kuhn_2018_using_structural_topic_modeling_to_identify_latent_topics_and_trends_in_aviation_incident_reports

Year

2018

Author(s)

Kenneth D. Kuhn

Title

Using structural topic modeling to identify latent topics and trends in aviation incident reports

Venue

Transportation Research Part C

Topic labeling

Manual

Focus

Secondary

Type of contribution

Novel approach

Underlying technique

Manual labeling assisted by top words provided by multiple (3) metrics

Topic labeling parameters

Max labels per topic: 3
Nr of inspected words: 5

Label generation

Table 2 shows, for each topic found in the ASRS database, intuitive topic label(s) identified using expert judgement

The prevalence of the topic within the corpus, and the five highest ranked words when ordering by probability of occurrence conditional on topic (Prob), by lift (Lift), and by the FREX statistic (FREX). All three of the metrics described in the preceding paragraph are used, as there is no single correct statistic to use. The topic labels were selected based on the other data shown in Table 2

Some topics appear to cover distinct but related issues or systems and were therefore assigned more than one label. For example, one topic is linked to words describing passengers and other words describing aircraft cargo.

Table 2
Topics identified and labeled.

Topic Label	Topic Proportion	Criteria	Word 1	Word 2	Word 3	Word 4	Word 5
human factors	0.107	*Prob*	get	time	just	said	need
		Lift	deep	stupid	leadership	pride	imagine
		FREX	thing	something	think	know	realize
airspace	0.082	*Prob*	aircraft	control	traffic	sector	airspace
		Lift	apreq	datablock	dside	jurisdic	loa
		FREX	sector	carrier	train	dside	separate
ATC	0.078	*Prob*	arriv	clearance	atc	departure	cross
		Lift	domno	fms	mistook	sefr	trup
		FREX	fms	restrict	rnav	sid	waypoint
surface, routing	0.074	*Prob*	runway	tower	aircraft	taxi	clear
		Lift	backtaxi	ogg	quebec	foxtrot	papa
		FREX	taxiway	runway	hold	short	taxi
approach	0.071	*Prob*	approach	visual	final	land	runway
		Lift	phanom	mateo	glidepath	stable	loc
		FREX	approach	visual	tcas	sight	terrain
smoke, fire	0.064	*Prob*	land	emerg	airport	fire	declar
		Lift	smoke	midcabin	fire	fum	tailpipe
		FREX	fire	smoke	declare	emerg	divert
low-altitude traffic	0.061	*Prob*	airport	pilot	radio	flight	traffic
		Lift	civilian	foreflight	laser	tfr	tfrs
		FREX	helicopter	ctaf	class	tfr	pattern
fatigue	0.057	*Prob*	flight	plan	dispatch	crew	hour
		Lift	circadian	polar	nighter	fdp	awake
		FREX	fatigue	schedule	sleep	hour	duty
thrust, flaps	0.056	*Prob*	captain	flap	takeoff	first	officer
		Lift	dual	rto	asymmetric	thrust	autothrust
		FREX	flap	trim	thrust	autothrottle	lever

climb	0.052	*Prob*	altitude	climb	feet	level	atc
		Lift	barometric	altimeter	gyro	rime	compass
		FREX	climb	altitude	cloud	altimeter	feet
tug, brake	0.049	*Prob*	aircraft	brake	left	right	ramp
		Lift	tug	wand	rope	traction	towbar
		FREX	tug	wheel	brake	deice	snow
maintenance, fault	0.047	*Prob*	maintain	aircraft	system	control	mel
		Lift	veil	dmi	nef	mel	elac
		FREX	mel	inop	fault	maintain	breaker
passengers, cargo	0.047	*Prob*	flight	passenger	door	attend	captain
		Lift	clinic	csr	lightheaded	mail	monoxide
		FREX	agent	door	galley	cargo	bag
weather	0.041	*Prob*	aircraft	speed	wind	weather	turbulence
		Lift	pub	recat	vortex	chop	turbulence
		FREX	turbulence	wake	wind	encounter	moderate
landing gear, fuel	0.041	*Prob*	fuel	gear	land	engine	tank
		Lift	sputter	desert	enrich	pump	imbalance
		FREX	tank	gear	pump	fuel	hydraulic
mechanic	0.040	*Prob*	aircraft	mechanic	inspect	install	remove
		Lift	jobcard	rii	washer	bolt	bundle
		FREX	install	card	cable	repair	bolt
engine, oil, pressure	0.034	*Prob*	engine	pressure	cabin	start	oil
		Lift	buy	outflow	psi	bleed	pressure
		FREX	oil	bleed	pressure	mask	temperature

Motivation

“the goal of the topic labels is to succinctly describe the topics”

Topic modeling

STM

Topic modeling parameters

Nr of topics: 6 to 100

Nr. of topics

97

Label

One to three distinct manually assigned single or multi word labels

Label selection

\

Label quality evaluation

No formal evaluation, but:

Fig. 5 shows the start of the ASRS narratives that are most strongly linked to each topic. Fig. 5 also provides some intuitive evidence that topic modeling and topic label assignment were successful insofar as the narrative portions provided do appear, generally speaking, to match their topic labels but not other topic labels. The narrative linked to the climb topic appears to be an exception.

Topic: human factors Someone needs to look into our Air Carrier deeply before someone gets hurt. This is a serious safety issue. . . .
Topic: airspace I handed off Aircraft X to HTO sector. I still had communications with Aircraft X. I had radar and communications on Aircraft Y. I climbed a departure off of JFK to FL190. . .
Topic: ATC There is a discrepancy in altitudes between the charted arrival and the FMC database for the TELLR 1 STAR into DEN. The altitude at CREDE Intersection on the charted arrival is. . .
Topic: surface, routing After landing and departing Runway 2L; I taxied onto Taxiway Alpha; an active taxiway; prior to receiving clearance from Ground Control to do so. I saw no clear. . .
Topic: approach We were cleared for visual approach to Runway 6 from (approximately) a 6 mile right downwind. [We] flew base over antenna tower. We had excellent visibility and clearly had. . .
Topic: smoke, fire Avionics smoke ECAM Came on as we climbed toward FL250. We handled the ECAM and determined that there were no other indications of smoke; fire or instrument. . .
Topic: low-altitude traffic I was flying VFR from RIV and my iPad overheated and my panel mount GPS lost signal and I might have entered Class Bravo airspace while transitioning from PDZ. . .
Topic: fatigue While preplanning the release for Flight ABC ZZZ-ZZZ1 arrival time XA45Z; [Flight Planning Software] advised the Dispatcher with the following message: NO ALTERNATE. . .
Topic: thrust, flaps We received erroneous takeoff configuration warning when thrust levers advanced for takeoff. Immediately discontinued takeoff at approximately 30 KIAS and taxied clear of. . .
Topic: climb During descent; after leveling at ATC assigned 12;000 FT; ATC queried about our assigned altitude. Pilot not flying indicated 12;000 FT assigned and asked what ATC was showing and asked. . .
Topic: tug, brake I was pushing aircraft to park at hangar. Once chocked was going to pull the push tractor back and in to set it up for a crewmate to pull the pin out. As I thought I had my foot. . .
Topic: maintenance, fault We arrived at the airplane and noticed Fire Detection Loop A for the APU was deferred. I tested the squibs and noticed no squibs messages (1 and 2) for the APU. . .
Topic: passengers, cargo I noticed a strong metallic smell and at the same time I could hear the bell going off for the interphone. I picked up the interphone and I think it was Flight Attendant B. . .
Topic: weather During arrival into ATL we were descending thru approximately 10000 FT when we encountered wake turbulence. Over an approximate 5 minute period the wake induced three. . .
Topic: landing gear, fuel On May 2013 at approximately XA:00 hours; in good VFR weather; on base leg to the airport; the engine lost power. I switched the fuel selector to the opposite tank; but the. . .
Topic: mechanic I; as an Inspector; and Mechanic X were involved with the Left-hand Elevator Assembly that was removed for damage and a replacement Serviceable [Elevator]; according to the. . .
Topic: engine, oil, pressure Received 'ENG 2 OIL TEMP HI' in descent. Temperature started in the yellow then jumped to the red range. Temperature top [at] 225 degrees. Temperature was erratic. . .

Fig. 5. Example narratives for each topic.

Assessors

\

Domain

Paper: Aviation
Dataset: Aviation

Problem statement

The Aviation Safety Reporting System includes over a million confidential reports describing aviation safety incidents.
This article describes the application of structural topic modeling to Aviation Safety Reporting System data. The application identifies known issues. The method also reveals previously unreported connections.

Corpus

Origin: Aviation Safety Reporting System (ASRS)
Nr. of documents: 25,706
Details:

- ASRS records available to the public based on incidents that took place between January 2011 and December 2015.

Document

Narrative from an ASRS record.

Flight “Mission” (“Passenger” 68% of the time. 14% of the relevant incident reports focused on a “Personal” flight, 6% on a “Cargo/Freight” flight, and the remaining 12% on some other category of flight such as “Skydiving.”)

Reporting organization (“Air Carrier” for 58% of reports, as “Government” for 16%, as “Personal” for 12%, and as some other category of organization for the remaining 13% of reports).

Other structured data fields within ASRS records include some containing in formation on the month when the incident was reported, the locale (e.g., LGA.Airport, TUL.TRACON),

meteorological conditions, and phase of flight (e.g., Climb, Descent).

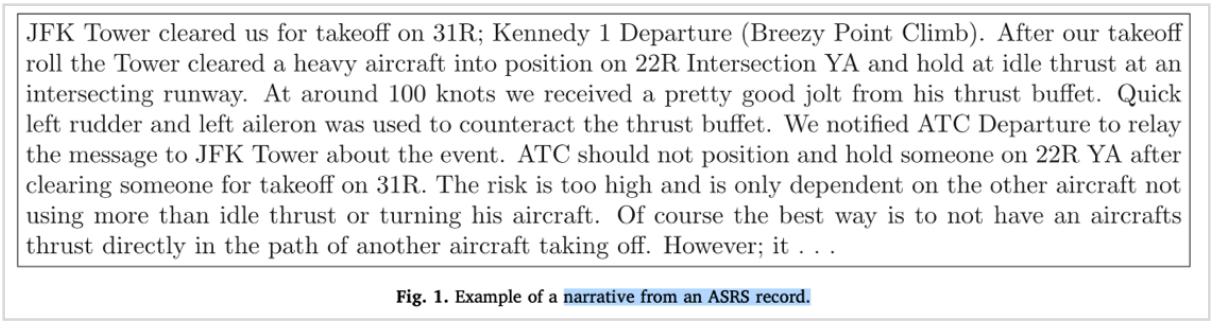
JFK Tower cleared us for takeoff on 31R; Kennedy 1 Departure (Breezy Point Climb). After our takeoff roll the Tower cleared a heavy aircraft into position on 22R Intersection YA and hold at idle thrust at an intersecting runway. At around 100 knots we received a pretty good jolt from his thrust buffet. Quick left rudder and left aileron was used to counteract the thrust buffet. We notified ATC Departure to relay the message to JFK Tower about the event. ATC should not position and hold someone on 22R YA after clearing someone for takeoff on 31R. The risk is too high and is only dependent on the other aircraft not using more than idle thrust or turning his aircraft. Of course the best way is to not have an aircrafts thrust directly in the path of another aircraft taking off. However; it . . .

Fig. 1. Example of a narrative from an ASRS record.

Pre-processing

Punctuation, whitespace, and stop words were removed from the corpus.

```
@article{kuhn_2018_using_structural_topic_modeling_to_identify_latent_topics_and_trends_in_aviation_incident_reports,
  abstract = {The Aviation Safety Reporting System includes over a million confidential reports describing aviation safety incidents. Natural language processing techniques allow for relatively rapid and largely automated analysis of large collections of text data. Interpretation of the results and further investigations by subject matter experts can produce meaningful results. This explains the many commercial and academic applications of natural language processing to aviation safety reports. Relatively few published articles have, however, employed topic modeling, an approach that can identify latent structure within a corpus of documents. Topic modeling is more flexible and relies less on subject matter experts than alternative document categorization and clustering methods. It can, for example, uncover any number of topics hidden in a set of incident reports that have been, or would be, assigned to the same category when using labels and methods applied in earlier research. This article describes the application of structural topic modeling to Aviation Safety Reporting System data. The application identifies known issues. The method also reveals previously unreported connections. Sample results reported here highlight fuel pump, tank, and landing gear issues and the relative insignificance of smoke and fire issues for private aircraft. The results also reveal the prominence of the Quiet Bridge Visual and Tip Toe Visual approach
```

```
paths at San Francisco International Airport in safety incident reports. These results would, ideally, be verified by subject matter experts before being used to set priorities when planning future safety studies.},
  author = {Kenneth D. Kuhn},
  date-added = {2023-04-03 14:27:38 +0200},
  date-modified = {2023-04-03 14:27:38 +0200},
  doi = {https://doi.org/10.1016/j.trc.2017.12.018},
  issn = {0968-090X},
  journal = {Transportation Research Part C: Emerging Technologies},
  keywords = {Aviation, Aviation safety, Natural language processing, Topic modeling, Text mining, Aviation safety reporting system},
  pages = {105-122},
  title = {Using structural topic modeling to identify latent topics and trends in aviation incident reports},
  url = {https://www.sciencedirect.com/science/article/pii/S0968090X17303881},
  volume = {87},
  year = {2018}}
```

#Thesis/Papers/BS