apt advancing practices of teachers

Applications of Machine Learning applications in support of improving the everyday pedagogical practices of teachers

Data and project overview

APT draws on a corpus of video of K-12 teachers of all grade and subjects. Teachers who are seeking National Board certification (the highest recognition of professional practice) submit video and a number of associated artifacts for reviewers to consider. Over the years this corpus has been looked at by education researchers for evidence of particular practices. In some cases, reviewers have looked across the corpus for examples of science teachers meeting aspects of the Next Generation Science Standards. Other education researchers, who are seeking to advance the profession, have identified particular practices as being highly associated with effective or ambitious teaching, e.g., TeachingWorks High Leverage Practices or Danielson's Framework. In every case domain experts reviewed the video and identified the relevant portion with a tag. The APT project was conceived as a way to automate and scale this identification process, and offer it in near realtime to educators. The idea is relatively simple, how can we build a system that rapidly recognizes relevant practices, which teacher education experts have identified as important, in what teachers do everyday in their classrooms so that these teachers can improve upon their practices?

For the purposes of our work over the last year, we have focused on a subset of the corpus that has been tagged with TeachingWorks High Leverage Practices. Rather than using speech to text processors to generate transcripts, we relied on videos in the collection that were already transcribed. To date, all of our analysis has focused on a few tags.

Why is this worth doing?

Though teaching is a public and social practice, in many instances educators (from pre-K through graduate school) find themselves teaching in learning environments alone with students. Whereas some K-12 school districts have formal induction programs (where more accomplished teachers mentor new recruits), regardless of whether we are thinking of the typical graduate student instructor, or the recently certified K-12 teacher, much of the early years—and the vast majority of the remainder of a teacher's experiences—are solo and siloed from a development of practice perspective. Recently however many programs are beginning to use digital video recordings of teaching practice as a basis to engage in conversations about improving practice. Advancing the Practices of Teachers (APT) seeks to build on the underlying concepts of reflecting with a more experienced colleague on digital video and the power of "big data" and machine learning to support the pedagogical development of teachers.

What is the current state of art?

Students studying to become teachers spend time in classrooms in apprenticeship roles of a variety of modes, with time under tutelage being the biggest variance. Some new K-12 teachers engage in induction programs where more experienced educators observe their teaching, but these observations do not happen during every lesson that the new teachers teach. In higher education, the selection of graduate student instructors (GSI) frequently is not based on pedagogical expertise, but rather might be based on financial need of the GSI, or in limited instances demonstrated expertise with the subject matter. While GSI might receive post class student evaluations, it is rare for a GSI to get direct teaching support by a more experienced educator.

As the success of using digital video in sports improvement has grown, educators are beginning to look to video as a method for improving practice of teachers.

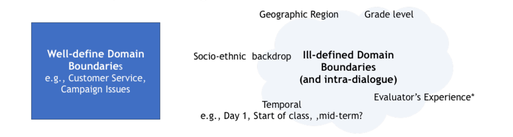
Challenges

Limited (training) set size
Noisy; discourse is messy
Inconsistent lengths
No "bad" practice examples

Does sufficient information exist in the audio transcripts to ascertain best practice behavior?
Can such information be extracted as features useful to a machine learning model?
What would such a model be, and can it be efficiently calculated?
What are the unknowns?
Do we sufficiently understand the problem?
What are the adjacent fields of Inquiry?
What information and methods are useful?

Research

"Boundary" Challenge…



Tennis –
Badminton, Ping Pong, Volleyball…Handball

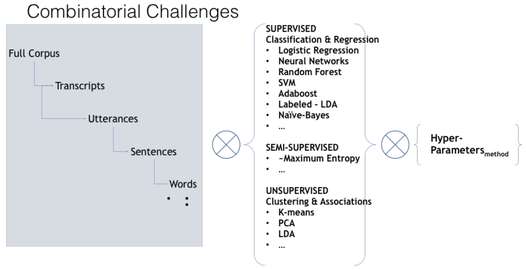
Um, "Injustice anywhere is a threat to justice everywhere."

"He said.. She said" vs whatever the underlying topic of discussion may be about.

…suggests we might need to incorporate features from outside the transcripts , manually annotate a training corpus, or get more data.

Boundary

Combinatorial Challenges



Combinatorial

Experiments

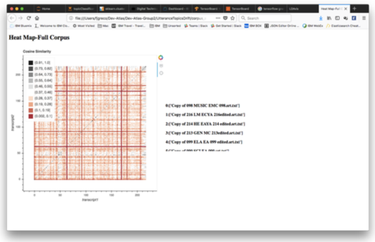
Term Frequency – Inverse Document Frequency Matrix TF-IDF

D1: This is a dog.
D2: This is a cat
D3: Cat[s] and dogs[s] fight.
D4. A cat is a cat, not a dog.

	D1	D2	D3	D4		IDF Factor
This	.25 (.576)	.25 (.576)	0	0	2	2.303
is	.25 (.474)	.25 (.474)	0	.125 (.237)	3	1.897
a	.25 (.347)	.25 (.347)	0	.375 (.520)	5	1.386
fight	0	0	.25 (.749)	0	1	2.996
cat	0	.25 (.402)	.25 (.402)	.25 (.402)	4	1.609
dog	.25 (.474)	0	.25 (.474)	.135 (.256)	3	1.897
not	0	0	0	.125 (.375)	1	2.996
and	0	0	.25 (.749)	0	1	2.996
	4	4	4	8	20	

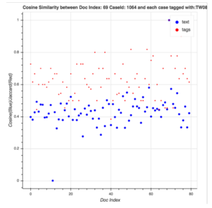
TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).
IDF(t) = ln(Total number of documents / Number of documents with term t in it).

Full Corpus Transcript to Transcript Cosine Similarity View



Inquiry 5 -Tone Analysis

Figure 5: Similarity Check Example using Cosine Metric and Jaccard Index



TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).
IDF(t) = log_e(Total number of documents / Number of documents with term t in it).

TF-IDF matrix :: $\text{TF-IDF}_{\text{document, term}}$

Cosine Similarity: Inner product of document vectors from the TF-IDF matrix

Jaccard Index: Intersection of Tag Sets / Union of Tag Sets

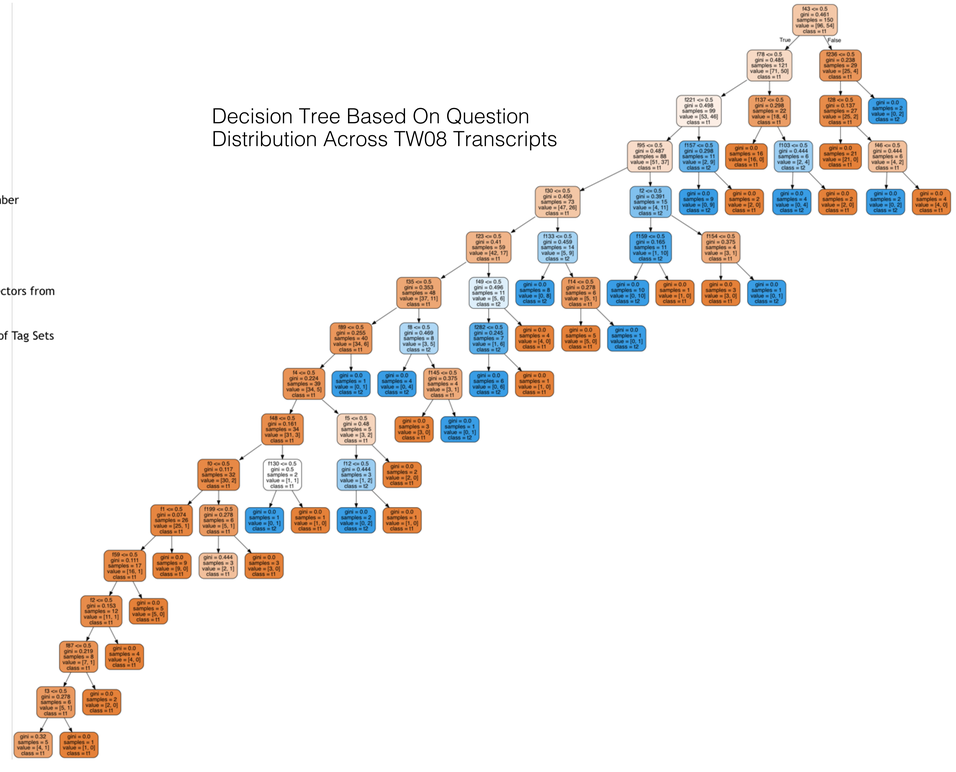
K-Nearest Neighbor Results Example>

```
# retained documents: 217

vectorize
Tag: TW04, Count: 217
Tag: TW12, Count: 217
multilabel.shape: (217, 2)
Tags: [TW04', TW12']
Run:0
Run:1
Run:2
Run:3
Run:4
Total Test Successes: 76 Total Test Failures: 9
test_success_ratio: 0.894117647059

Total Train Successes: 889 Total Train Failures: 111
train_success_ratio: 0.889
```

Decision Tree Based On Question Distribution Across TW08 Transcripts



---

team: Orrin Murray UEI, Franck Greco IBM/UChicago alum, Orest Xherija UChicago, Cornelia Bailey, IT Services