ML Week

Features and Modeling

Jeff Abrahamson

20–22 juillet 2016

Vector spaces

Vector spaces

Features are dimensions

Feature extraction

Feature engineering

Feature extraction

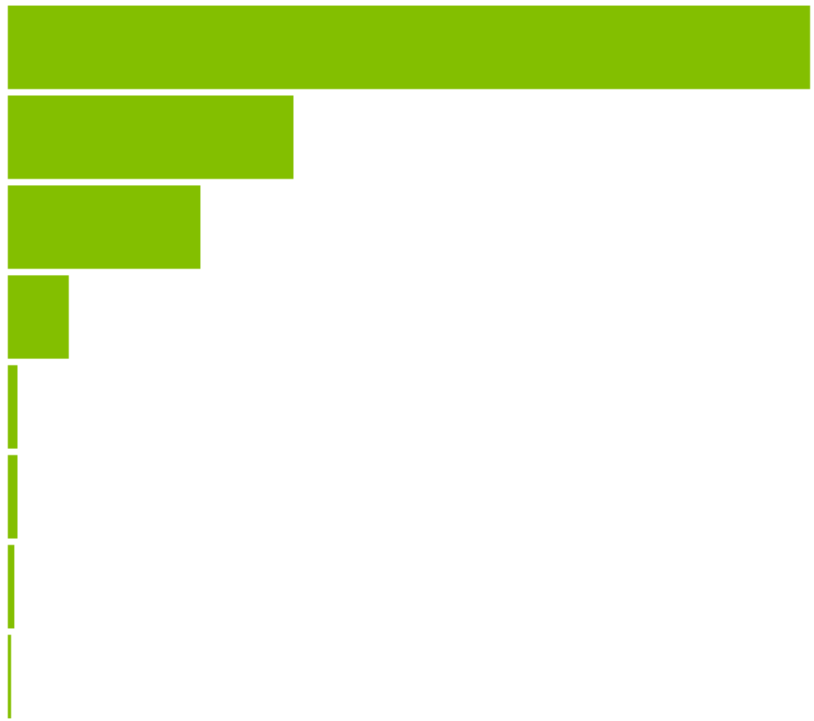
Feature engineering

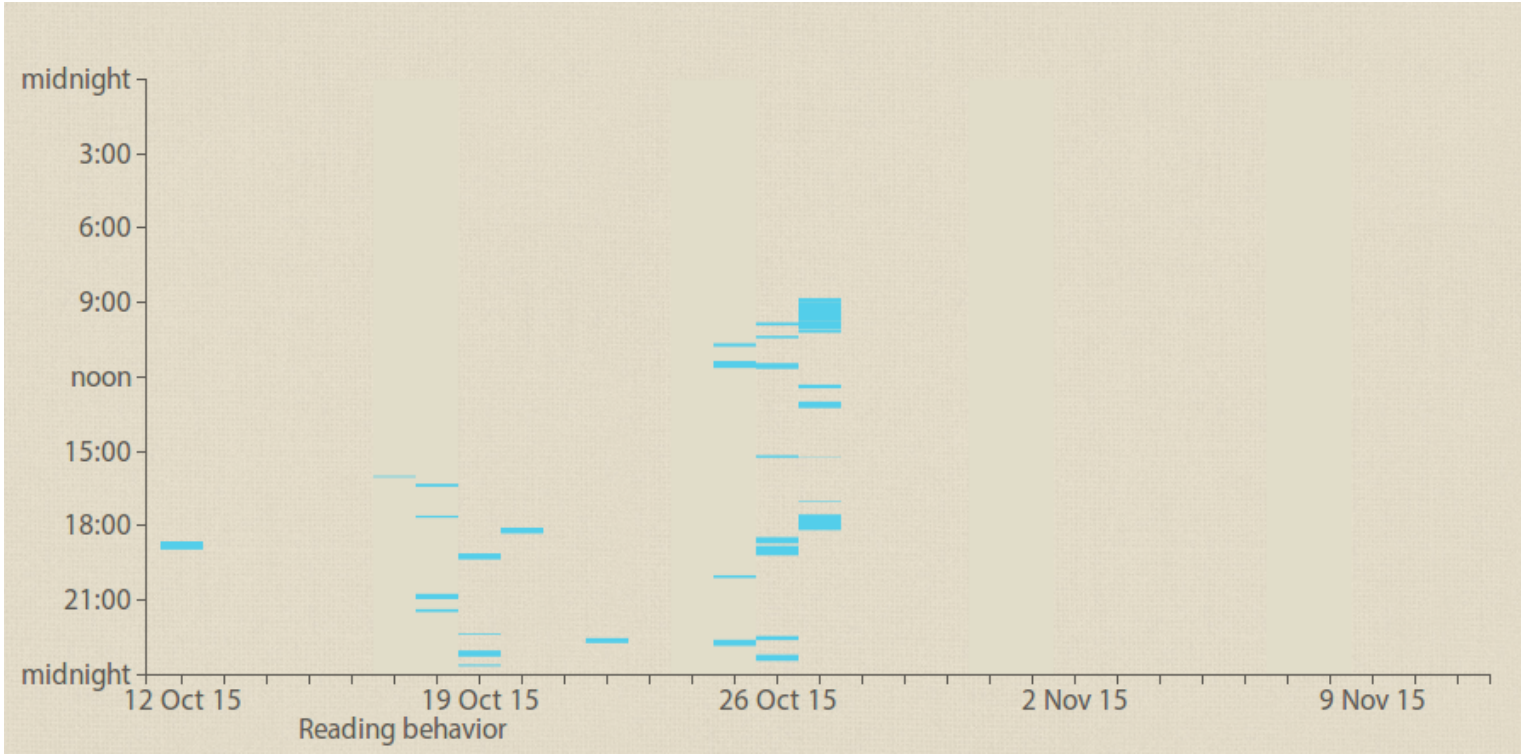
Synthetic features

Feature Engineering

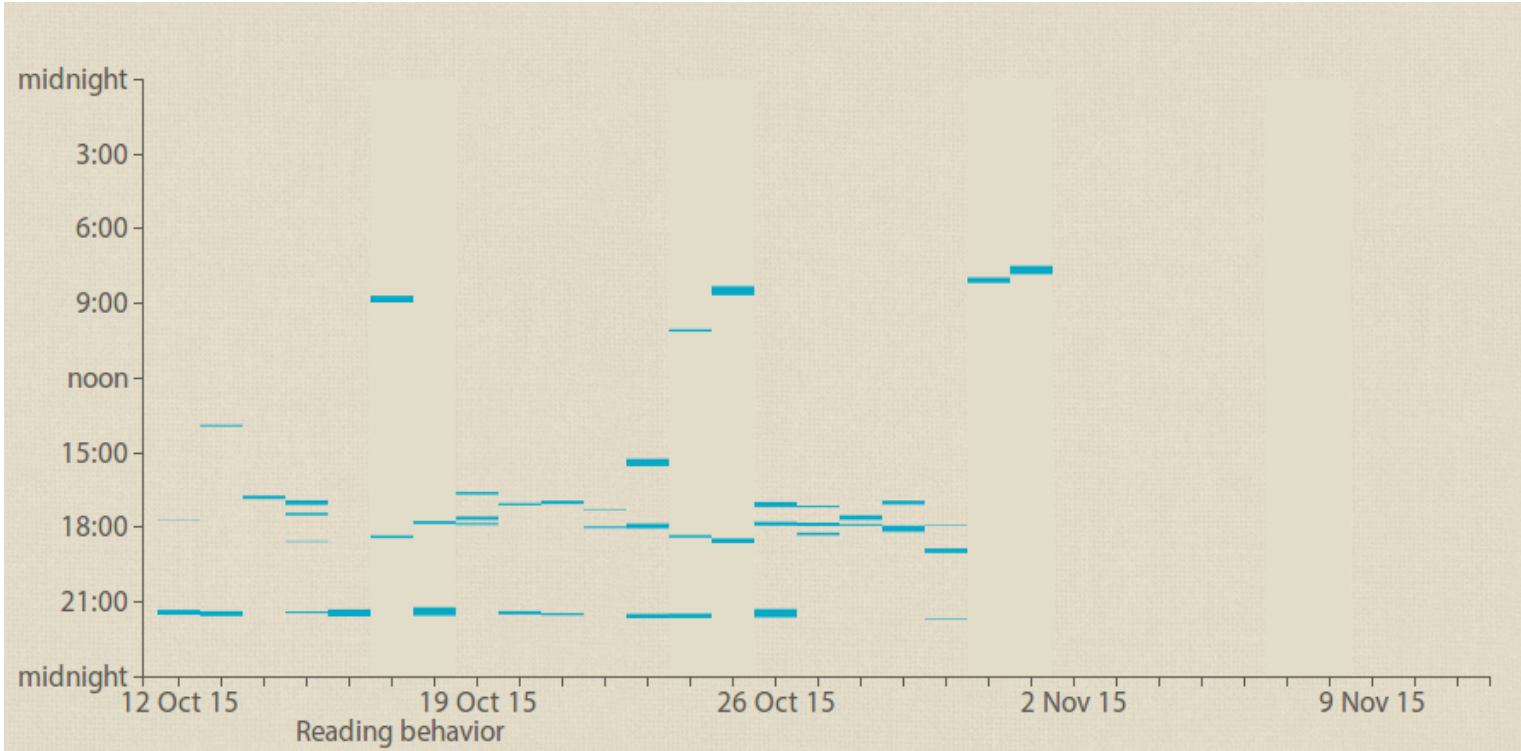
1. Brainstorm
2. Pick some
3. Make them
4. Evaluate
5. Repeat

Value	Count	Percent
Mr.	517	58.025%
Miss.	185	20.763%
Mrs.	125	14.029%
Master.	40	4.489%
Dr.	7	0.786%
Rev.	6	0.673%
Sir.	5	0.561%
Col.	2	0.224%
Jonkheer.	1	0.112%
Lady.	1	0.112%
the Countess.	1	0.112%
Ms.	1	0.112%





Jellybooks



Jellybooks

One of K = one-hot encoding

Text features

Bag of words

- Corpus (documents)
- Vocabulary (set of unique words)
- Words

Text features

Bag of words

- Order doesn't matter
- Stop words
- Stemming *(racinisation, désuffixation)*
- Lemmatisation *(transformer en lemme)*

Image features

- Corners, edges (rotation invariant, but scaling can hide)
- More complex: scale space or RNN
- Point matching is easy

Image features

Problems

- Illumination
- Scale
- Rotation
- Skew (perspective)
- Data size (matrices not sparse)

Questions?

purple.com/talk-feedback