More temporal dynamics
This week

Temporal models

This week we’ll look back on some of the topics already covered in this class, and see how they can be adapted to make use of temporal information

1. Regression – sliding windows and autoregression
2. Classification – dynamic time-warping
3. Dimensionality reduction - ?
4. Recommender systems – some results from Koren

Today:

1. Text mining – “Topics over Time”
2. Social networks – densification over time
Monday: Time-series regression

Also useful to plot data:

BeerAdvocate, ratings over time

Scatterplot

BeerAdvocate, ratings over time

Sliding window (K=10000)
seasonal effects
long-term trends

Code on:
http://jmcauley.ucsd.edu/cse190/code/week10.py
As you recall...
The longest-common subsequence algorithm is a standard dynamic programming problem

<table>
<thead>
<tr>
<th></th>
<th>-</th>
<th>A</th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

- = optimal move is to delete from 1\textsuperscript{st} sequence

- = optimal move is to delete from 2\textsuperscript{nd} sequence

↑ = either deletion is equally optimal

= optimal move is a match
Monday: Temporal recommendation

To build a reliable system (and to win the Netflix prize!) we need to account for **temporal dynamics**:

Netflix ratings by movie age

(People tend to give higher ratings to older movies)

Netflix ratings over time

(Netflix changed their interface)

---

Figure from Koren: “Collaborative Filtering with Temporal Dynamics” (KDD 2009)
yeast and minimal red body thick light a Flavor sugar strong quad. grape over is molasses lace the low and caramel fruit Minimal start and toffee. dark plum, dark brown Actually, alcohol Dark oak, nice vanilla, has brown of a with presence. light carbonation. bready from retention. with finish. with and this and plum and head, fruit, low a Excellent raisin aroma Medium tan

**Bags-of-Words**

What we would like:

87 of 102 people found the following review helpful

**Sentiment analysis**

Dimensionality reduction

(review of "The Chronicles of Riddick")
8. Social networks

Hubs & authorities

Power laws

Small-world phenomena

Strong & weak ties
9. Advertising

Matching problems

AdWords

Bandit algorithms
CSE 190 – Lecture 17
Data Mining and Predictive Analytics

Temporal dynamics of text
Bag-of-Words representations of text:

$$F_{\text{text}} = [150, 0, 0, 0, 0, 0, 0, ..., 0]$$

- a
- aardvark
- zoetrope
In week 5/7, we tried to develop low-dimensional representations of documents:

What we would like:

87 of 102 people found the following review helpful

⭐⭐⭐⭐⭐ You keep what you kill, December 27, 2004
By Schtinky "Schtinky" (Washington State) - See all my reviews

This review is from: The Chronicles of Riddick (Widescreen Unrated Director's Cut) (DVD)

Even if I have to apologize to my friends and favorites, and my family, I have to admit that I really liked this movie. It's a Sci-Fi movie with a "Mad Maxx" appeal that, while changing many things, left Riddick from 'Pitch Black' to be just Riddick. They did not change his attitude or soften him up or bring him out of his original character, which was very pleasing to 'Pitch Black' fans like myself.

First off, let me say that when playing the DVD, the first selection to come up is Convert or Fight, and no explanation of the choices. This confused me at first, so I will mention off the bat that they are simply different menu formats, that each menu has the very same options, simply different background visuals. Select either one and continue with the movie.

(review of “The Chronicles of Riddick”)
Topics over Time (Wang & McCallum, 2006) is an approach to incorporate temporal information into low-dimensional document representations

e.g.
• The topics discussed in conference proceedings progressed from neural networks, towards SVMs and structured prediction (and back to neural networks)
• The topics used in political discourse now cover science and technology more than they did in the 1700s
• With in an institution, e-mails will discuss different topics (e.g. recruiting, conference deadlines) at different times of the year
**Topics over Time** (Wang & McCallum, 2006) is an approach to incorporate temporal information into low-dimensional document representations.

- Timestamps \( t_{di} \) are drawn from Beta(\( \psi_{z_{di}} \)).

- There is now one Beta distribution **per topic**

Beta distributions are a flexible family of distributions that can capture several types of behavior – e.g. gradual increase, gradual decline, or temporary “bursts”.

\[
p.d.f: \quad x^{\alpha - 1} (1-x)^{\beta - 1} \quad \frac{1}{B(\alpha, \beta)}
\]
Results:
Political addresses – the model seems to capture realistic “bursty” and gradually emerging topics

<table>
<thead>
<tr>
<th>Mexican War</th>
<th>Panama Canal</th>
<th>Cold War</th>
<th>Modern Tech</th>
</tr>
</thead>
<tbody>
<tr>
<td>states 0.02032</td>
<td>government 0.02928</td>
<td>world 0.01875</td>
<td>energy 0.03902</td>
</tr>
<tr>
<td>mexico 0.01832</td>
<td>united 0.02132</td>
<td>states 0.01717</td>
<td>national 0.01534</td>
</tr>
<tr>
<td>government 0.01670</td>
<td>islands 0.01167</td>
<td>security 0.01710</td>
<td>development 0.01448</td>
</tr>
<tr>
<td>united 0.01521</td>
<td>canal 0.01014</td>
<td>soviet 0.01664</td>
<td>space 0.01436</td>
</tr>
<tr>
<td>war 0.01059</td>
<td>american 0.00872</td>
<td>united 0.01491</td>
<td>science 0.01227</td>
</tr>
<tr>
<td>congress 0.00951</td>
<td>cuba 0.00834</td>
<td>nuclear 0.01454</td>
<td>technology 0.01227</td>
</tr>
<tr>
<td>country 0.00906</td>
<td>made 0.00747</td>
<td>peace 0.01408</td>
<td>oil 0.01178</td>
</tr>
<tr>
<td>texas 0.00852</td>
<td>general 0.00731</td>
<td>nations 0.01069</td>
<td>make 0.00994</td>
</tr>
<tr>
<td>made 0.00727</td>
<td>war 0.00660</td>
<td>international 0.01024</td>
<td>effort 0.00969</td>
</tr>
<tr>
<td>great 0.00611</td>
<td></td>
<td>america 0.00987</td>
<td>administration 0.00957</td>
</tr>
</tbody>
</table>
Latent Dirichlet Allocation

Results:
e-mails & conference proceedings
Results:
conference proceedings (NIPS)

Relative weights of various topics in 17 years of NIPS proceedings
http://people.cs.umass.edu/~mccallum/papers/tot-kdd06.pdf
CSE 190 – Lecture 17
Data Mining and Predictive Analytics

Temporal dynamics of social networks
How can we characterize, model, and reason about the structure of social networks?

1. Models of network structure
2. Power-laws and scale-free networks, “rich-get-richer” phenomena
3. Triadic closure and “the strength of weak ties”
4. Small-world phenomena
5. Hubs & Authorities; PageRank
Temporal dynamics of social networks

Two weeks ago we saw some processes that model the generation of social and information networks

- Power-laws & small worlds
- Random graph models

These were all defined with a “static” network in mind. But if we observe the order in which edges were created, we can study how these phenomena change as a function of time.

First, let’s look at “microscopic” evolution, i.e., evolution in terms of individual nodes in the network.
Q1: How do networks grow in terms of the number of nodes over time?

(From Leskovec, 2008 (CMU Thesis))

A: Doesn’t seem to be an obvious trend, so what do networks have in common as they evolve?
**Q2:** When do nodes create links?
- *x*-axis is the age of the nodes
- *y*-axis is the number of edges created at that age

**A:** In most networks there’s a “burst” of initial edge creation which gradually flattens out. Very different behavior on LinkedIn (guesses as to why?)
**Q3:** How long do nodes “live”?  
- x-axis is the difference between date of last and first edge creation  
- y-axis is the frequency

**A:** Node lifetimes follow a power-law: many many nodes are shortlived, with a long-tail of older nodes
Temporal dynamics of social networks

What about “macroscopic” evolution, i.e., how do global properties of networks change over time?

Q1: How does the # of nodes relate to the # of edges?

- A few more networks: citations, authorship, and autonomous systems (and some others, not shown)
- A: Seems to be linear (on a log-log plot) but the number of edges grows faster than the number of nodes as a function of time
Q1: How does the # of nodes relate to the # of edges?

A: seems to behave like

\[ E(t) \propto c N(t)^a \]

where

\[ 1 \leq a \leq 2 \]

- a = 1 would correspond to **constant** out-degree – which is what we might traditionally assume
- a = 2 would correspond to the graph being fully connected
- What seems to be the case from the previous examples is that a > 1 – the number of edges grows faster than the number of nodes
Q2: How does the degree change over time?

- A: The average out-degree increases over time.
Q3: If the network becomes **denser**, what happens to the (effective) diameter?

- **A:** The diameter seems to decrease.
- In other words, the network becomes **more** of a small world as the number of nodes increases.
Q4: Is this something that **must** happen – i.e., if the number of edges increases faster than the number of nodes, does that mean that the diameter must decrease?

A: Let’s construct random graphs (with \( a > 1 \)) to test this:

- **Erdos-Renyi** – \( a = 1.3 \)
- **Pref. attachment model** – \( a = 1.2 \)
So, a decreasing diameter is not a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model.

**Q5:** is the degree distribution of the nodes sufficient to explain the observed phenomenon?

**A:** Let’s perform random rewiring to test this.

Random rewiring preserves the degree distribution, and randomly samples amongst networks with observed degree distribution.
Temporal dynamics of social networks

So, a decreasing diameter is not a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model.

Q5: is the degree distribution of the nodes sufficient to explain the observed phenomenon?

(c) Affiliation network (ATP-ASTRO-PH)  
(d) US patent citation network (CIT-PATENTS)
So, a decreasing diameter is not a “rule” of a network whose number of edges grows faster than its number of nodes, though it is consistent with a preferential attachment model.

Q5: is the degree distribution of the nodes sufficient to explain the observed phenomenon?

A: Yes! The fact that real-world networks seem to have decreasing diameter over time can be explained as a result of their degree distribution and the fact that the number of edges grows faster than the number of nodes.
Temporal dynamics of social networks

Other interesting topics...

"memetracker"
Alignment query data with disease data – Google flu trends: https://www.google.org/flutrends/us/#US

Questions?

Further reading:
“Dynamics of Large Networks” (most plots from here)
Jure Leskovec, 2008

“Microscopic Evolution of Social Networks”
Leskovec et al. 2008

“Graph Evolution: Densification and Shrinking Diameters”
Leskovec et al. 2007
Some incredible assignments
Bike Stalking

Charles McKay and Kimberly Ly

- Predict the **end location** of a bicycle commute
- Use regression to predict lat/lon, and map it to a station
- Features based on location, distance, time (hour/day)
Predicting Censorship on Weibo

Brian Tsay and John Kuk

• Predict whether a tweet will be censored based on its content
• Features based on the user, retweets, and daily censorship
Wordles!

Amazon Video Games: Alexander Ishikawa

Wine: Alexander Ishikawa
Wordles!

Shashank Uppoor and Shreyas Pathre Balakrishna

• Predict **hygiene scores** on Yelp from text
Energy Demand Prediction

Shubham Saini, Jonathan Cervantes, Vyom Shah, Kenneth Vuong

- Energy consumption data from 6 houses
- Forecast next-day power use
- Weather, time, clustering, appliances, occupancy
Crime Type Prediction

Jeffery Wang, Jesse Gallaway, Matthew Schwegler

- Predict crime type (statutory, property, personal)
- Features based on lat/lon, day/night, population, streetlamp distance (!), and clustering
Is There a Time for Crime?

David Thomasson

- **Use only** temporal data to forecast crimes
- (Saturdays+Sundays), (Hours 1,2,3,4,18,19,20,21,22,23), (January, March, December) are +‘ve for crime
Ho-Wei Kang

- Predict post popularity on reddit
- Features include author, time, title, content, comments, age, gender
- Other related projects included predicting response time in long-distance relationships, and predicting “view changes” in /r/changemyview
Fill out those evaluations!

- Please evaluate the course on http://cape.ucsd.edu/students!
Want more data mining?

- I am running a workshop on “Big Graphs” on January 6-8
- **Registration (and lunch) is free!**
- See [http://cseweb.ucsd.edu/~slovett/workshops/big-graphs-2016/](http://cseweb.ucsd.edu/~slovett/workshops/big-graphs-2016/)
Thanks!