Project Proposal

Faith, Hope and Love in TED Talks

Sean Davern
Seattle, Fall Cohort, weeks 6.5-9

The words faith, hope and love show up prominently in a collection[1] of 2,467 TED talk titles, descriptions and transcripts[2]:

```
Titles that contain the words:
  faith 5
  hope 5
  love 35
  faith, hope or love 45
```

```
Descriptions that contain the words:
  faith 13
  hope 66
  love 114
  faith & hope 2
  hope & love 2
  faith & love 1
  faith, hope & love 0
```

```
Transcripts that contain the words:
  faith 168
  hope 1022
  love 1175
  faith & hope 88
  hope & love 520
  faith & love 101
  faith, hope and love 56
```

I propose using NLP to process the languange content (transcripts) of the collection of TED talks to assess the types of faith, hope and love the talks describe. This analysis would use unsupervised learning techniques, e.g. clustering, to hopefully differentiate and group the kinds of love, hope and faith the talks were discussing. For example, are the authors describing their love for something, encouraging the expression love, or talking about love in some other way? I'd then use supervised learning with this output and talk ratings and views as a way to assess popularity and/or audience

reaction to the groups.

The goal of the overall analysis is to see if viewers are resonating with or enjoying any particular type or collections of what I believe are core human values.

Data

The data I propose[1] to use is available from kaggle. The data comes in two files containing the indicated fields:

ted_main.csv: comments, **description**, duration, event, film_date, languages, name, num_speaker, publication_date, **ratings**, related_talks, speaker_occupation, tags, **title**, url, **views**

transcripts.csv: **transcript**, url