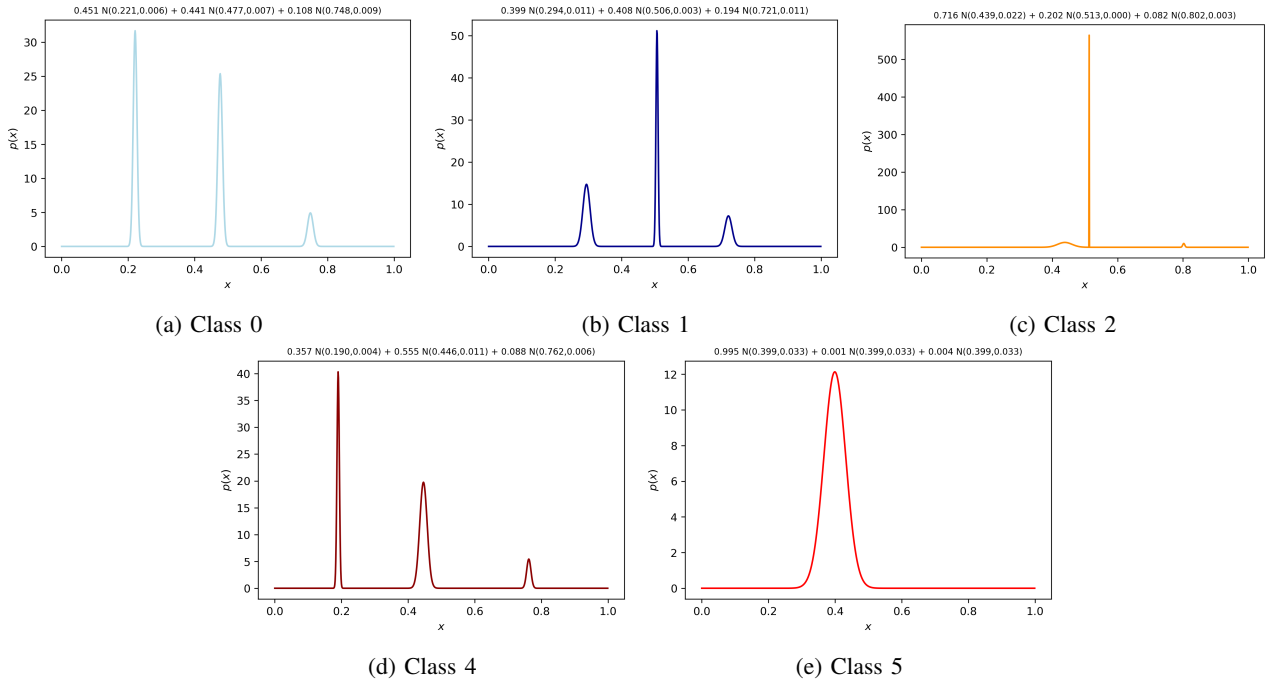


Figure 5: Gaussian mixture representation of the class centroids for passenger tweets

then possible to plot in a 2D plane the different days in the delay space using the location and scale parameters associated. The location parameter represents how much the distribution is shifted from 0 and the scale parameter gives an information on the width of the distribution. Figure 7 shows the airline daily delay distributions in this 2D plane along with a color code associating each day to its passenger sentiment class.

Looking at Figure 7, there are nine days with a location greater than ten minutes separated in two classes, with three days in the clearly *negative* class 4 and six days in the main class 0. This indicates that airlines managed to mitigate the effect of delays on passenger mood for six of these nine days. On the opposite spectrum, Figure 8 zooms into days with a delay location of less than ten minutes. What appears clearly here is that days with good flight performance, e.g. days with a negative average delay and a low scale are not necessarily experienced as positive for passengers. More precisely, all the class 5 days are located in this good flight performance zone, indicating that leaving early is not necessarily well perceived by passengers. Most of the class 4 days (89.5%) are days with a negative location and a scale lower than 5 minutes, highlighting the opposition between flight performance and passenger experience.

A similar representation is shown in Figure 8 using the airline sentiment class color code. The near totality (97.1%) of the two positive classes 3 and 6 concern days with a negative location and a scale lower than 5 minutes. This concentration suggests that important delays do have an impact on airline communication, in the sense that they cannot afford to express a mood too *positive* with respect to their customers.

#### IV. CONCLUSION

This paper aimed at presenting and leveraging a novel method for processing results from airline sentiment analysis applied to Twitter. Once sentiment classifiers are trained on well defined datasets, transforming them into regressors allows to obtain a Gaussian Mixture representation of the daily sentiment distribution. This representation can then be easily categorized in seven classes clearly defined and with an understandable signification. Separating and comparing the analysis of passenger generated tweets with airline generated tweets highlights the opposition in perception and experience of air travel between passengers and airlines. This opposition is even more visible when comparing these sentiment classes to the usual flight-centric metrics, since it

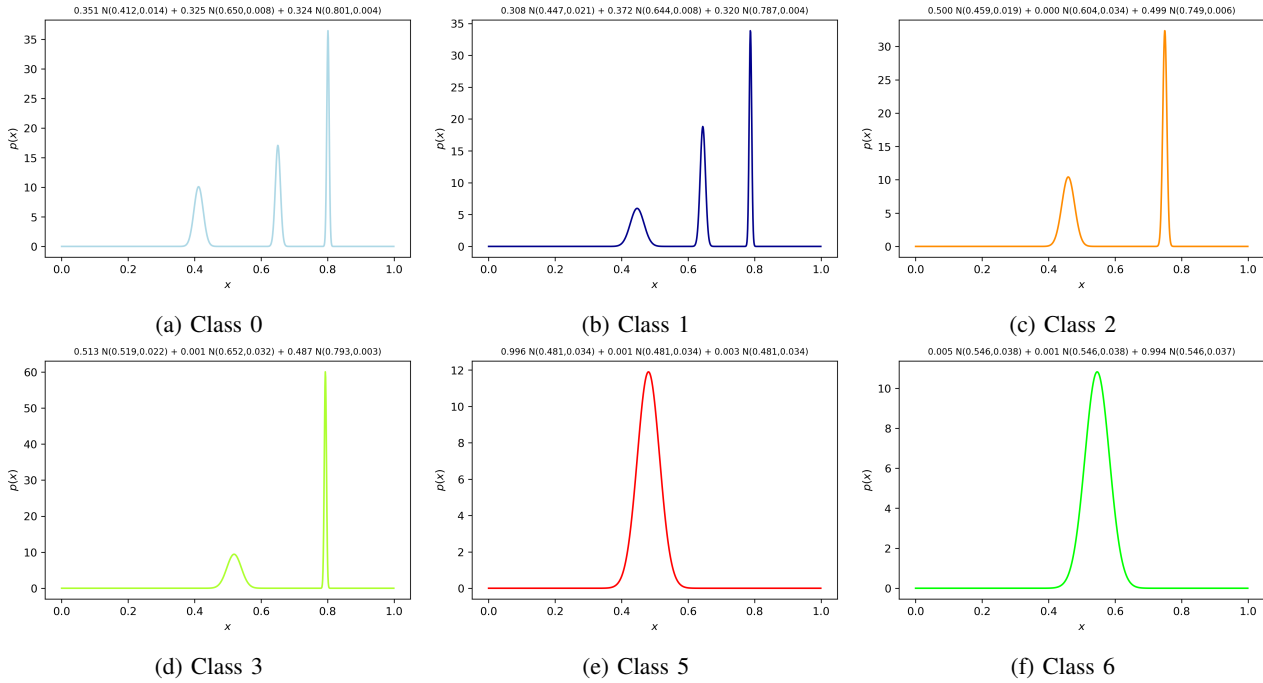


Figure 6: Gaussian mixture representation of the class centroids for company tweets

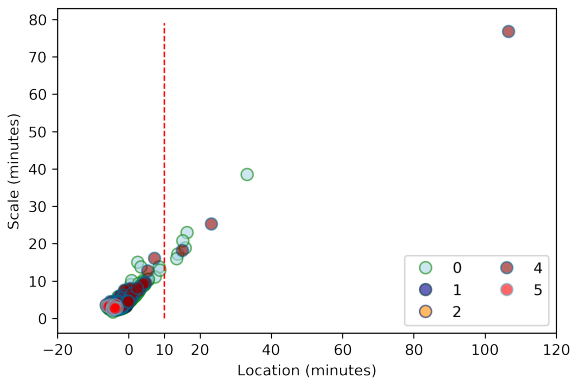


Figure 7: A 2D representation of the daily distributions of the amount of delay with the associated passenger sentiment class color code

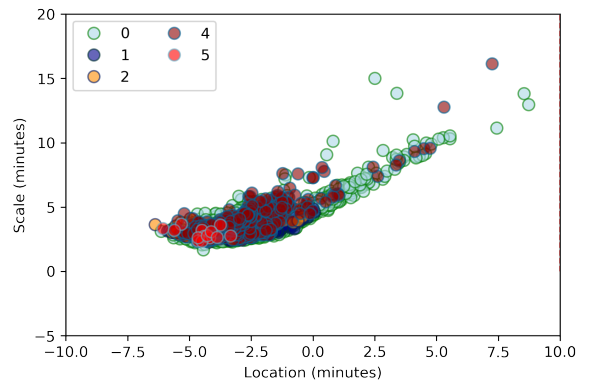


Figure 8: Zoom into the 2D representation of the daily distributions of the amount of delay with the associated passenger sentiment class color code

clearly shows that on time and early departures are not a sufficient condition for a *positive* passenger experience.

Future studies should focus on the frontier between the different sentiment class, in order to better understand when and how a day shifts between *positive* and *negative* classes, enabling airlines to prevent unwanted class shifts and thus

improving passenger experience.

#### ACKNOWLEDGMENT

The authors would like to thank Nikunj Oza from NASA-Ames, the BDAI team from Verizon, Palo Alto as well as the French government for their financial support.



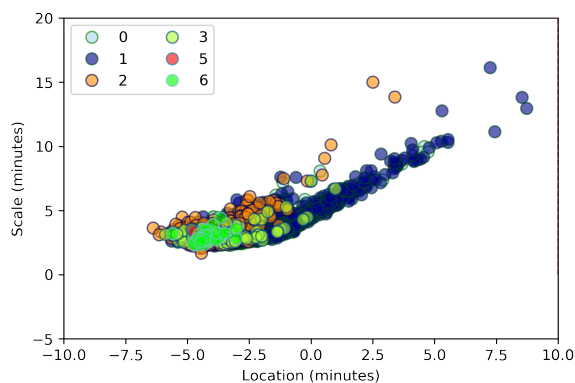


Figure 9: Zoom into the 2D representation of the daily distributions of the amount of delay with the associated airline sentiment class color code

#### REFERENCES

- [1] Bureau of Transportation Statistics, "Bureau of Transportation Statistics, About BTS." [Online]. Available: <http://www.rita.dot.gov/bts/about>
- [2] *NextGen Priorities - Joint Implementation Plan Update Including the Northeast Corridor*. Federal Aviation Administration, Oct. 2017.
- [3] E. Commission, Ed., *Flightpath 2050: Europe's Vision for Aviation ; Maintaining Global Leadership and Serving Society's Needs ; Report of the High-Level Group on Aviation Research*, ser. Policy / European Commission. Luxembourg: Publ. Off. of the Europ. Union, 2011, oCLC: 930887434.
- [4] Y. O. Gawdiak and T. Diana, "NextGen Metrics for the Joint Planning and Development Office," p. 10, 2011.
- [5] A. Cook, G. Tanner, S. Cristóbal, and M. Zanin, "Passenger-Oriented Enhanced Metrics," p. 8, 2012.
- [6] A. Marzuoli, E. Boidot, E. Feron, and A. Srivastava, "Implementing and validating air passenger-centric metrics using mobile phone data," *Journal of Aerospace Information Systems*, 2018.
- [7] A. Marzuoli, P. Monmousseau, and E. Feron, "Passenger-centric metrics for Air Transportation leveraging mobile phone and Twitter data," in *Data-Driven Intelligent Transportation Workshop - IEEE International Conference on Data Mining 2018*, Singapore, Nov. 2018.
- [8] Statista, "Monthly active Twitter users in the United States," <https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/>.
- [9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, vol. 10, pp. 79–86, May 2002.
- [10] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," *LREc*, vol. 10, no. 2010, pp. 1320–1326, May 2010.
- [11] N. F. da Silva, E. R. Hruschka, and E. R. Hruschka, "Tweet sentiment analysis with classifier ensembles," *Decision Support Systems*, vol. 66, pp. 170–179, Oct. 2014.
- [12] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, 2008.
- [13] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, 2008.
- [14] M. J. C. Samonte, J. M. R. Garcia, V. J. L. Lucero, and S. C. B. Santos, "Sentiment and opinion analysis on Twitter about local airlines," in *Proceedings of the 3rd International Conference on Communication and Information Processing - ICCIP '17*. Tokyo, Japan: ACM Press, 2017, pp. 415–422.
- [15] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," *arXiv:cs/0205028*, May 2002.
- [16] C.-O. Truica, J. Velcin, and A. Boicea, "Automatic Language Identification for Romance Languages Using Stop Words and Diacritics," in *2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*. Timisoara, Romania: IEEE, Sep. 2015, pp. 243–246.
- [17] P. Monmousseau, A. Marzuoli, E. Feron, and D. Delahaye, "Predicting and Analyzing US Air Traffic Delays using Passenger-centric Databases," in *Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)*, Vienna, Austria, 2019.
- [18] Kaggle, "Twitter US airline sentiment," <https://www.kaggle.com/crowdflower/twitter-airline-sentiment>.
- [19] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," in *Proceedings of the ACL Student Research Workshop on - ACL '05*. Ann Arbor, Michigan: Association for Computational Linguistics, 2005, p. 43.
- [20] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," *CS224N project report, Stanford*, vol. 1, 2009.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine Learning in Python," *Machine Learning in Python*, 2011.
- [22] S. Roberts, D. Husmeier, I. Rezek, and W. Penny, "Bayesian approaches to Gaussian mixture modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1133–1142, Nov./1998.
- [23] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of machine learning research*, vol. 9, pp. 2579–2605, 2008.