Overview of Machine Learning Research

Anima Anandkumar

U.C. Irvine

Learning with Big Data

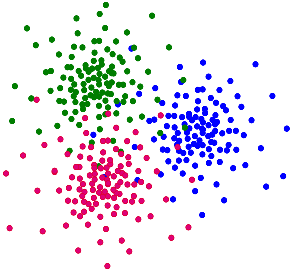
High Dimensional Regime

- Missing observations, gross corruptions, outliers, ill-posed problems.
- Needle in a haystack: finding low dimensional structures in high dimensional data.

Principled approaches for finding low dimensional structures?

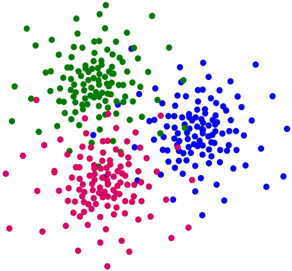
Application 1: Clustering

- Basic operation of grouping data points.
- Hypothesis: each data point belongs to an unknown group.



Application 1: Clustering

- Basic operation of grouping data points.
- Hypothesis: each data point belongs to an unknown group.



Probabilistic/latent variable viewpoint

- The groups represent different distributions. (e.g. Gaussian).
- Each data point is drawn from one of the given distributions. (e.g. Gaussian mixtures).

Application 2: Topic Modeling

JANINE DI GIOVANNI
The New York Times - Thursday, October 25, 2012

Bleary-Eyed Syrian Troops Fight a Building at a Time

HOMS, Syria — For more than 24 hours, President Bashar al-Assad's soldiers fought their way through this city, bleary-eyed men, worn down by months of combat. Afraid to go into the streets, where snipers pick their targets, the government men snaked their way through "mouse holes" punched in walls of blown-out buildings. Their goal was to retake one building, just one, a former school controlled by the rebel Free Syrian Army.

"We will eventually get this school," said Rifaf, part of a small group of soldiers on the mission. "But it's a matter of time."

In many ways, Homs serves as a microcosm of Syria, a community of Sunnis, Shiites, Alawites and Christians who lived side by side, only to find that the collapse of the police state quickly dissolved any sense of common purpose. Civil war came to Homs early, and it has ground on ever since, one street, one building, one apartment at a time. It has been a seesaw, with the opposition gaining ground, then the government taking it back.

That is the story of the school, this group of government soldiers and a nation caught in a contest of attrition. Half of Homs exists as it did before, where some Christians and Alawites, allied with President Assad, still cling to a normal routine, visit cafes and enjoy strolling the streets.

The other half is rubble.

"See the snipers? They have been shooting at us all day," Rifaf said as he took cover in an abandoned apartment, crouched beneath a window as the familiar crack of automatic gunfire echoed in the courtyard.

"The sniper is there," he said, nodding toward a blasted-out building a few blocks away. Like the others

interviewed for this a
identified for securi

That was the sch
a weapons dep

Homs, Syri
around th
play a lo
government

The government's men
opposition, at least as fighte

Bab al-Sebaa Street, since a
desired address, is in ruins. The
New York Times

The New York Times

The News-Review

Roseburg, Oregon Vol. 146 No. 166 MONDAY, NOVEMBER 12, 2012 www.nrtoday.com 75 cents

Callers to 211 seek help with food, rent

Information line has 150 users per month, but caller sample shows mixed results

CARISA CEGAVSKE
The News-Review

Since the 211 information line for service launched in May, about 150 Douglas County residents have reportedly called each month.

Jared Paben, 211info resource manager, said he hopes to boost those numbers by spreading the word that 211 referrals are available. At the same time, he expressed concern that a sampling of callers showed more than half those who called the information line reported they still weren't able to get the help they needed.

Callers are connected with a Portland adviser who searches a database of 112 Douglas County nonprofit and government organizations that may help them. The nonprofit group 211info, which runs the 211 call center for Oregon and Southwest Washington, recently released statistics about who has been using the service in Douglas County.

Nearly three-quarters of the county's 211 callers are women. More than four in five are living below the federal poverty level. The top two requests were for help with food and shelter. Twelve percent of callers are looking for help putting food on the table and 9 percent want help paying the rent. Other top requests are for help with heat bills and health care.

Paben is responsible for updating lists of resources in Douglas and Josephine counties.

He said the number of monthly callers has remained steady since May. He said he hopes more people will learn about the 211 line and use it to find the services they need.

"It's stayed consistent, but I was hoping the numbers would go even higher," he said.

Paben said he anticipates more calls this winter when people need help purchasing Christmas gifts and paying for heat.

Not all the callers' needs were met. A follow-up with a sample group of callers found that even after the referral, 62 percent didn't get the help they needed. Half of those said they were told funds were gone or services full at the agencies to which they were referred.

Turn to **211**, page A7

Corrections

Douglas County has more 211 agency listings than Linn or Yamhill counties. It has fewer than Deschutes County. The comparisons were incorrect in an article on page A1 Monday.

The News-Review attempts to quickly and prominently correct mistakes that have appeared in the paper. Errors may be brought to the attention of the city editor by calling 541-957-4201 or by email at newsdesk@nrtoday.com

211: Douglas County trails others in info line usage

Continued from page A1

United Community Action Network and Greater Douglas United Way worked with phone companies to make the 211 service available in Douglas County.

UCAN is one of the top five agencies to which Douglas County callers are referred. The other four were the Salvation Army, FISH food pantry of Roseburg, St. Vincent de Paul and the Roseburg Dream Center, an outreach ministry of the New Life Christian Center that provides clothing and food to those in need.

Some calls were for services for which 211 wasn't able to recommend a provider. Unmet needs included bus fare and school clothes and supplies.

Douglas County is trailing behind many counties in 211 line usage, with one call for every 703 residents, Paben said. Deschutes County had about one call for every 473 people and Yamhill County had one for every 483. He said Douglas County has more service agencies than either Deschutes or Yamhill counties.

"These comparisons tell me we might be a little behind the curve in terms of usage and I probably need to step up my efforts getting the word out about 211," Paben said. "We know 211 and the agencies in our database could help lots of people in our community. The key is going to be letting them know it's now available to them."

Information about resources and service is also available at 211info.org. Visits to the Douglas County portion of the website have averaged about 235 per month.

• You can reach reporter Carisa Cegavske at 541-957-4213 or ccegavske@nrtoday.com.

Hippo eats dwarf

A HIPPOPOTAMUS has swallowed a dwarf in a circus accident in northern Thailand.

"A dwarf, nicknamed Od, died when he bounced sideways from a trampoline and was swallowed by a yawning hippopotamus, which was waiting to appear in the next act," the *Pattaya Mail* reported.

"Vets on the scene said Hilda the Hippo had a gag reflex that automatically caused her to swallow."

The vets said it was the first time the hefty vegetarian had ever eaten a circus performer.

"Unfortunately, the 1000-plus spectators continued to applaud wildly until commonsense dictated there had been a tragic mistake.

"Police said the trampoline had been sent for forensic analysis."

Document modeling

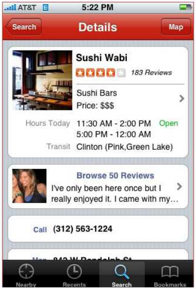
- Observed: words in document corpus.
- Hidden: topics.
- Goal: carry out document summarization.

Application 3: Understanding Human Communities

Social Networks

- Observed: network of social ties, e.g. friendships, co-authorships
- Hidden: groups/communities of actors.

Application 4: Recommender Systems



Sushi Wabi

183 Reviews

Sushi Bars

Price: $$$

Hours Today 11:30 AM - 2:00 PM Open

5:00 PM - 12:00 AM

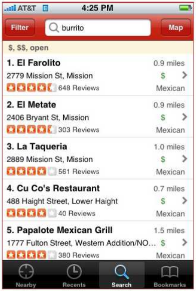
Transit Clinton (Pink,Green Lake)

Browse 50 Reviews

I've only been here once but I really enjoyed it. I came with my…

Call (312) 563-1224

Map 842 W Randolph St



$, $$, open

1. El Farolito 2779 Mission St, Mission 648 Reviews	0.9 miles $	Mexican
2. El Metate 2406 Bryant St, Mission 303 Reviews	0.9 miles $	Mexican
3. La Taqueria 2889 Mission St, Mission 561 Reviews	1.0 miles $	Mexican
4. Cu Co's Restaurant 488 Haight Street, Lower Haight 40 Reviews	0.7 miles $	Mexican
5. Papalote Mexican Grill 1777 Fulton Street, Western Addition/NO… 380 Reviews	1.5 miles $	Mexican

Recommender System

- Observed: Ratings of users for various products, e.g. yelp reviews.
- Goal: Predict new recommendations.
- Modeling: Find groups/communities of users and products.

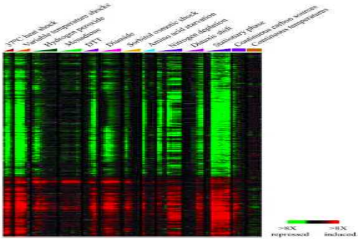
Application 5: Feature Learning

$$\begin{array}{cc} \text{Label} & \text{Features} \\ \left[\begin{array}{c|ccccc} 0 & 2.1 & 5.2 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 2 & 1 & \cdots \\ 1 & 1.1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 7 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{array}\right] \end{array}$$

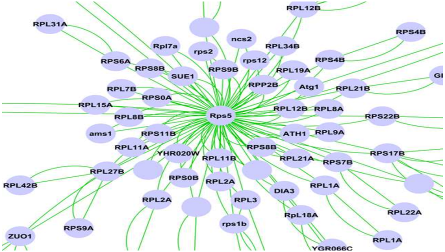
Feature Engineering

- Learn good features/representations for classification tasks, e.g. image and speech recognition.
- Sparse representations, low dimensional hidden structures.

Application 6: Computational Biology

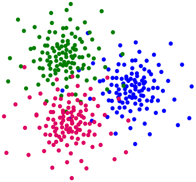


Gasch *et al.* *Mol Biol Cell* 2000.



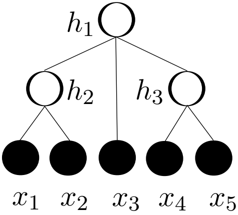
- Observed: gene expression levels
- Goal: discover gene groups
- Hidden variables: regulators controlling gene groups

How to model hidden effects?



Basic Approach: mixtures/clusters

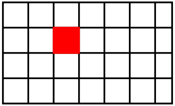
- Hidden variable h is categorical.



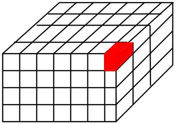
Advanced: Probabilistic models

- Hidden variable h has more general distributions.
- Can model mixed memberships.

Learning Algorithms through Tensor Factorization



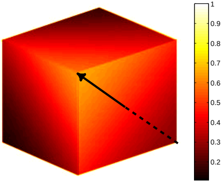
vs.



- Co-occurrence of three-words in a document, e.g. [apple, orange, banana].

Tensor Eigenvectors

- Can learn the hidden topics by finding tensor eigenvectors.
- Common friends (neighbors) of triplets of nodes in a social networks.

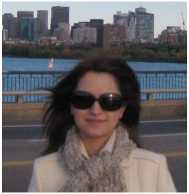


My Research Group

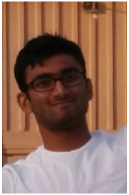
Furong Huang

Majid Janzamin

Hanie Sedghi



Niranjan UN



Forough Arabshahi

Resources and Course Information

Resources and Course Information

- ML summer school lectures available at `http://newport.eecs.uci.edu/anandkumar/MLSS.html`
- Publications at `http://newport.eecs.uci.edu/anandkumar/`

Resources and Course Information

- ML summer school lectures available at `http://newport.eecs.uci.edu/anandkumar/MLSS.html`
- Publications at `http://newport.eecs.uci.edu/anandkumar/`

Courses

- EECS 298: Large scale ML: theory and practice.
  - Cloud-based programming, spectral and tensor methods. Hadoop framework.
- EECS 298 (formerly 251B): Statistical learning theory
  - Theoretical course. Non-parametrics, optimization, regularization, concentration bounds.