CLASSIFICATION OF ATC UTTERANCE TRANSCRIPTS VIA WARM-START NON-NEGATIVE MATRIX FACTORIZATION

Marco Enriquez
menriquez@mitre.org
MITRE CAASD

2014 ICRAT Conference
Istanbul, Turkey

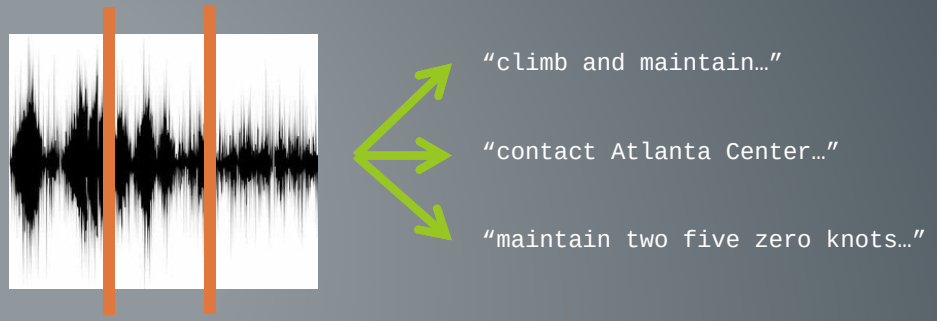
VOICE DATA AND AVIATION ANALYSIS

- *ATC voice data is often under-utilized in various aviation-related research*
- *Challenges:*
  - *Facility access and Permissions*
  - *Storage and Data Transfer*
  - *Quality of recording and Transcription*
- *Rewards:*
  - *Gain unique insight to operations*
    - *Advanced procedure analysis*
  - *Enhance trend/outlier activities*
- *We propose a methodology to automatically analyze ATC voice data for various aviation analyses*

Workflow

'Approved for Public Release; Distribution Unlimited. Case Number 14-1940'

1. Voice Data is stored to disk and is collected from partner facilities



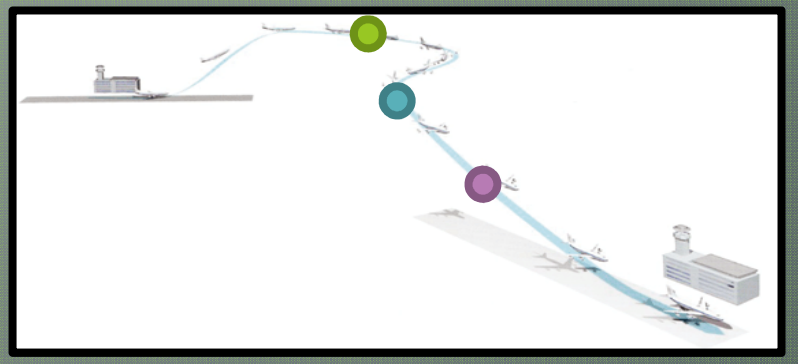
2. Voice data is segmented acoustically, and transcribed

Altitude “climb and maintain…”

Transfer “contact Atlanta Center…”

Speed “maintain two five zero knots…”

3. Algorithms categorize transcribed utterances



4. We fuse categorized utterances to trajectory data (Threaded Track)

Contribution

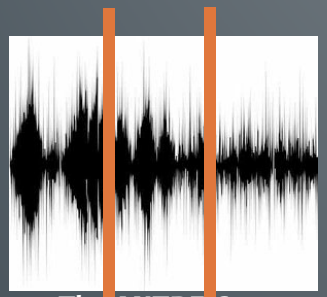
ACQUISITION

- *MITRE has a steadily growing archive of ATC voice data*
  - *Partnership with other towers and centers*
- *Not a streamlined operation (yet):*
  - *Occasionally collect hard drives from various centers*
  - *Upload data to our own repository*
- *Clean up of noise and other artifacts*
- *Affix metadata to voice data (e.g., time)*

SEGMENTATION AND TRANSCRIPTION

- Segmentation:
  - Voice data comes in a continuous stream
  - Must be segmented into 'utterances‘
    - Pauses, change of speaker, models, etc.

- Transcription:
  - Assuming audio is not of poor quality, we can use ASR software to transcribe utterances
  - Some examples: dragon naturally speaking, Nuance, etc.



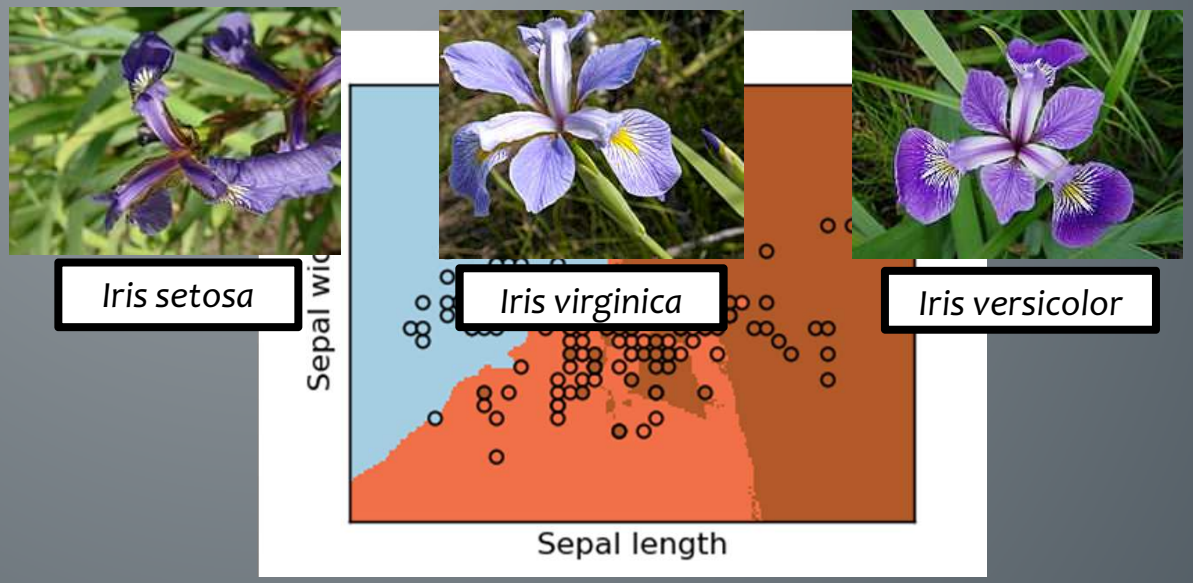
“climb and maintain…”

“contact Atlanta Center…”

“maintain two five zero knots…”

CLASSIFICATION

- "Classification": assigning a label or group to a data element



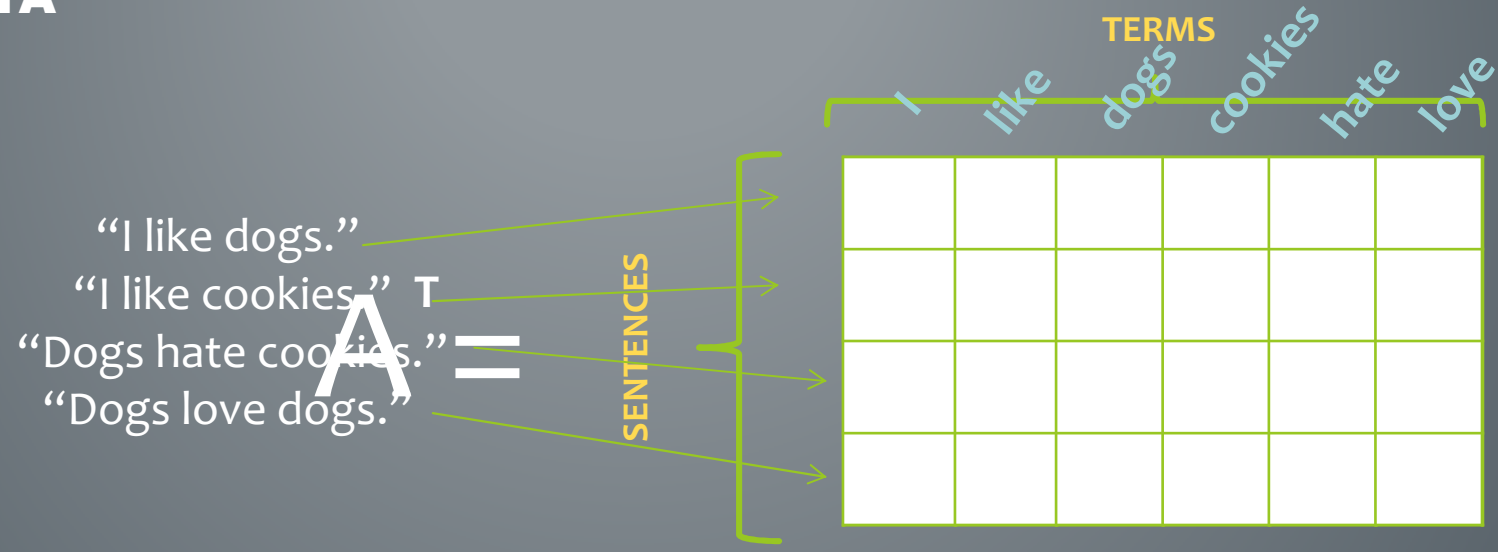
Altitude "climb and maintain…"

Transfer "contact Atlanta Center…"

Speed "maintain two five zero knots…"

CLASSIFICATION AND TEXTUAL DATA

- Classification can also be applied to textual data
  - *Transform sentences into a point in n-dim.*



- This work:
  - Non-negative matrix factorization (NMF) for utterance classification
  - Exploit domain knowledge we can to increase classification robustness

NMF PROBLEM STATEMENT

- *Non-negative matrix factorization (NMF):*
  - *Given a non-negative matrix, we can induce the following decomposition, for a fixed r:*

$$A \approx WH$$

Where:

$W \in \Re^{m \times r}$ *"basis matrix"*

$H \in \Re^{r \times n}$ *"weight matrix"*

- *To compute the NMF, typically one solves:*

$$\min_{W \in \mathbb{R}^{m \times r}_{+\cup\{0\}}, H \in \mathbb{R}^{r \times m}_{+\cup\{0\}}} \|A - WH\|_F^2$$

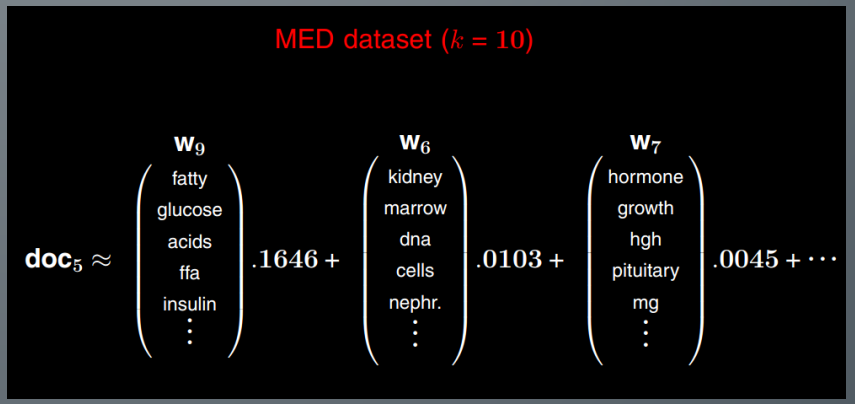
NMF IN THE LITERATURE

- *Previous uses of NMF include:*
  - *Facial detection and representation*

input images

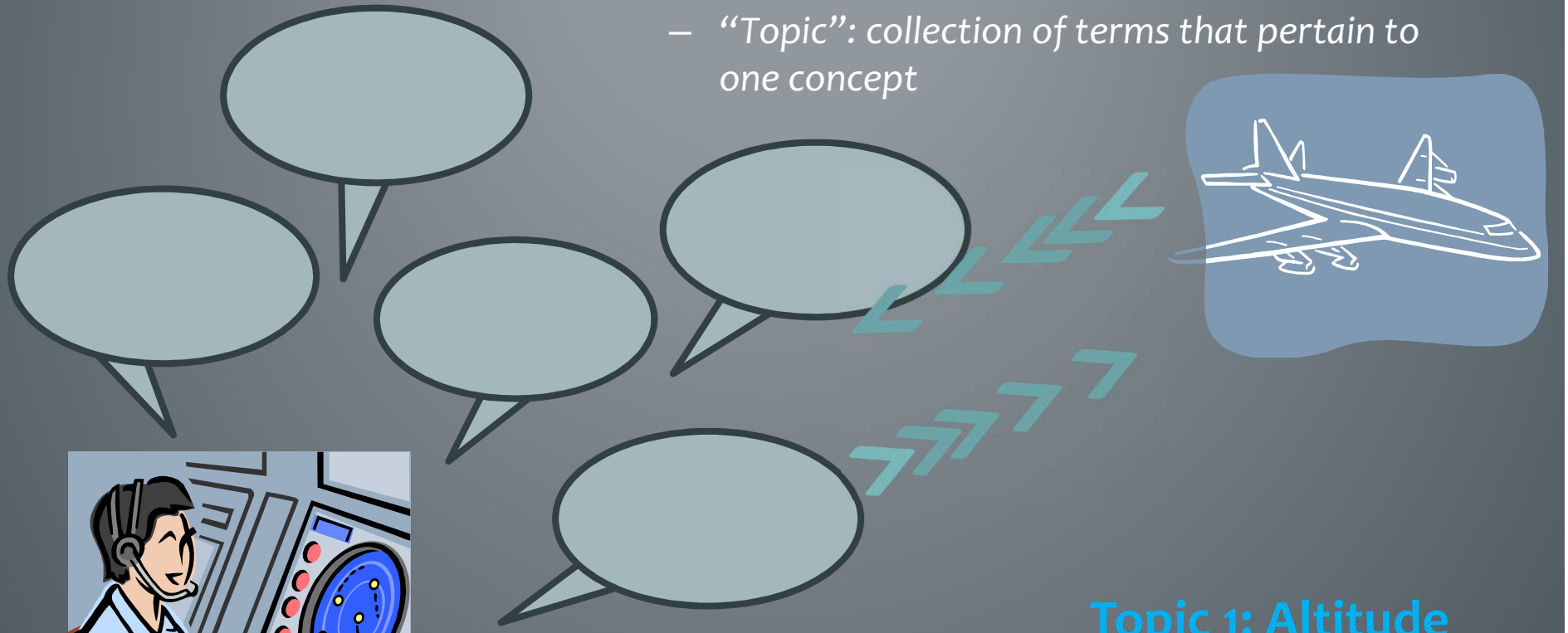
"face bases"

  - *Summarization of text documents*



NMF AND UTTERANCES

- ***Key Observation:*** *Air traffic controller utterance transcripts are "low-rank" in topic space*
  - *"Topic": collection of terms that pertain to one concept*



Topic 1: Altitude

Topic 2: Speed

NMF AND UTTERANCES

- *NMF a suitable technology*
  - *Explicit optimization scheme:*
    - *Capability to insert "smart" initial guess*
  - *Lee and Seung's multiplicative update rules:*

$$H \leftarrow H \frac{(W^T A)}{(W^T W H)} \qquad W \leftarrow W \frac{(A H^T)}{(W H H^T)}.$$

    - *Provably convergent to local min.*

Theorem 1. *The Euclidean distance $\|V - WH\|$ is non-increasing under the update rules (2). The Euclidean distance is invariant under these updates if and only if W and H are at a stationary point of the distance.*

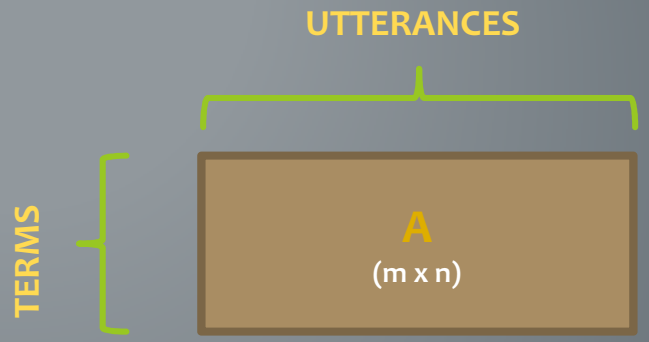
  - *Minimally dependent on tunable parameters!*

CHALLENGES ASSOCIATED WITH THE NMF

- *NMF has some drawbacks:*
  - *Rank is a user input*
  - *Decomposition is non-unique!*
    - *Unlike the SVD, which is unique*
  - *Explicit optimization: starting guess greatly affects convergence to (local) min.*
- *Domain knowledge :*
  - *How many topics to expect*
  - *Some terms that comprise the topic*
    - *Example: Altitude topic includes "feet"*
- *Upshot: domain knowledge translates to superior initial guess for* ***W*** *and* ***H***

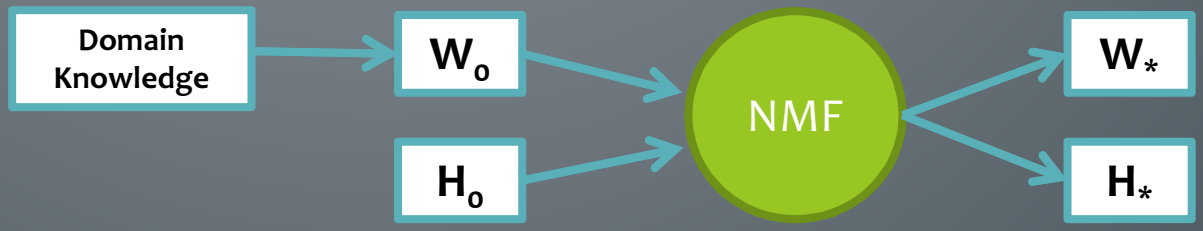
TOPIC INFERENCE FOR ATC UTTERANCES

- *Transform utterances into a matrix*



  - *Term Frequency-Inverse Document Frequency (tf-idf) term weighing scheme*

- *Apply the NMF to A, using initial guess for W and H:*



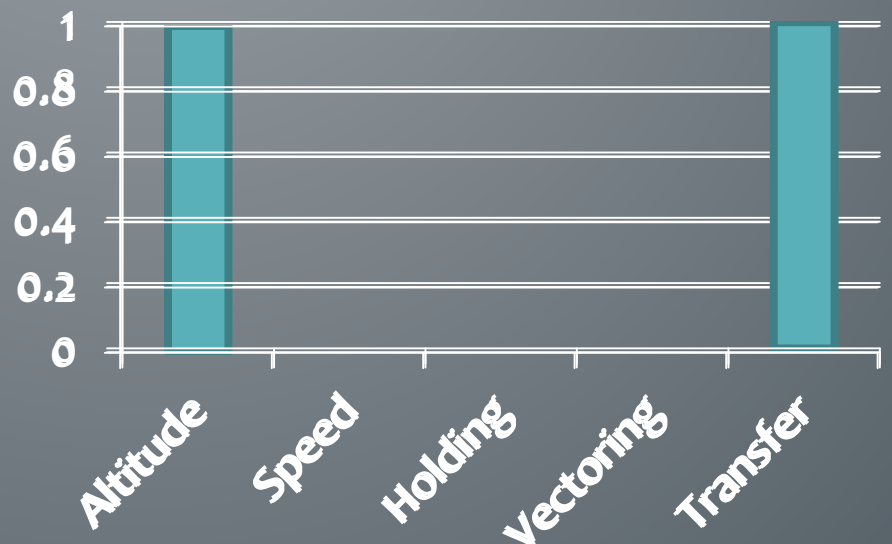
TOPIC INFERENCE FOR ATC UTTERANCES

- *Post-process W_* and H_**

"southwest 944 the next sector is reporting just light chop at 39 did you want to stay at 37"

"southwest 1242 contact seattle center 135.15 thank you"

"stay at 37"



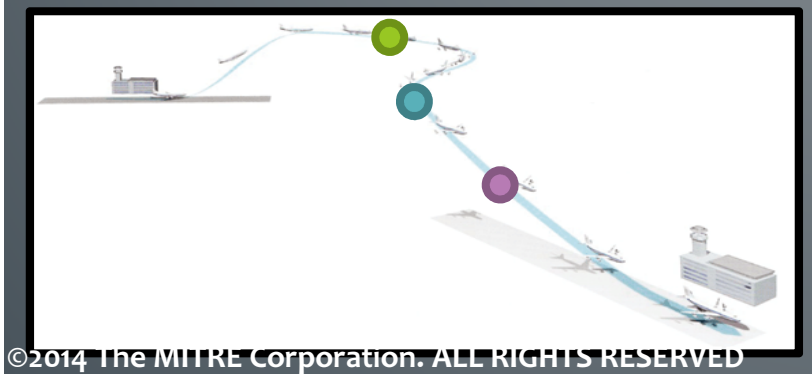
ACCURACY RESULTS

Test	Description	n	m	WS-NMF Accuracy
A	Human-transcribed utterances from ZTL Sector 16. Covers the en-route portion of the DAWGS SID from FL150 to FL230.	804	6735	99.20% (fp: 11)
B	Human-transcribed utterances from ZTL Sector 32. Covers the en-route portion of the DAWGS SID from FL240 to FL290	633	5350	98.66% (fp: 4)
C	Human-transcribed utterances from ZTL Sector 50. Covers the en-route portion of the FLCON star from FL240 to FL340.	1821	11222	99.72% (fp: 2)
D	ASR-transcribed utterances from a northern California HITL. The software Loquendo was used to transcribe the utterances.	276	1384	86.58%

Future: ASR + Classification

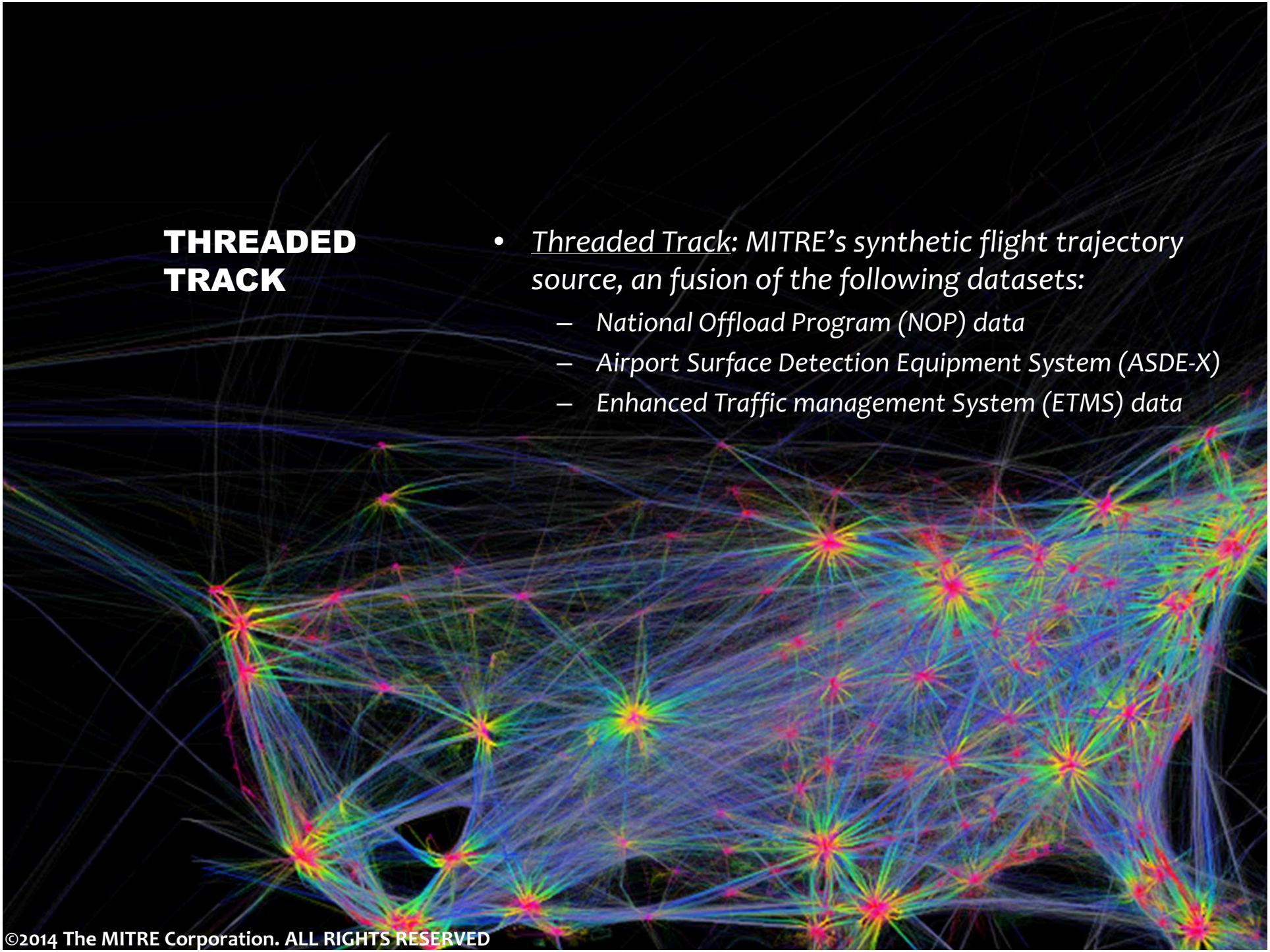
FUSION

- We can affix the utterance labels to their likely trajectory, at the correct spatio-temporal point
- Components:
  - Voice metadata (timestamps, tail ID)
  - "Threaded track“
- Can fuse even more: ATSAP/ASAP report, TAWS and TCAS alerts, etc.



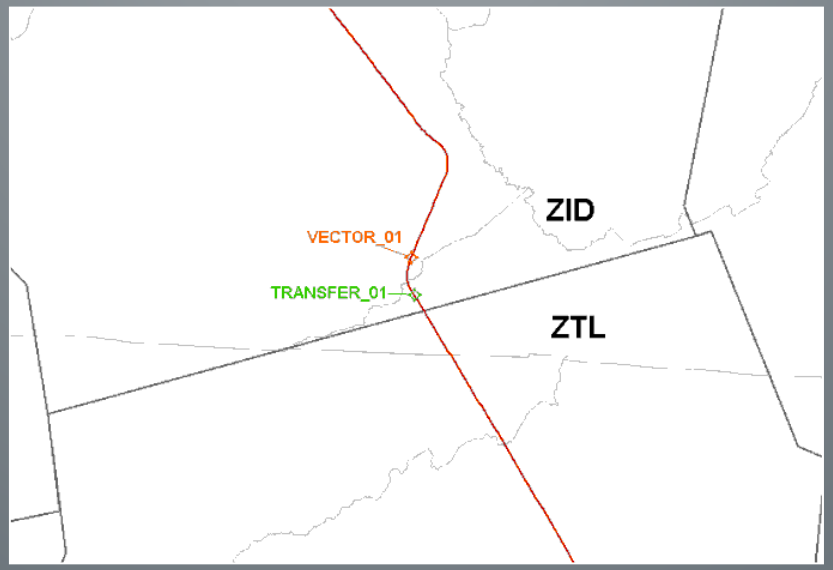
THREADED TRACK

- *Threaded Track: MITRE's synthetic flight trajectory source, an fusion of the following datasets:*
  - *National Offload Program (NOP) data*
  - *Airport Surface Detection Equipment System (ASDE-X)*
  - *Enhanced Traffic management System (ETMS) data*



LOOKING UP FUSED DATA

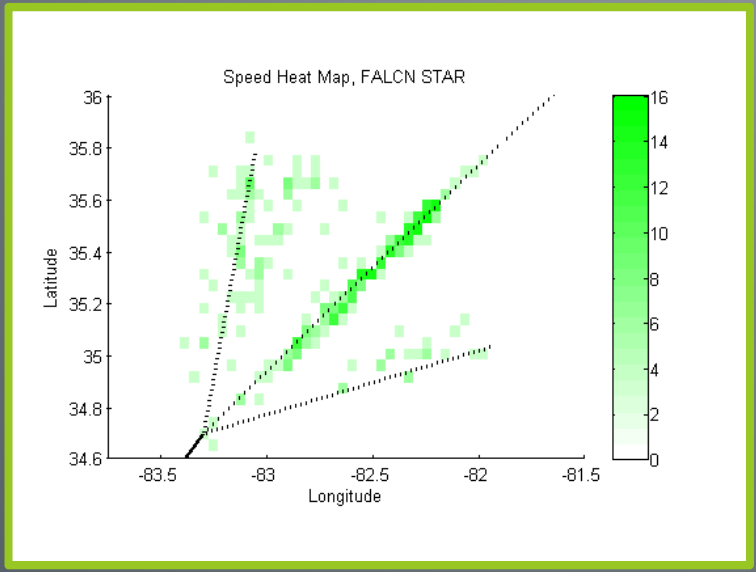
- *A "Google-like" service for flights:*
  - *Integrates trajectory lookups with available voice data and its identified label(s)*

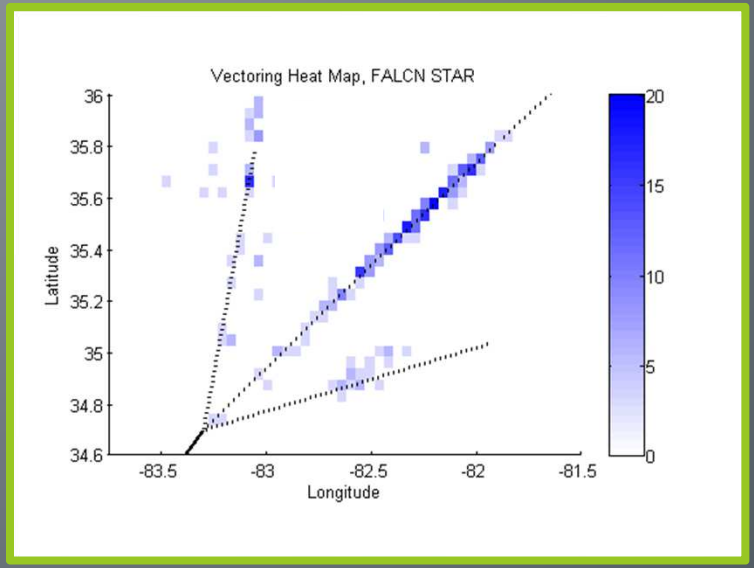


Time	ACID	Label	Category	Utterance	Duration
00:08:06.83650	BTA5900	VECTOR_01	Turn/Reroute	"Clear direct"	00:00:09
00:35:29.73470	BTA5900	TRANSFER_01	Transfer of Control	"Contact"	00:00:04

UTTERANCE HEATMAPS

- *Aggregate, and analyze utterance-fused trajectories*
  - *Visualize, with respect to a procedure, number of utterances*
  - *Another possibility: visualize, and quantify, temporal variation*





CONCLUSION

- *ATC utterance data will eventually become accessible*
  - *Worthwhile to invest technologies to analyze such massive datasets*
- *We proposed a classification algorithm based on the NMF*
  - *Explicit optimization scheme allows us to leverage our domain knowledge*
  - *Accurate, and easily implemented/scaled*
- *"Fusion": ATC voice data + label + trajectory*
  - *Towards telling the full "flight story"*
  - *Will enable advanced analysis (e.g., procedure pre-/post- analysis)*

FUTURE WORK

- *An initial guess for the weight matrix, **H***
  - *ASR transcription is not perfect:*
    - *Could "confidence" scores from transcription be translated as an initial guess?*
  - *Some can belong to multiple topics, though may be more likely to pertain to one topic*
    - *Again, could this preference be translated to an initial guess for H?*
  - ***Both involve interpreting H as a "probability"***