

CMSC 132 H

Oct 12th, 2009

Bill Pugh

CS Colloquium

Thursday, 4pm, CSIC 1115

Meme-tracking, Scheduling, and the Dynamics of the News Cycle

Jon Kleinberg, Cornell University

The flow of news through on-line networks has created a complex landscape of media sources and led to rich datasets that provide glimpses into how news is produced, shaped, and consumed. We begin by discussing methods for studying how news stories spread through such a system, using an approach that tracks short pieces of as they travel and mutate across news sources. This type of analysis can be effective at capturing temporal patterns in the news over a daily time-scale --- in particular, the succession of story lines that evolve, compete for attention, and collectively produce an effect that commentators refer to as the 'news cycle.' We then show how a detailed analysis of temporal dynamics can suggest novel optimization problems in the scheduling of news stories and other on-line media. Specifically, given a supply of featured content and data on user attention over time, we consider how to sequence the content in a way that maximizes the size of the audience.

Code review

- Move to front code review
- Due today; some issues

Regrade requests

- Correctly enabled now in grades.cs.umd.edu?
- Any grade issues should get posted there, not email

Recursion, edit distance

- Going to spend most of today doing coding exercises related to recursion and edit distance

Making change

- How many ways can you make change for 10¢ using American coins?
- How many ways can you make change for 35¢?
 - number of ways you can make change for 10¢ plus number of ways you can make change for 35¢ without using any quarters

Making change

- For simple problems, direct recursion suffices
- But if you want to know the number of ways to make change for \$100, it doesn't
 - you need to use caching
 - you need to also need to either combine some iteration with your recursion, or expand the loop
 - recursive calls could go 10000 calls deep
 - Use `-Xss10m` to set aside 10 megabytes for stack

Edit distance

- Implementing a dynamic programming solution
- Cost is 1 to insert, delete or change a character
- Cost is also 1 to transpose two adjacent characters

Spell correction

- Read list of 21,877 English words
- For a string, generate the set of words that have minimal distance from the string
- those might be your spell correction suggestions

Example

- for “billpugh”, there are 21 words at distance 4
- ballpark, bigmouth, bill, billfold, billion, billionth, bilow, biology, blush, borough, bough, bulldog, bullfight, bullish, filling, killing, laugh, pillage, village, willful, willing
- Distance 4 is probably too high for spell correction suggestion
- How long do you think it would take to compute distance to 21,877 words?