



Identifying topics in the **TRAINERROAD** cycling podcast

Jason Pizzollo

11/13/2020



TRAINERROAD Overview:

App



Podcast Format

Listener Q&A and Interviews

Data

3 years of podcast transcripts

- Median length 115 min
- 141 episodes

Problem Statement

Podcast episodes are a mix of topics

Can we use NLP and topic modeling to ...

- Determine the composition of a podcast episode?
- Help frame the structure of a conversation?

Topics



Training Plan

Plan Volume Recovery Season Phase



Nutrition

Food Bottle Calorie Gram Carbohydrate



Physiology

Muscle Study Blood Cell Effect



Strength Training

Weight Strength Fiber Gym Force



Equipment

Tire Wheel Rim Pressure Valve



Heat & Hydration

Sodium Sweat Water Heat Hot



MTB Racing

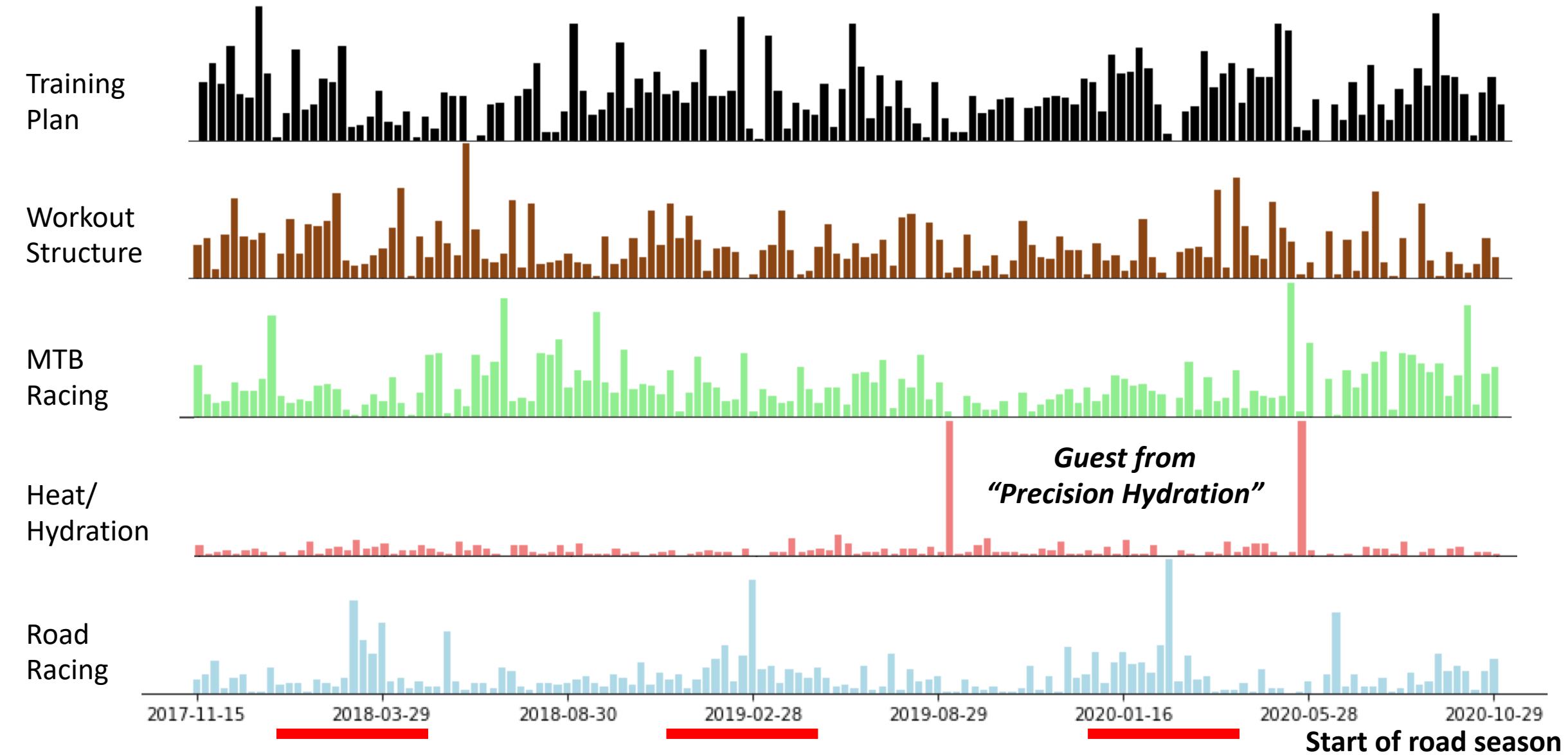
Mountain Cross Section Trail Skill



Road Racing

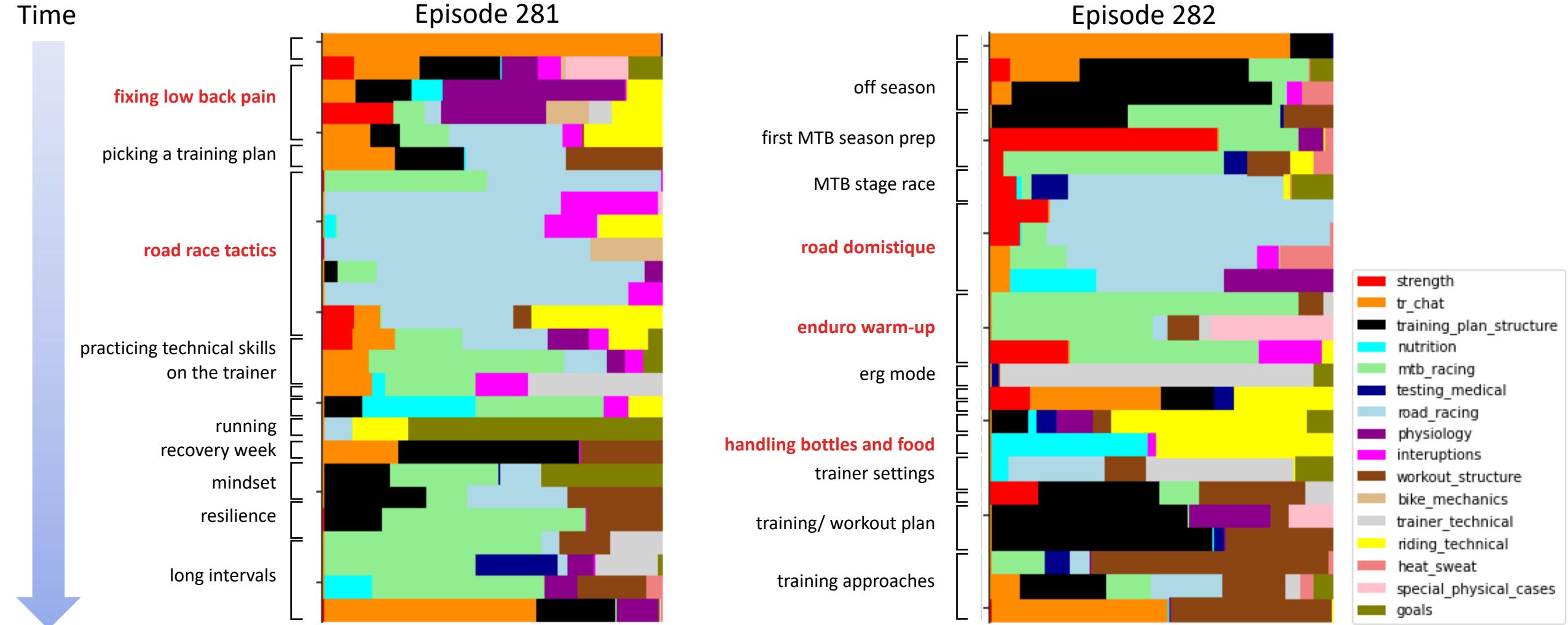
Team Group Field Attack Crit

Topic distributions change between episodes

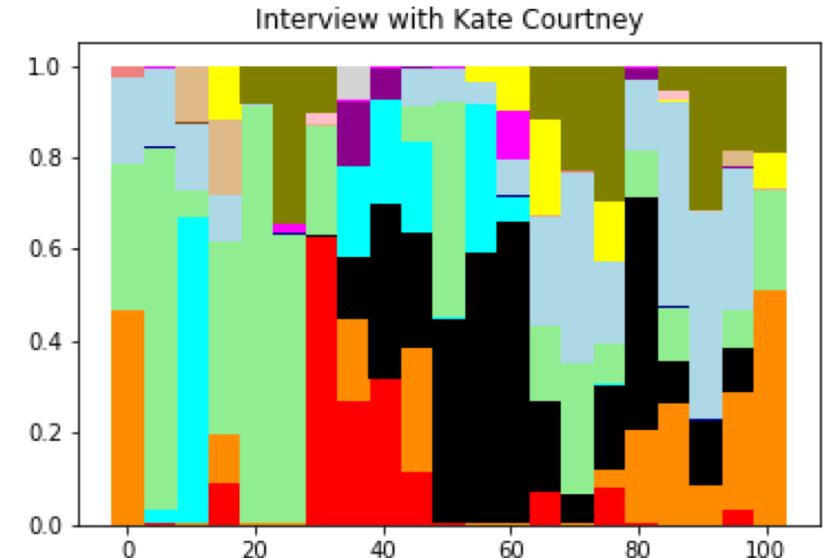
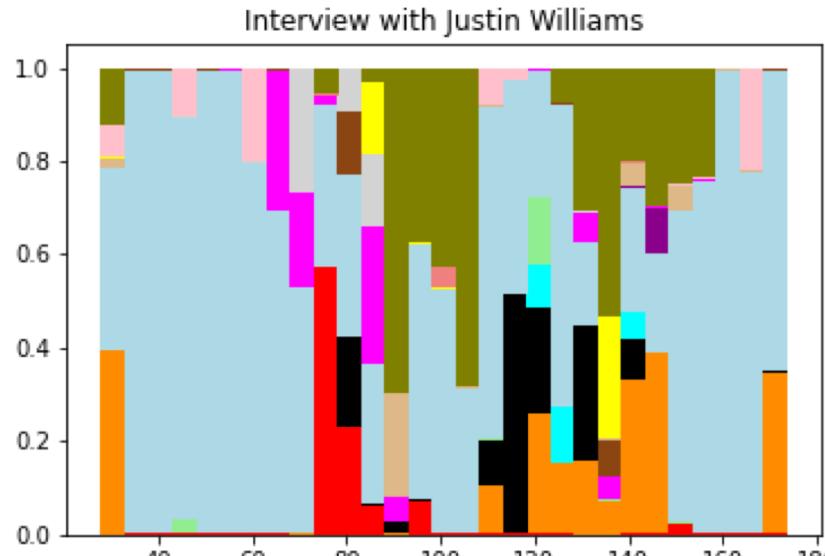


Changes in topics throughout an episode

...at 5 min resolution



Common interview structure



Structure of interview:

- ~25% mostly sport (discipline) specific
- Strength training
- Training plan structure
- Goals
- Bookend with informal chat

strength
tr_chat
training_plan_structure
nutrition
mtb_racing
testing_medical
road_racing
physiology
interruptions
workout_structure
bike_mechanics
trainer_technical
riding_technical
heat_sweat
special_physical_cases
goals

Summary

NLP-based topic modeling can identify...

Composition of a podcast episode

Coarse structure of conversations

Limitations...

Some confusion about topics

- balance of accuracy at large and small scale