



Sentiment: Beyond the basics

Fall 2011

Sentiment: the easy stuff is already done

- Lots of places to get document level sentiment
 - Linguistic
 - Model based
- Entity Sentiment is becoming more prevalent
- It's time to tackle the hard problems:
 - Tricky negation
 - Misplaced sentiment
 - Relative sentiment
 - Contextual sentiment
 - Multi-axis sentiment
- Short content and other social media issues

Negation and Intensifiers

- Negation

 - “*Not always* the best idea”

 - “It *wasn't* horrible”

 - “*Not necessarily* a good movie”

- Intensifiers

 - “*Spectacular* failure of a meal”

 - “*Really* bad actors”



Complex Sentiment Patterns and relative sentiment

- Complex patterns

Congrats to [@JacobyEllsbury](#) who was named Players Choice Awards AL Comeback Player of the Year! [#RedSox](#). Oh, and the Yankees still suck

- Relative sentiment

“Pepsi sucks, unlike coke”

“Love my Bruins, even more than the Sox”

“iPhones are better than Android”



Identifying Contextual vs. Polar

- Examples
 - “Good morning world” - contextual
 - “WTF the Sox blew it... again !!!” – Polar
 - “Still no power, waiting...” - contextual
- Need to decide if it's Contextual or Polar, and then proceed



Multi-meaning words

- Examples:

Hit (good) – He hit a mammoth blast over the wall for the win

Hit (neutral) – Hit me with your best shot

Hit (bad) – He hit me in the head

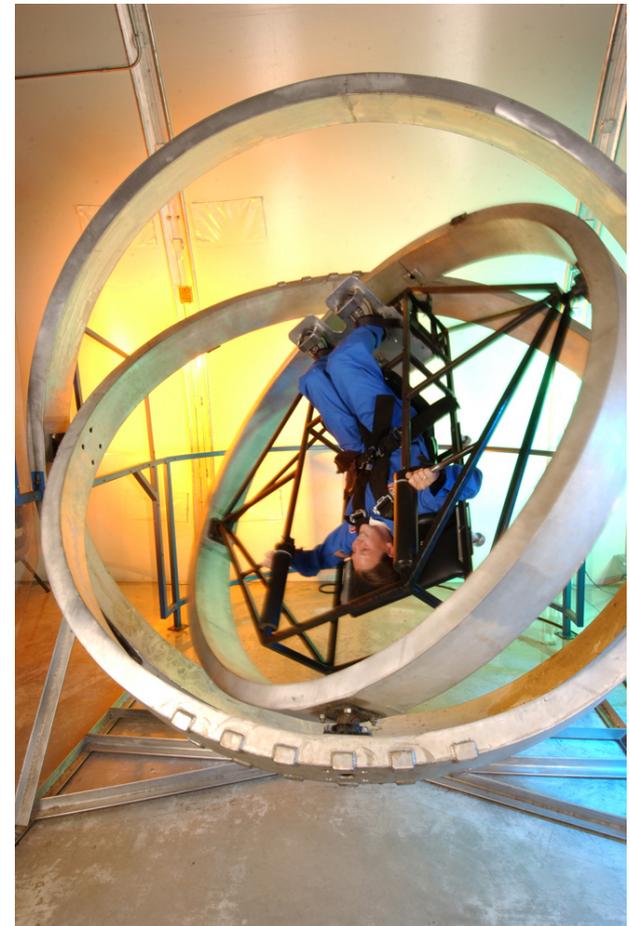
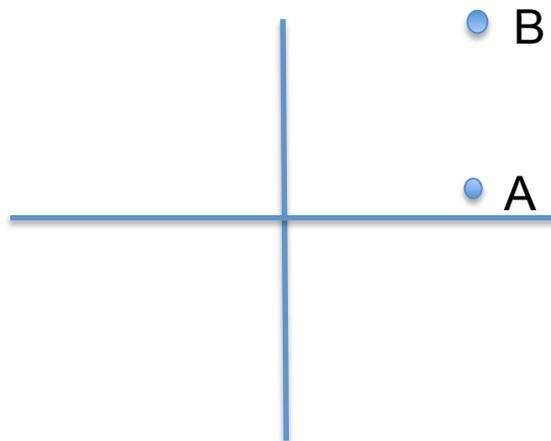
Chicken (noun) – Love chicken nuggets

Chicken (adj) – Bob is such a chicken



Multi-axis Sentiment

- A - We had a decent game today, solid hitting and we didn't make any mistakes in the field (positive, but not intense)
- B - We KILLED em today, hit as well as I've ever seen, and the fielding was perfect (positive and intense)



Putting it all together: Social Media

- Problems:

U Sck

they don't rite too goodly IMHO

LUV DEM acronyms and ☹

WHACKY CAPitalization AND GRUMPY peeps ;-)

- Needs:

Social media is the driver of consumer opinion now, so making sentiment work in this domain is of paramount importance

- Better contextual understanding of what's being said
- Need to ignore the vast majority of the content
- New POS taggers for this type of content
- Many of the core sentiment improvements described herein...



Thank you.

*Lexa*LYTICS

