Mining Revisions to Questions and Answers on StackExchange.com

Felix GREZES & Michelle MORALES

StackExchange

Motivation

Broader implications and related work:

- Sentence compression
  - Yamangil and Nelken (2008)
- Text summarization
  - Yamangil and Nelken (2008)
- Machine translation
  - Wubben, Bosch, Krahmer (2012)
- Keyword extraction, text correction, edit classification, etc.

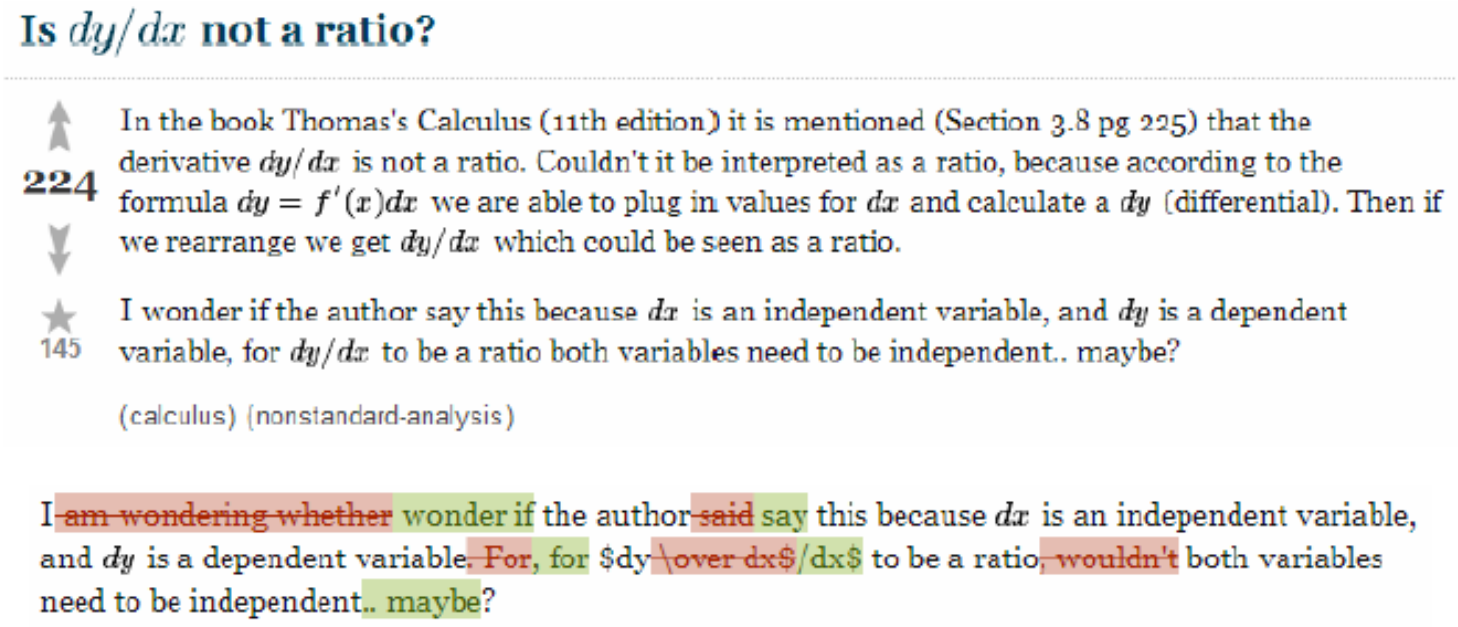
Project Outline

1. Data collection
2. Preliminary analysis
3. Task 1
    a. correlate score with features
4. Results
5. Task 2
    a. classify edits
    b. suggest meaning-preserving revisions

Integration of 3 core topics: NLP/ML/Web

Data Collection

- 111 Q&A sites
- collected ~6500 question/answer pairs



Is dy/dx not a ratio?

224

In the book Thomas's Calculus (11th edition) it is mentioned (Section 3.8 pg 225) that the derivative dy/dx is not a ratio. Couldn't it be interpreted as a ratio, because according to the formula $dy = f'(x)dx$ we are able to plug in values for dx and calculate a dy (differential). Then if we rearrange we get dy/dx which could be seen as a ratio.

145

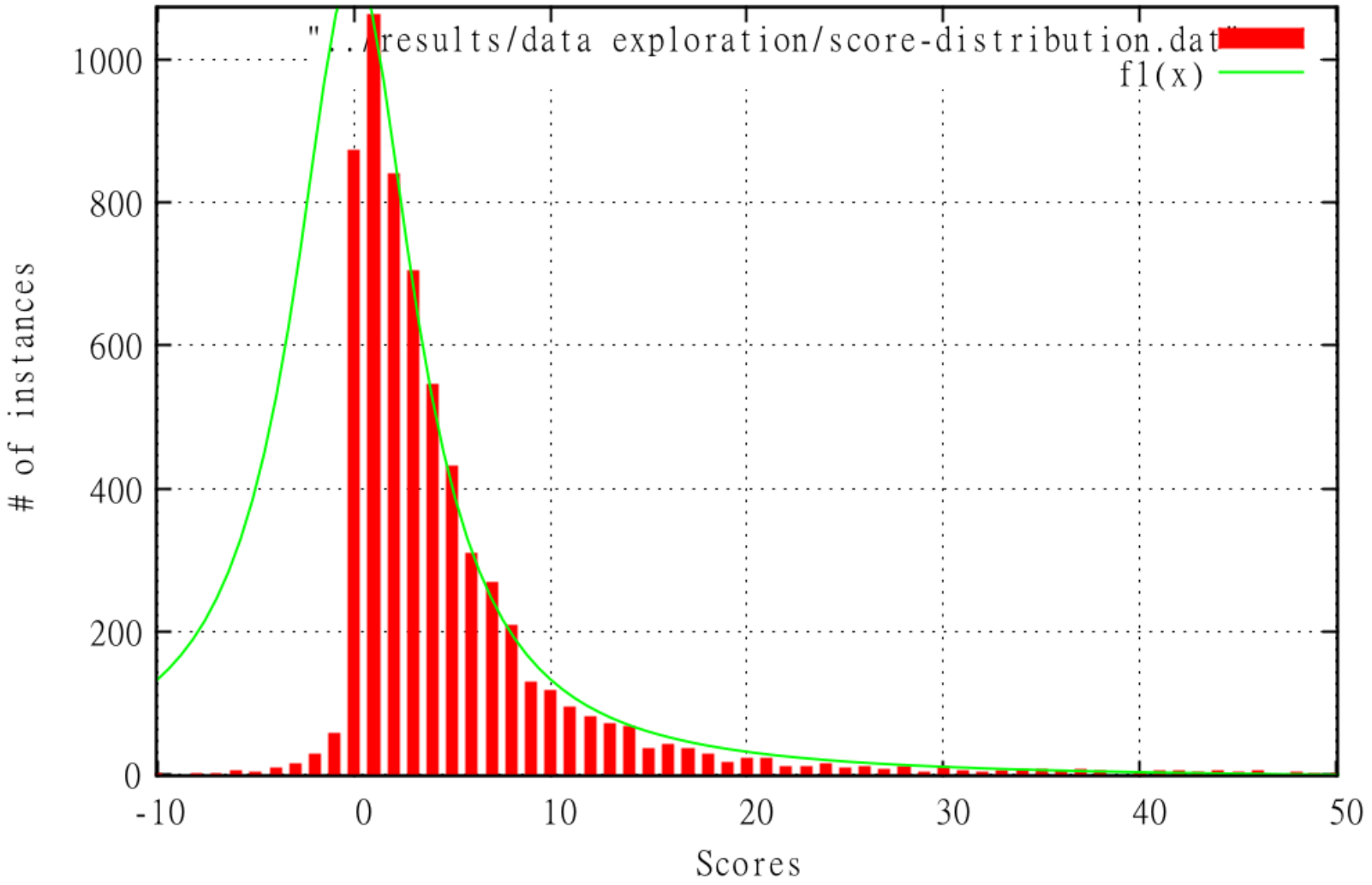
I wonder if the author say this because dx is an independent variable, and dy is a dependent variable, for dy/dx to be a ratio both variables need to be independent.. maybe?

(calculus) (nonstandard-analysis)

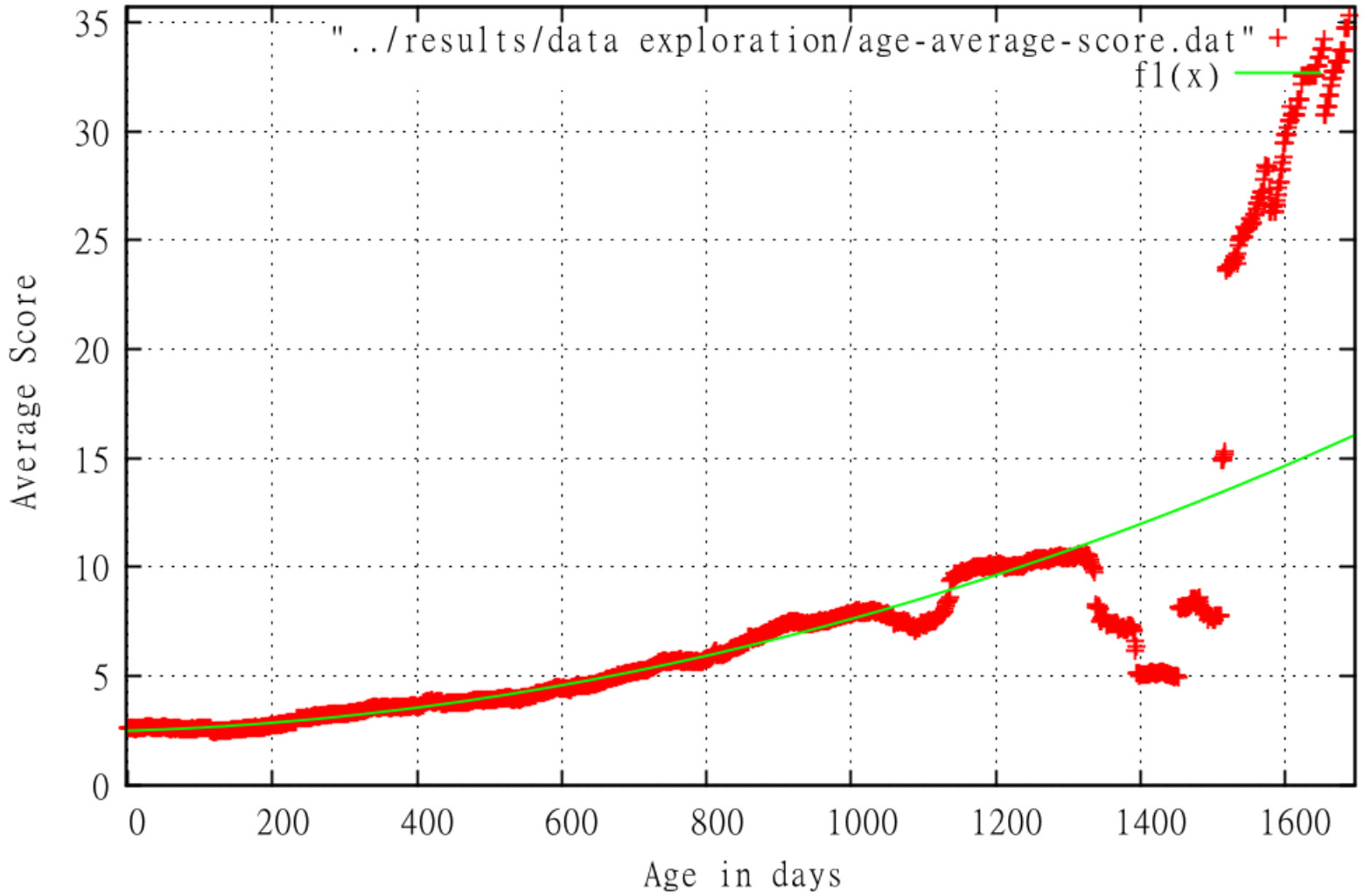
I ~~am wondering whether~~ wonder if the author ~~said~~ say this because dx is an independent variable, and dy is a dependent variable~~. For~~, for $dy ~~\over dx$~~/dx$ to be a ratio~~, wouldn't~~ both variables need to be independent.. maybe?

- title, text, tags, score, rank, edits, metadata (images, code, links)

Preliminary Data Analysis



Preliminary Data Analysis

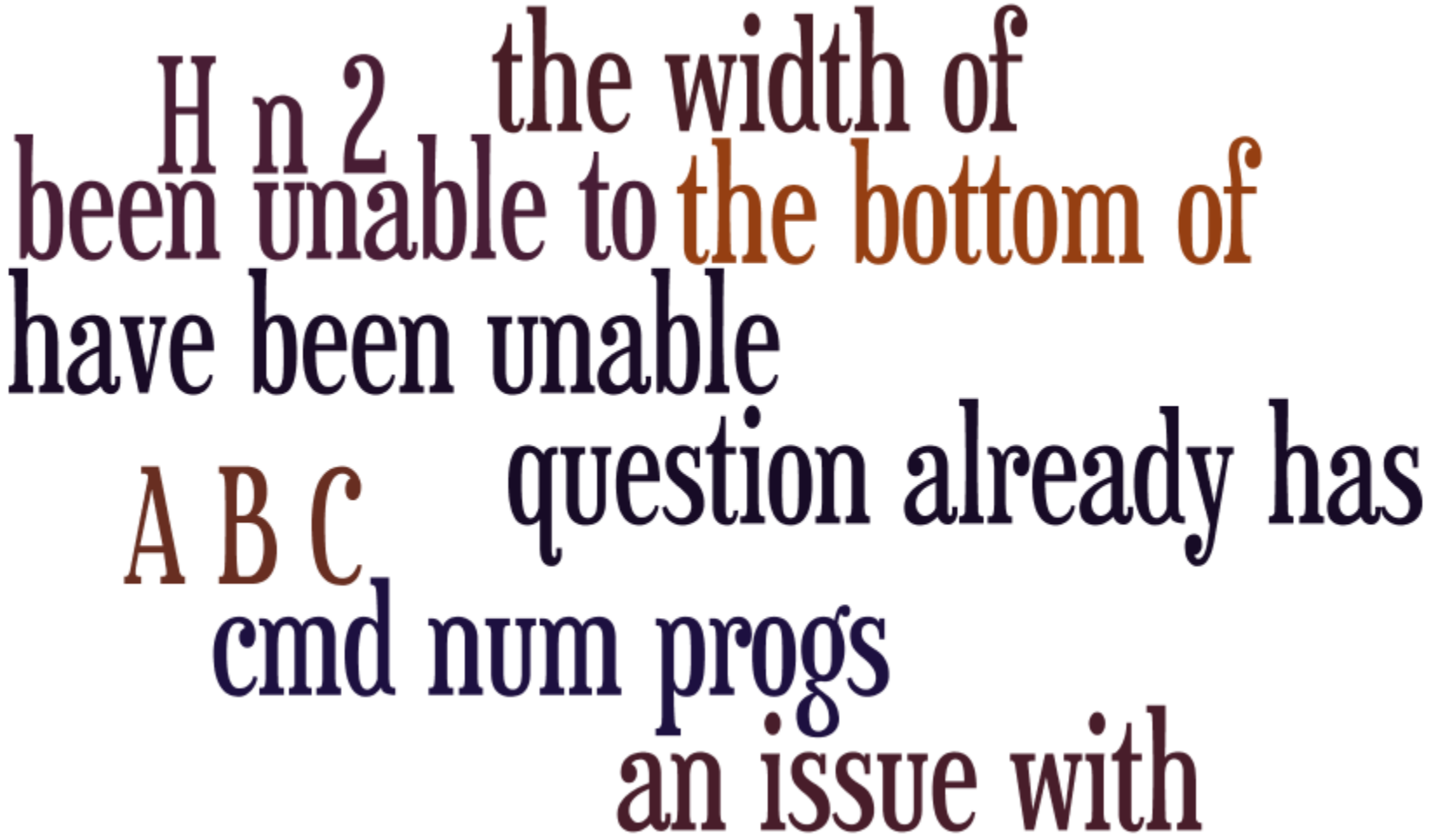


Task 1: Correlation Results

- Goal: find out if the textual information (text, title, tags) was responsible for high scores.
- Test of different representation
  - Bag-of-words, N-grams
  - Stemming, no Stemming
- Simple SMO on Weka

Low correlation on test sets: ~1%

Most prominent tri-grams



Next Task: Revisions

Revisions are inherently changes
in the positive direction.

- Unsupervised approach
- Use tools to classify edits into Meaning-preserving, Meaning-altering
  - Thesaurus, Wordnet
  - Hunspell as in Max, Wisniewski 2010
  - Adapt Daxenberger, Gurevychy 2013
- Suggest common, meaning-preserving revisions only

References

- Automatically Classifying Edit Categories in Wikipedia Revisions
  - Johannes Daxenberger and Iryna Gurevych 2013
- Mining Naturally-occurring Corrections and Paraphrases from Wikipedia's Revision History
  - Aurélien Max and Guillaume Wisniewski 2010
- Mining Wikipedia's Article Revision History for Training Computational Linguistics Algorithms
  - Rani Nelken and Elif Yamangil 2008
- Sentence Simplification by Monolingual Machine Translation
  - Sander Wubben, Antal van den Bosch, and Emiel Krahmer 2012