Machine Learning: CIS 472/572
Introduction

Instructor: Thien Huu Nguyen

Based on slides by Daniel Lowd, Vibhav Gogate, Pedro Domingos, and others.

Logistics

- **Instructor:** Thien Huu Nguyen
  - Email: thien@cs.uoregon.edu
  - Office: Deschutes 330
  - Office hours: Tuesdays, Thursdays (3:00-4:00pm via Zoom)
- **Teaching Assistant:** Viet Lai
  - Email: vietl@uoregon.edu
  - Office hours: Mondays, Wednesdays (4:00-5:00pm via Zoom)
- **Web:** http://www.cs.uoregon.edu/Classes/21S/cis472/
- **Discussion Board:** Piazza (link on web page)
  Please use this for discussion!

Evaluation

- 3-4 assignments (50%)
  - Some programming, some exercises
- One Midterm (20%)
  - 2/3rds of the way through
- One Project (30%)
  - Apply machine learning to a real problem of your choice *(Recommended: Participate in a contest on Kaggle.com.)*
  - Groups allowed (for undergrad students)
  - Written report
  - Presentations during final exam time (still considering)

Kaggle

Active Competitions

All Competitions

17 found, 17 active

Search competitions

- **All competitions**
- Enterable

Status

- **Active**
- Completed

Sponsor

- InClass (student competition)

Competition Name	Reward	Teams	Deadline
National Data Science Bowl Predict ocean health, one plankton at a time	$175,000	312	2 months
Driver Telematics Analysis Use telematic data to identify a driver signature	$30,000	453	2 months
Helping Santa's Helpers Jingle bells, Santa tells ...	$20,000	439	2.1 days
Click-Through Rate Prediction Predict whether a mobile ad will be clicked	$15,000	1203	35 days
BCI Challenge @ NER 2015 A spell on you if you cannot detect errors!	$1,000	161	50 days
Sentiment Analysis on Movie Reviews Classify the sentiment of sentences from the Rotten Tomatoes dataset	Knowledge	675	54 days
Finding Elo Predict a chess player's FIDE Elo rating from one game	Knowledge	100	2 months

Source Materials

Primary source of readings:

- H. Daume III, ***A Course in Machine Learning v0.9***. http://ciml.info
  (Good intro, focuses on machine learning concepts before math.
  Free online. Not finished.)

Excellent supplements:

- K. Murphy, ***Machine Learning: A Probabilistic Perspective***, MIT Press, 2012. (Great reference and in-depth coverage.)
- T. Mitchell, ***Machine Learning,***
  McGraw-Hill, 1997. (Great intro, but old and expensive.)
- Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016. (Good source for deep learning, free online: https://www.deeplearningbook.org/)

- C. Bishop, ***Pattern Recognition and Machine Learning***, Springer, 2006
- R. Duda, P. Hart & D. Stork, ***Pattern Classification*** (2nd ed.), Wiley, 2000

- D. Barber, ***Bayesian Reasoning and Machine Learning,*** Cambridge University Press, 2012. (Free online!)
- T. Hastie, R. Tibshirani, J. Friedman, ***The Elements of Statistical Learning***, Springer, 2009. (Free online!)

Why Study Machine Learning: A Few Quotes

"A breakthrough in machine learning would be worth ten Microsofts."
-Bill Gates, Microsoft

"Machine learning is the next Internet."
-Tony Tether, Former Director, DARPA

"Machine learning is the hot new thing."
-John Hennessy, President, Stanford

"Web rankings today are mostly a matter of machine learning."
-Prabhakar Raghavan, Dir. Research, Yahoo

"Machine learning is going to result in a real revolution."
-Greg Papadopoulos, CTO, Sun

So What Is Machine Learning?

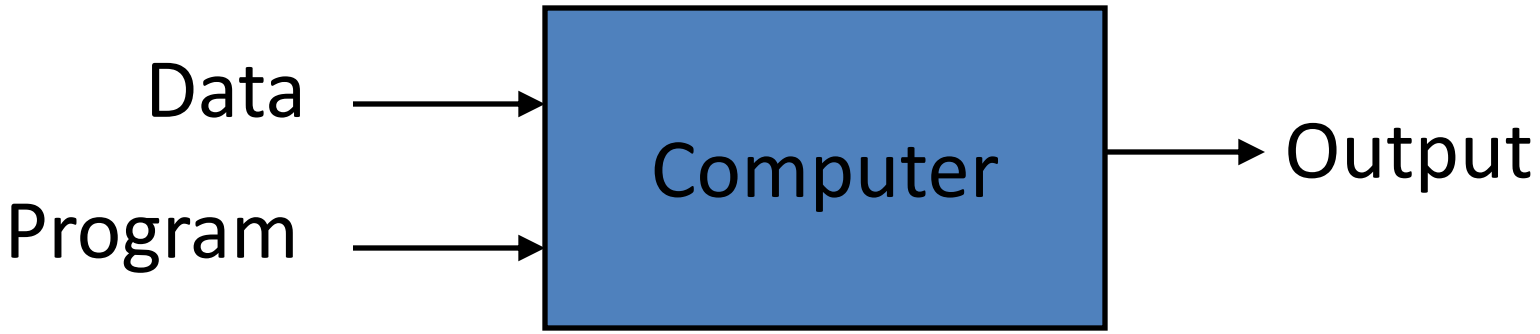
- Learning is acquiring and improving performance through experience.
- H. Simon: Any process by which a system improves its performance
- M. Minsky: Learning is making useful changes in our minds
- R. Michalsky: Learning is constructing or modifying representations of what is being experienced
- L. Viliant: Learning is the process of knowledge acquisition in the absence of explicit programming

So What Is Machine Learning?

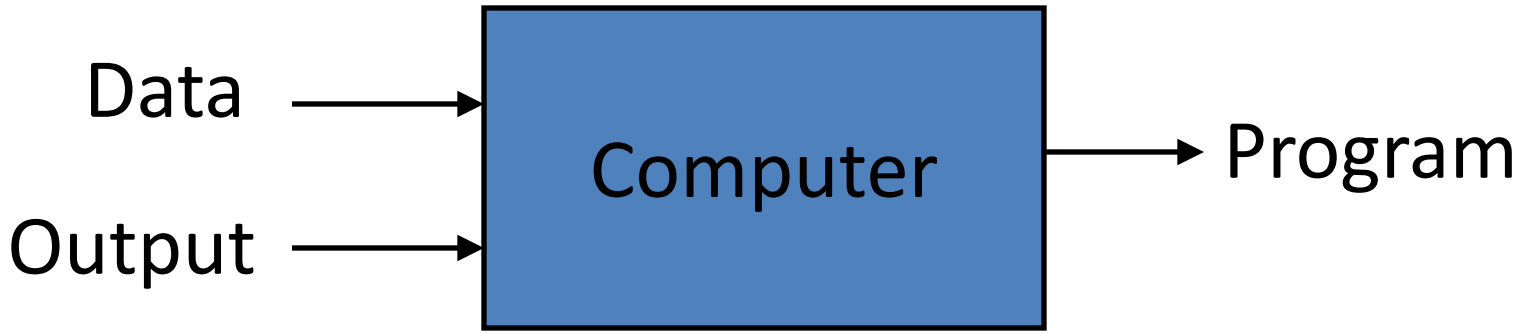
- What does it mean for a computer to learn? Why would we want them to learn? How do we get them to learn?
- We want computers to learn when it is too difficult or too expensive to program them directly to perform a task → Automating automation
- Get the computer to program itself by showing examples of inputs and outputs.
- In reality: we will write a "parameterized" program, and let the learning algorithm find the set of parameters that best approximates the desired function or behavior → Let the data do the work!

Traditional Programming

- Automating automation
- Getting computers to program themselves



Machine Learning



Another view on machine learning

- Machine learning = automated science (sort of)
- Goal is to go from raw data to useful knowledge
- An ML algorithm finds a theory to fit the data and background knowledge as well as possible.
- A theory is good if it has good predictive accuracy.

Related Fields

- Fields that use machine learning:
  - Artificial intelligence
  - Computer vision: object detection, tracking, segmentation, …
  - Natural language processing: machine translation, question answering, text generation, …
  - Computational biology: sequence alignment, information extraction, …
  - Robotics: state estimation, self-driving, …
  - Medicine: medical diagnosis, treatment prediction, drug design, …
  - Financial Analysis: portfolio allocation, option pricing (many people on Wall Street use machine learning)

Related Fields

- Fields with similar goals to machine learning:
  - Statistical Estimation
  - Data mining
  - Data science
  - Psychology (developmental, cognitive)
- Fields used by machine learning:
  - Information theory
  - Numerical optimization
  - Computational complexity

Definition: Machine Learning!

- T. Mitchell: Improving performance via experience
  - A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T as measured by P, improves with experience.

Example 1: A Chess learning problem

- Task T: playing chess
- Performance measure P: percent of games won against opponents
- Training Experience E: playing practice games against itself

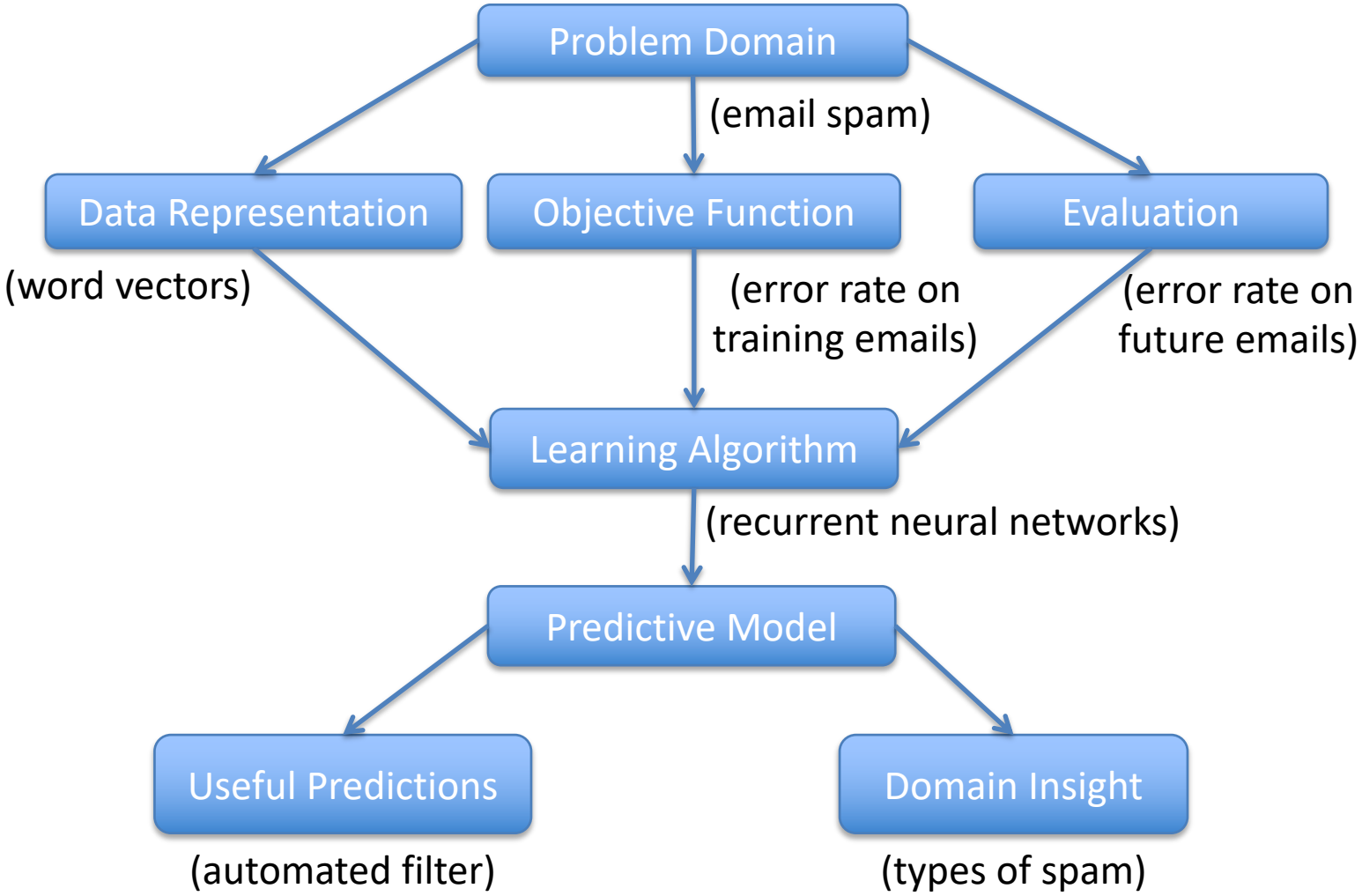
Example 2: Autonomous Vehicle Problem

- Task T: driving on a public highway/roads using vision sensors
- Performance Measure P: percentage of time the vehicle is involved in an accident
- Training Experience E: a sequence of images and steering commands recorded while observing a human driver

Example 3: Machine Translation

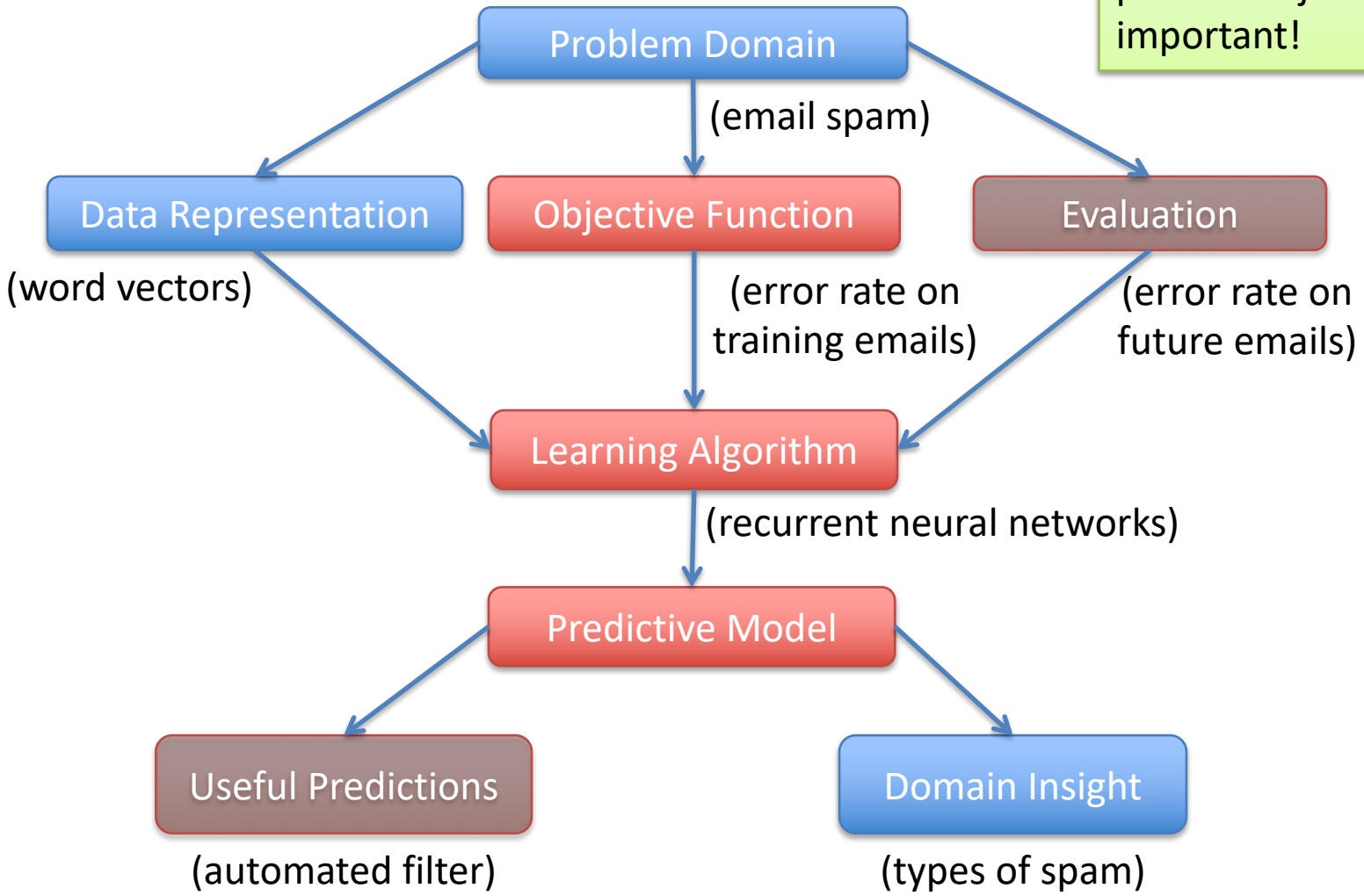
- Task T: translating sentences from one language to another language
- Performance Measure P: the similarities between the machine-translated sentences and the human-translated sentences (BLUE scores)
- Training Experience E: pairs of translated sentences between languages provided by human

The Big Picture



The Big Picture

This class mostly focuses on the red pieces, but the blue pieces are just as important!



ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Some algorithms are specific to application domains/data
- Every machine learning algorithm has three components:
  - **Model Representation**
  - **Evaluation**
  - **Optimization**

Model Representation

- Decision trees
- Instance-based models
- Linear function (hyperplane)
- Neural networks/Deep learning
- Support vector machines
- Kernel methods
- Model ensembles
- (Sets of rules / Logic programs)
- (Graphical models (Bayes/Markov nets))
- Etc.

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- KL divergence
- Etc.

Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

Types of Learning

- **Based on the information available**
  - **Supervised (inductive) learning**
    - Training data includes desired outputs
  - **Unsupervised learning**
    - Training data does not include desired outputs
    - Find hidden structure in data
  - **Semi-supervised learning**
    - Training data includes a few desired outputs
  - **Reinforcement learning**
    - The learner interacts with the world via "actions" and tries to find an optimal policy of behavior with respect to "rewards" it receives from the environment
- **Based on the role of the learner**
  - **Passive Learning**
    - Training solely relies on data
  - **Active Learning**
    - Training additionally involves interactions with human

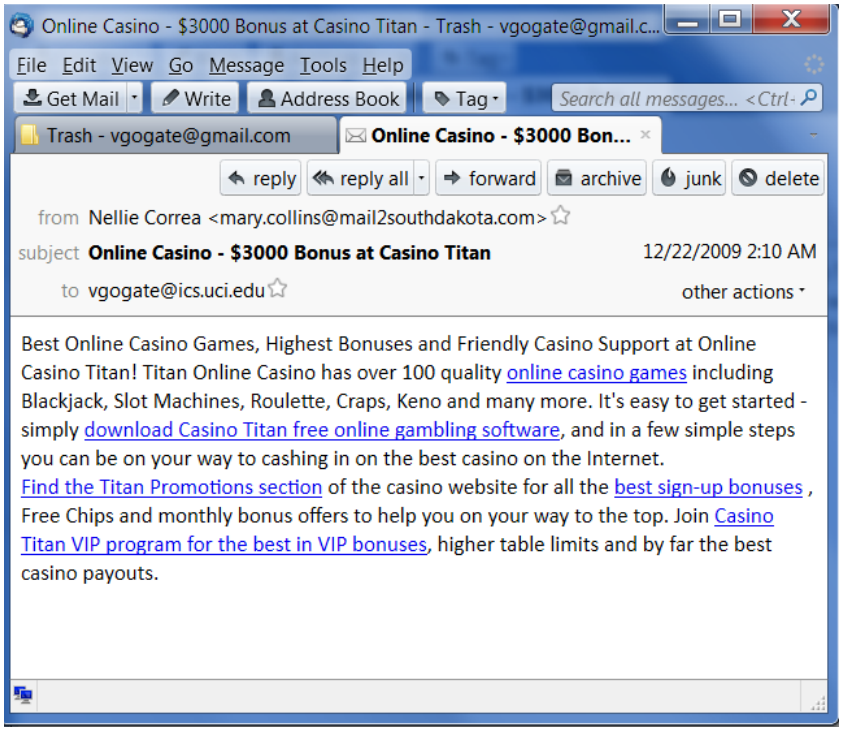
Types of Supervised Learning Problems

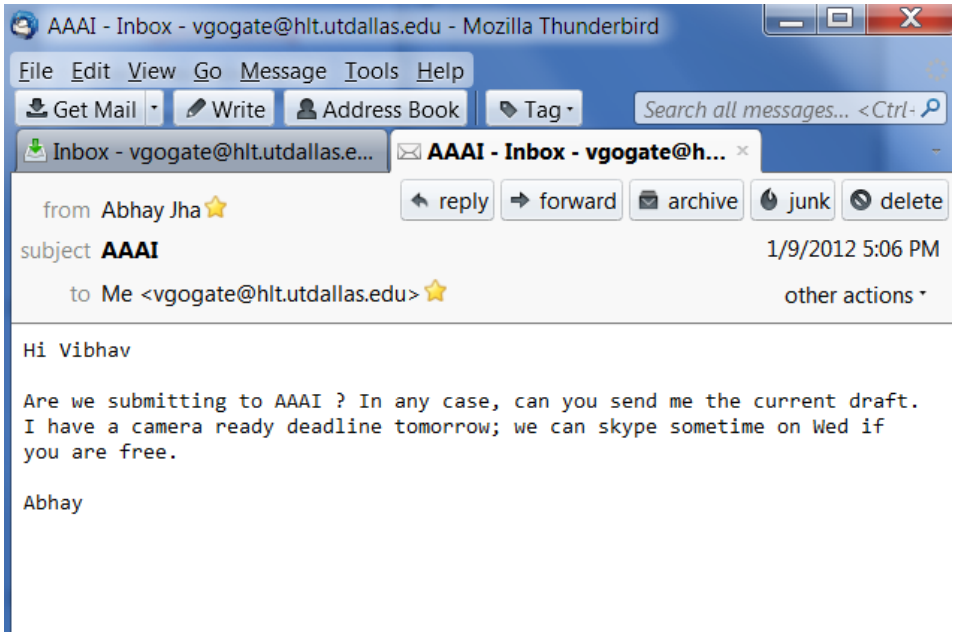
- **Classification:** predict a discrete/categorical value from a predefined set of values
- **Regression:** predict a continuous/real value
- **Structured prediction:** predict a complex output, such as a sequence or tree

Machine Learning: Applications

Examples of what you will study in class in action!

Classification Example: Spam Filtering





Classify as "Spam" or "Not Spam"

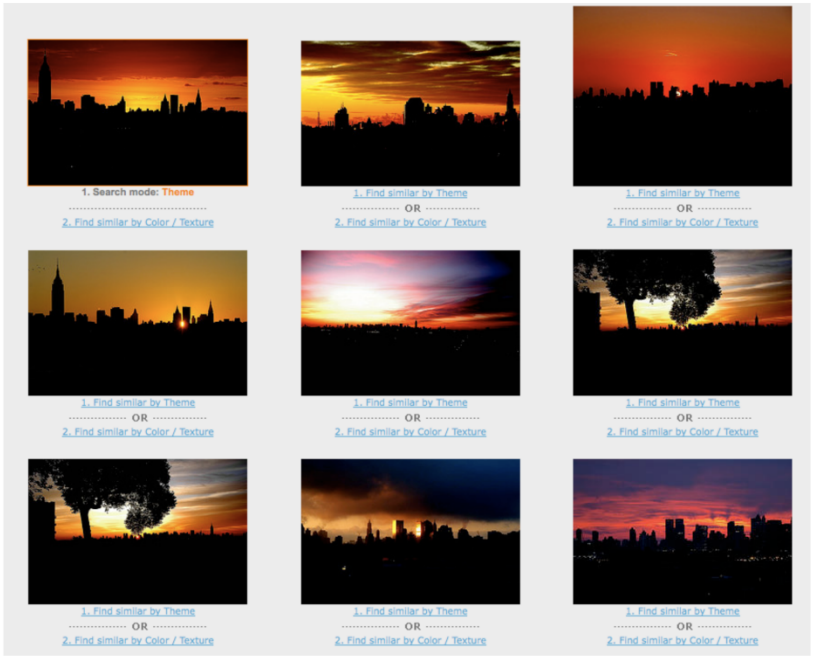
Classification Example: Weather Prediction

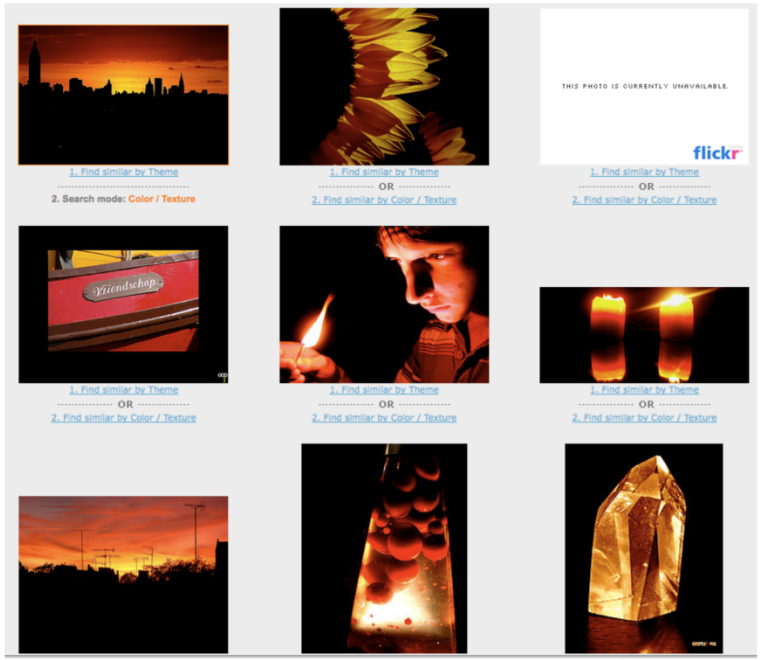
Regression example: Predicting Gold/Stock prices

Good ML can make you rich (but there is still some risk involved).

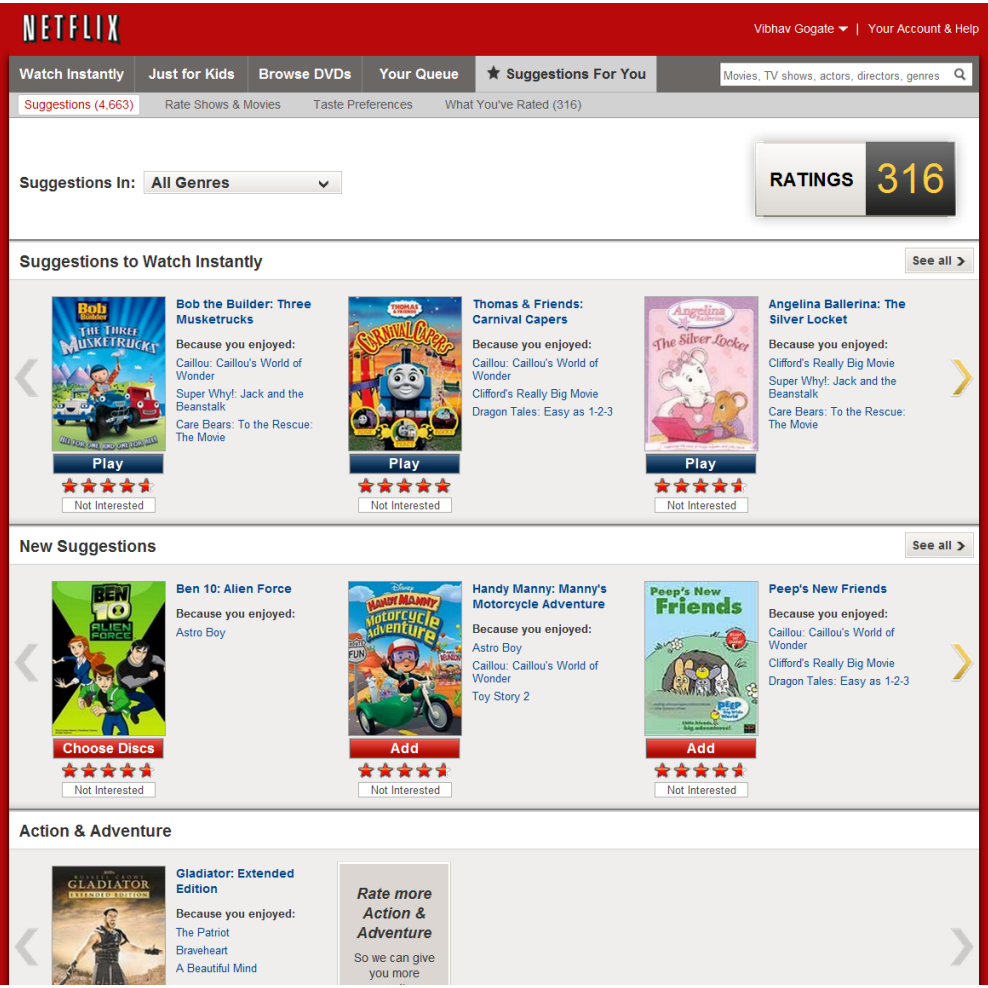
Given historical data on gold prices, predict tomorrow's price!

Similarity Determination

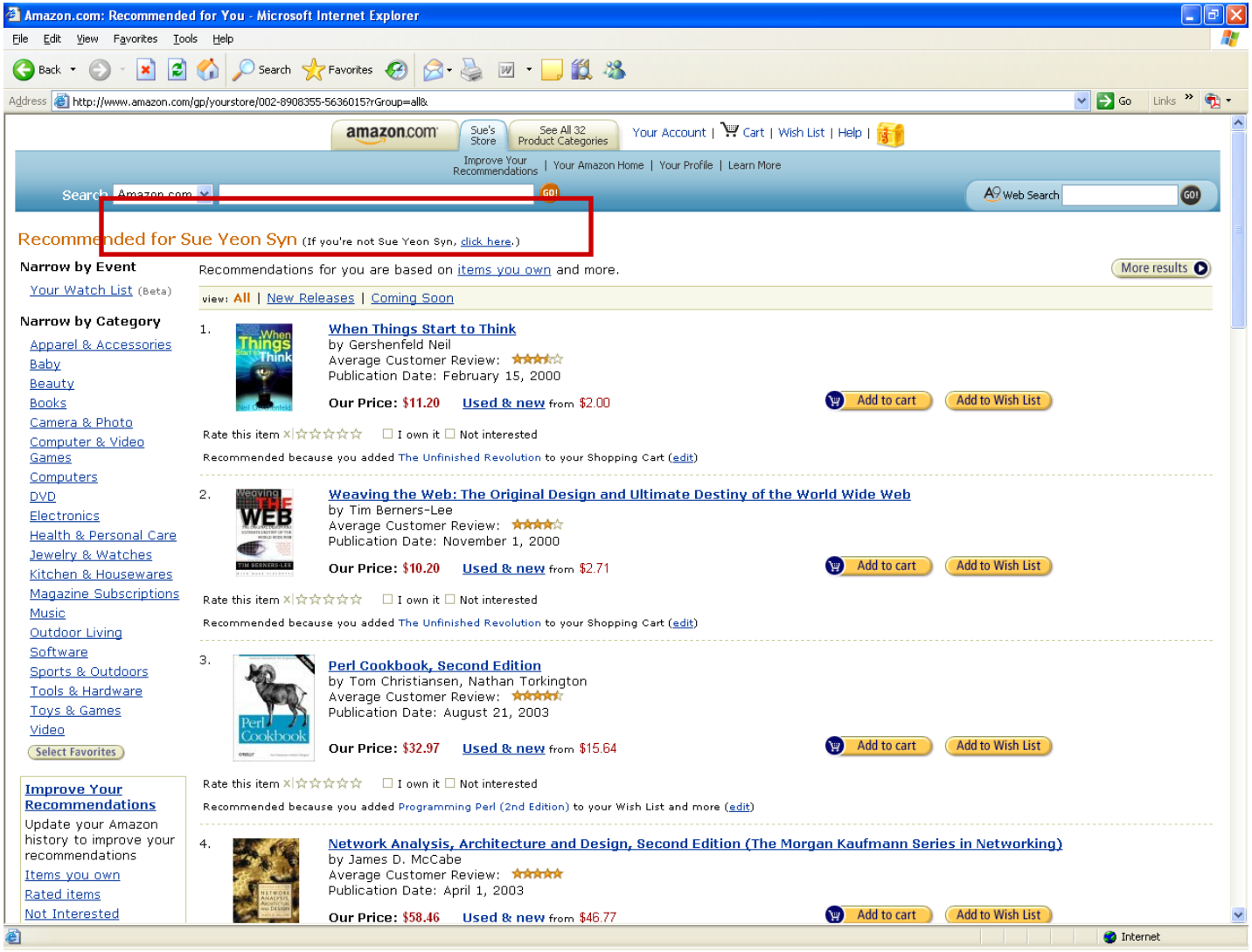




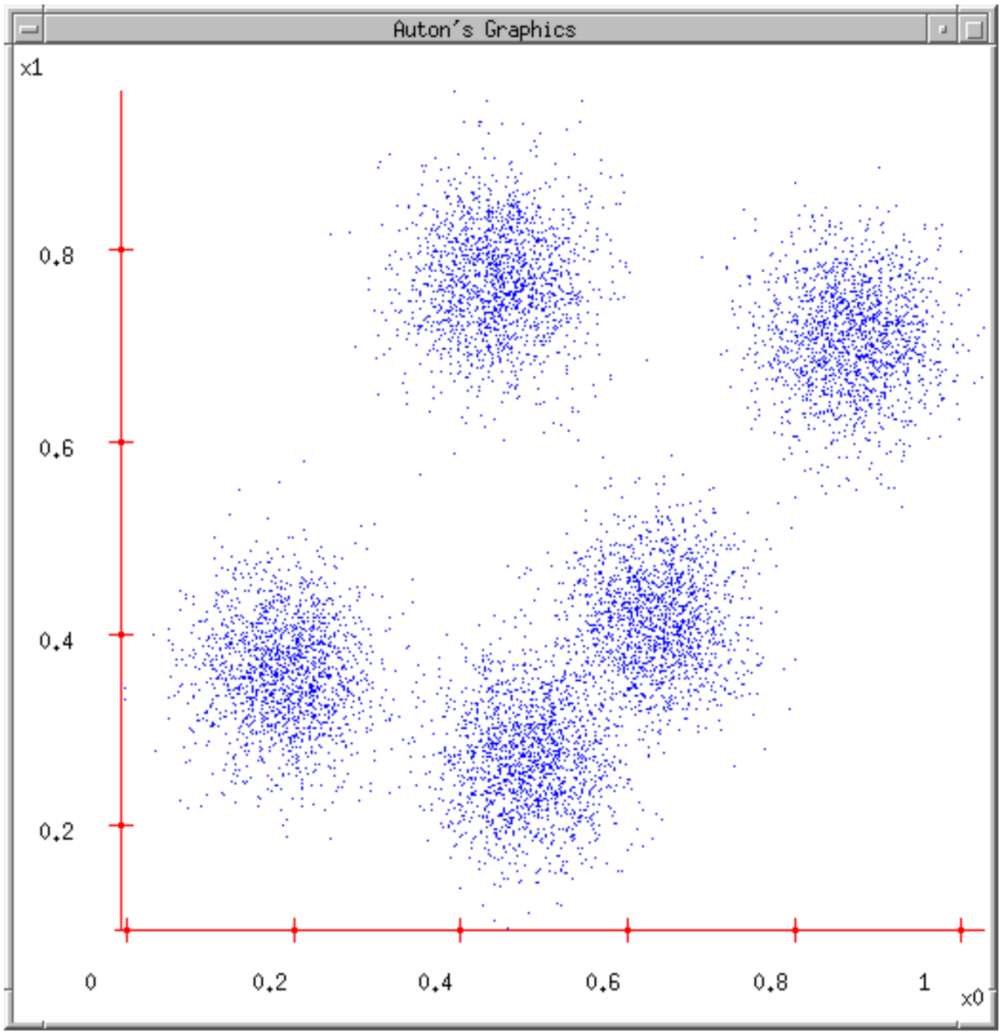
Collaborative Filtering



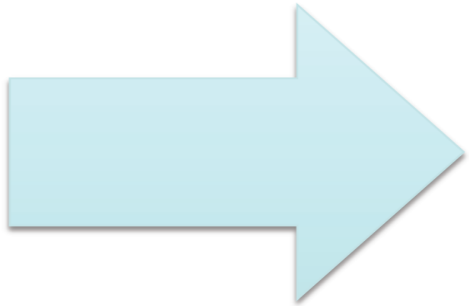
Collaborative Filtering



Clustering: Discover Structure in data

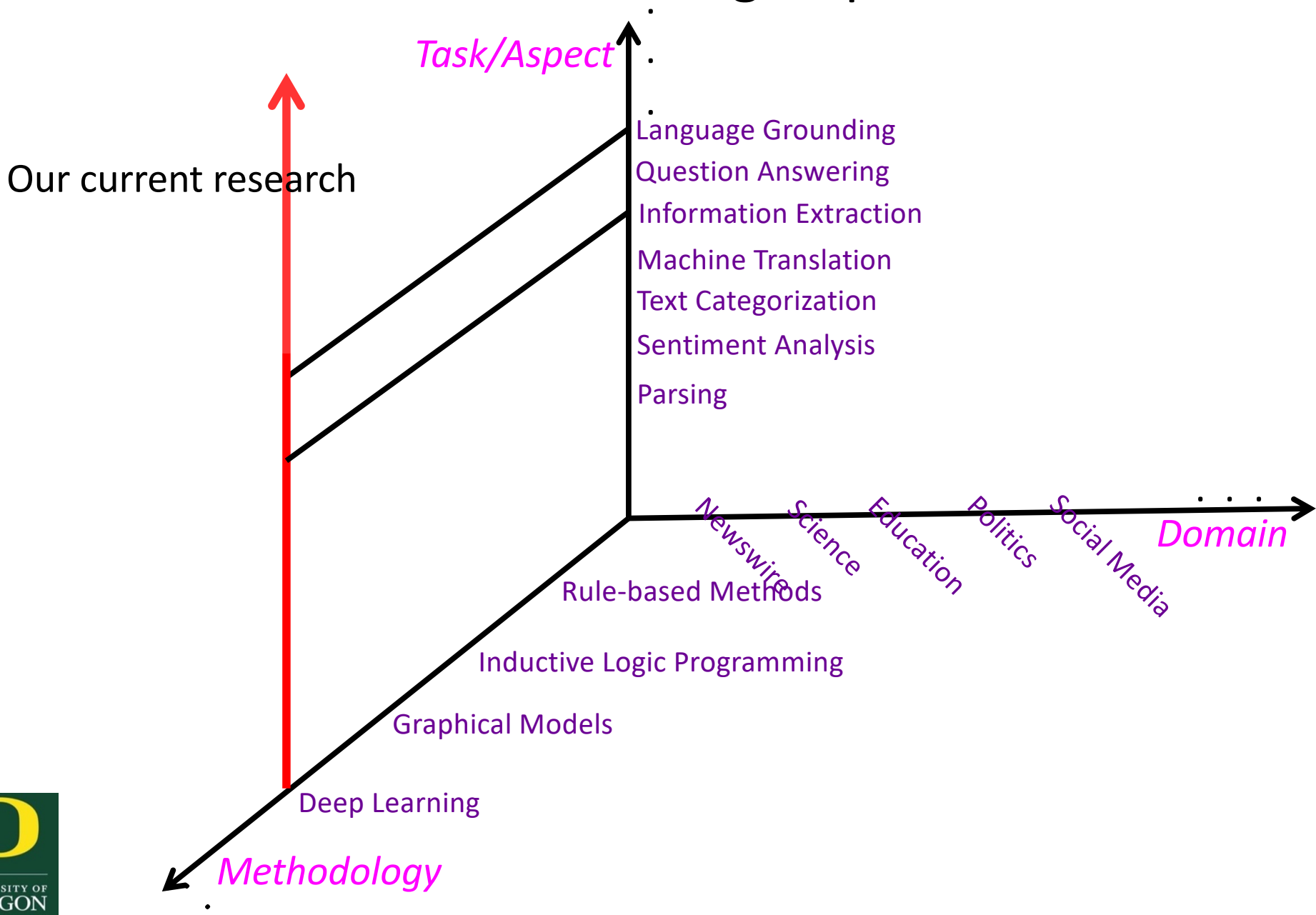


Clustering images



[Goldberger et al.]

Research in the NLP group at UO



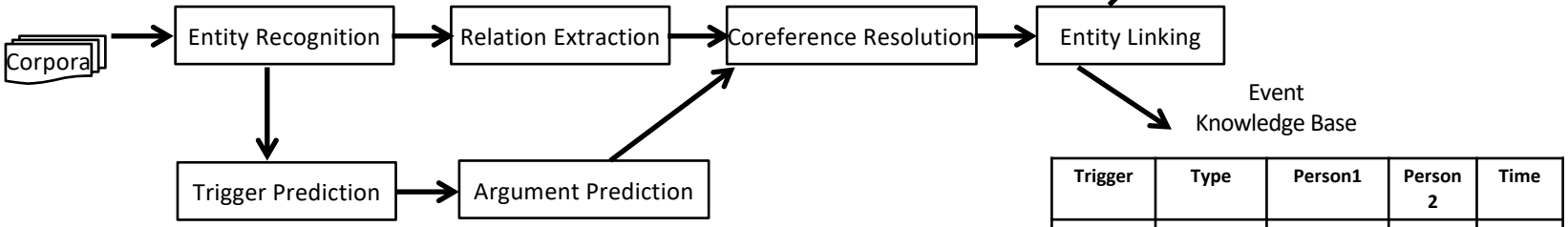
Information extraction

Giuliani, 58, proposed to Nathan, a former nurse, during a business trip to Paris _ five months after he finalized his divorce from Donna Hanover in July after 20 years of marriage.

In interviews last year, Giuliani said Nathan gave him ``tremendous emotional support'' through his treatment for prostate cancer and as he led New York City during the Sept. 11, 2001, terror attacks.

Relation Knowledge Base

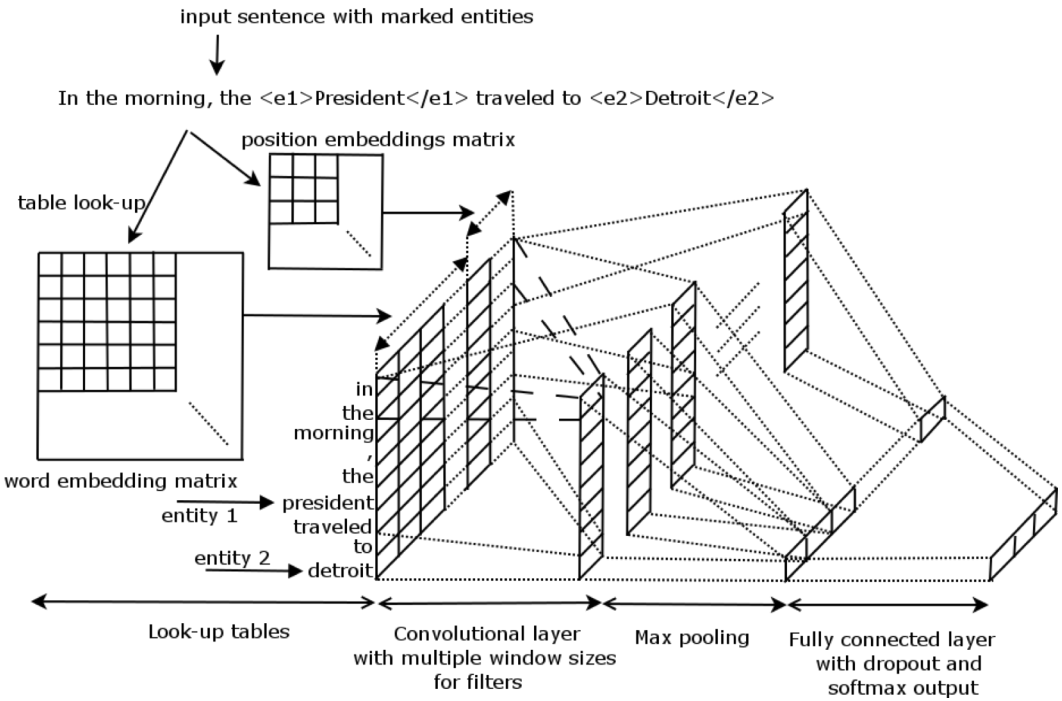
Name	leaderOf	....
Giuliani	New York City	
.....		



Event Knowledge Base

Trigger	Type	Person1	Person 2	Time
divorce	Divorce	Giuliani	Donna Hanover	July
.....				

Neural Networks for Relation Extraction



A Convolutional Neural Network (CNN) for Relation Extraction

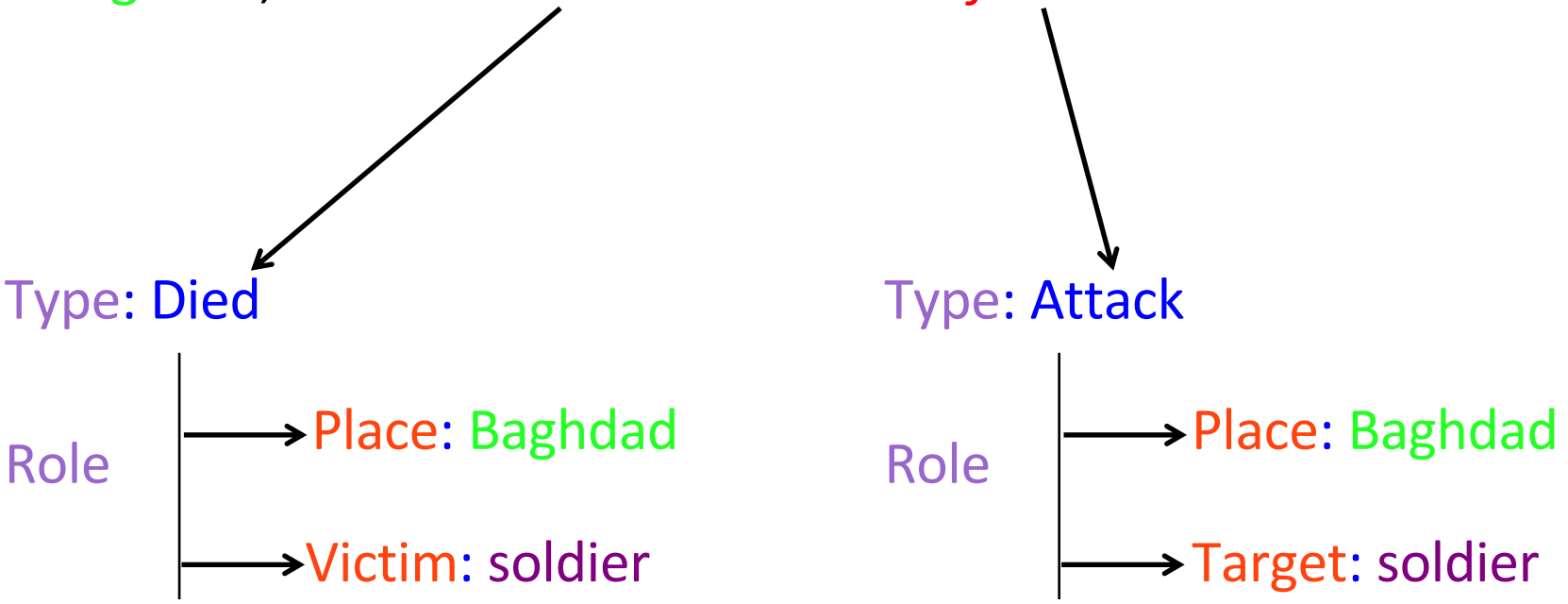
Nguyen and Grishman, 2015a

SemEval 2010 Dataset

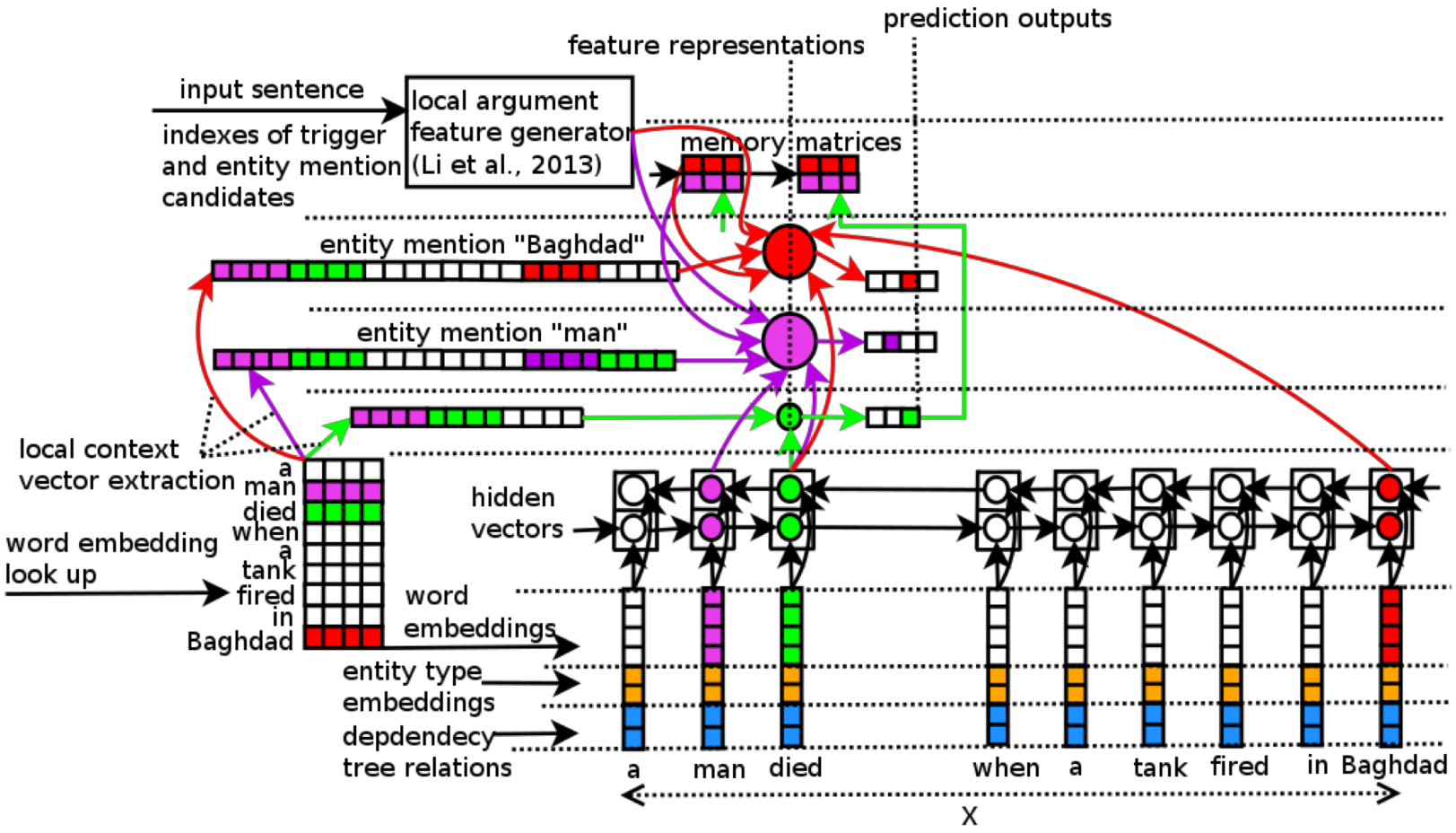
Classifier	Features	F
MaxEnt	POS, WordNet, morphological features, noun compound system, thesauri, Google n -grams	77.6
SVM	POS, WordNet, prefixes and other morphological fea- tures, dependency parse, Levin classes, PropBank, FrameNet, NomLex-Plus, Google n -grams, paraphrases, TextRunner	82.2
CNN (Zeng et al., 2014)	WordNet	82.7
CNN (Nguyen and Grishman, 2015a)	-	82.8

Event Extraction

In Baghdad, a soldier *died* when his tank *fired* on the street.

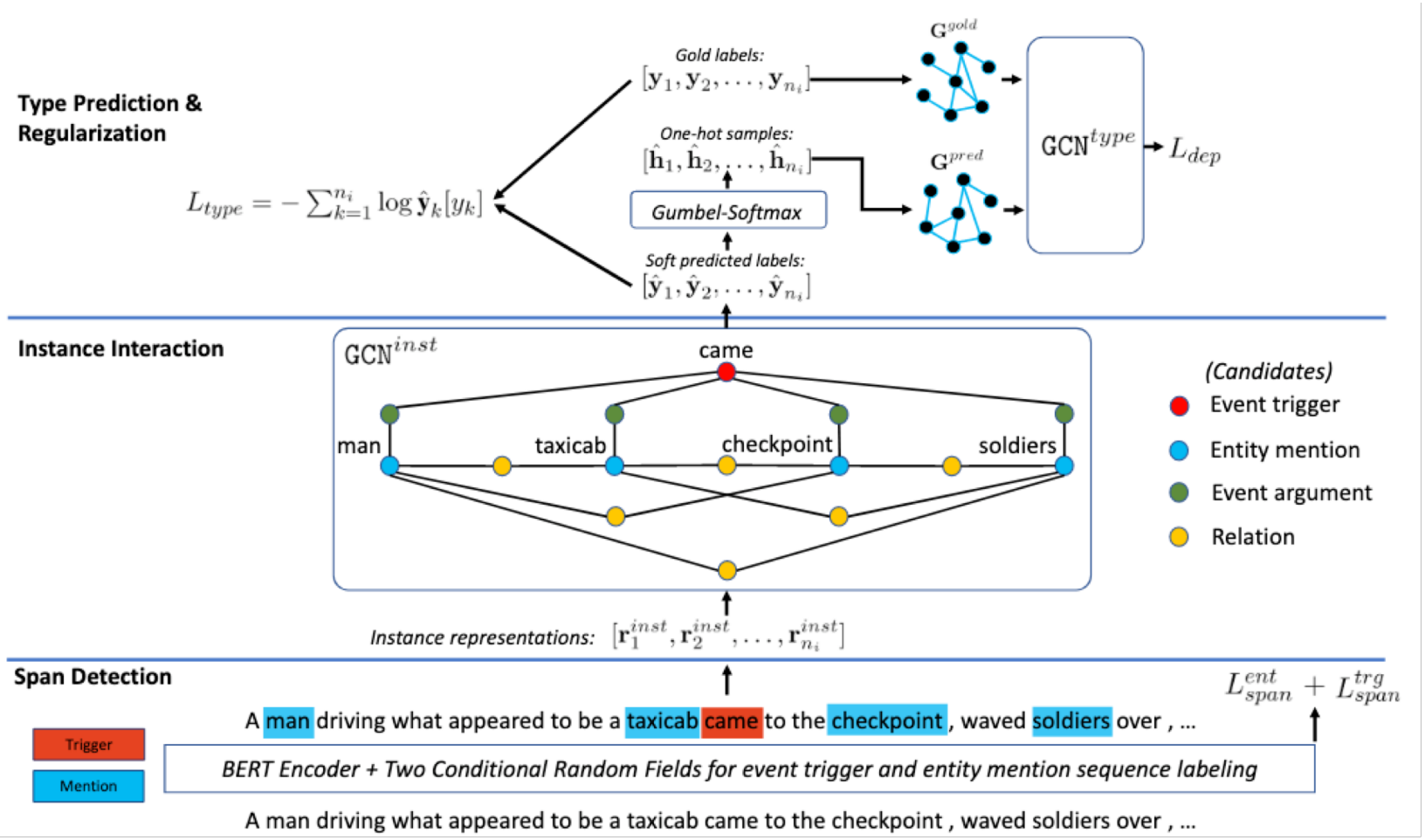


Joint Inference for Event Extraction



Nguyen et al., 2016 (NAACL 2016)

Joint Inference for Event Extraction



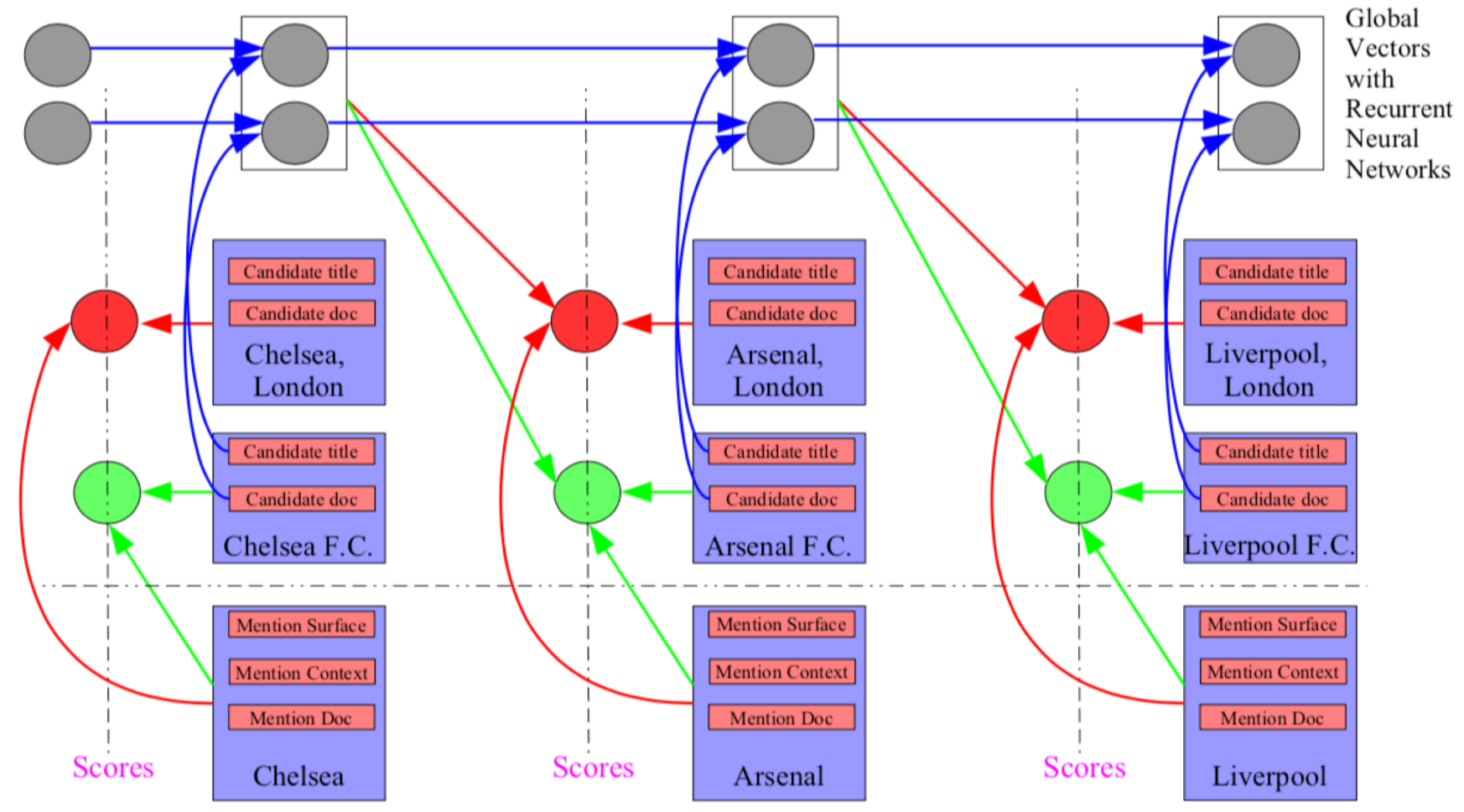
Nguyen et al., 2021 (NAACL 2021)

Demo: http://nlp.uoregon.edu/fourie

Entity Linking

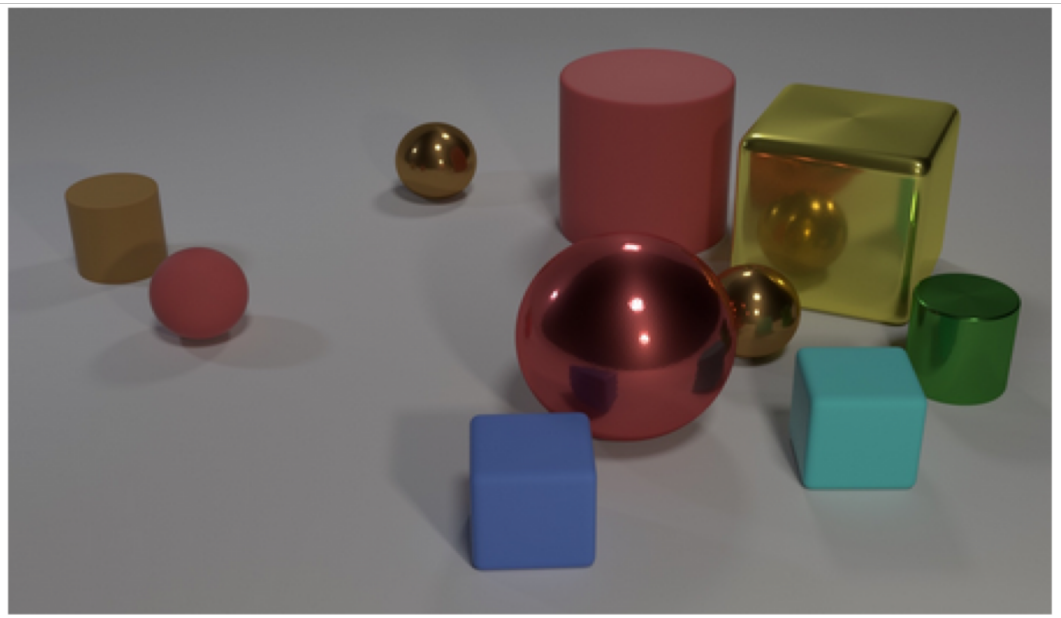
Chelsea have long-standing rivalries with *North London* clubs *Arsenal* and *Tottenham Hotspur*. A strong rivalry with *Leeds United* dates back to several heated and controversial matches in the 1960s and 1970s.

Joint Inference for Entity Linking



Nguyen et al., 2016 (COLING 2016)

Visual Question Answering (VQA)



Q: Are there an **equal number** of **large things** and **metal spheres**?

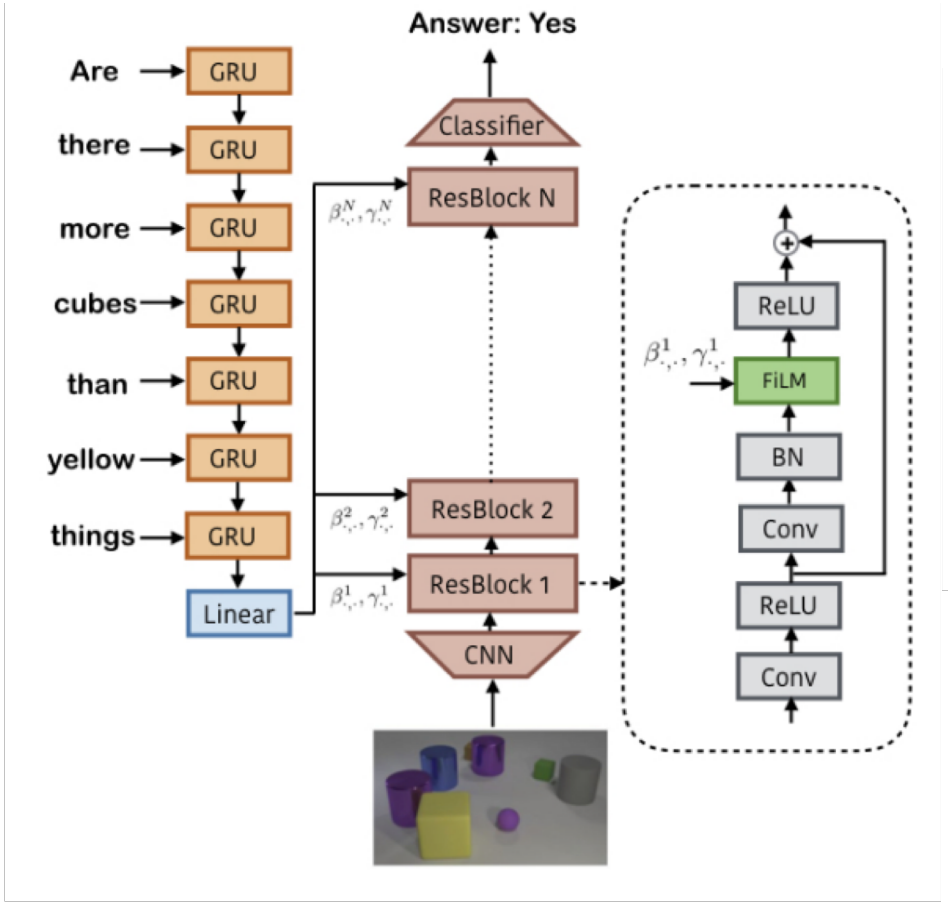
Q: **What size** is the **cylinder** **that is left of** the **brown metal** thing **that is left of** the **big sphere**?

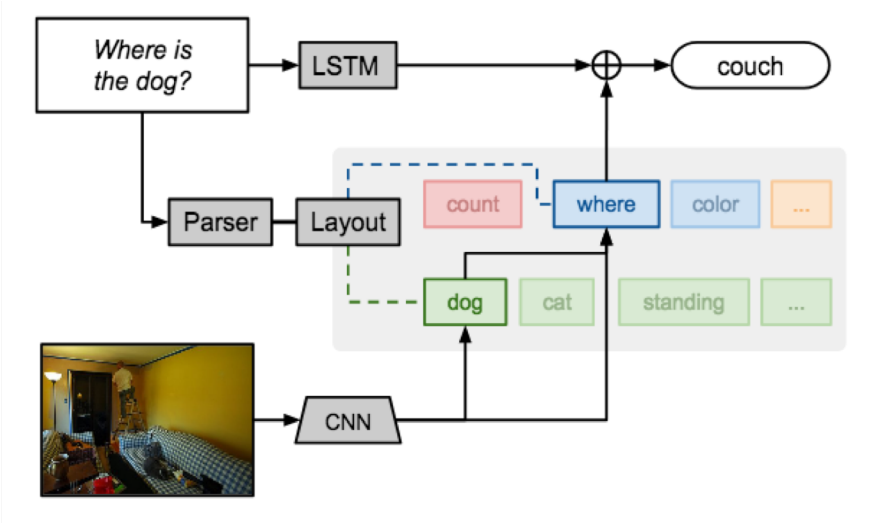
Q: There is a **sphere** with the **same size as** the **metal cube**; is it **made of the same material as** the **small red sphere**?

Q: **How many** objects are **either** **small cylinders** or **red** things?

Johnson et al., 2017

VQA Models





Neural Module Networks
Jacob et al., 2017

FiLM: Ethan et al., 2018

Machine learning has grown in leaps and bounds

- The main approach for
  - Speech Recognition
  - Robotics
  - Natural Language Processing
  - Computational Biology
  - Sensor networks
  - Computer Vision
  - Web
  - …and many more each year…

What We'll Cover

- **Supervised learning:** Decision tree induction, Instance-based learning, Neural networks/Deep Learning, Support vector machines, Linear Regression, Model ensembles, Learning theory, etc.
- **General machine learning concepts and techniques**: Feature selection, cross-validation, algorithm evaluation …
- **Unsupervised learning:** Clustering, Dimensionality reduction, Auto-Encoder, VAE, GAN (if time permitted)

Not covering:

- Reinforcement learning
- Probabilistic graphical models (410/510)
- Structured prediction (e.g., machine translation, image segmentation, multi-label classification)

What We'll Cover: Comparison

Computer Science

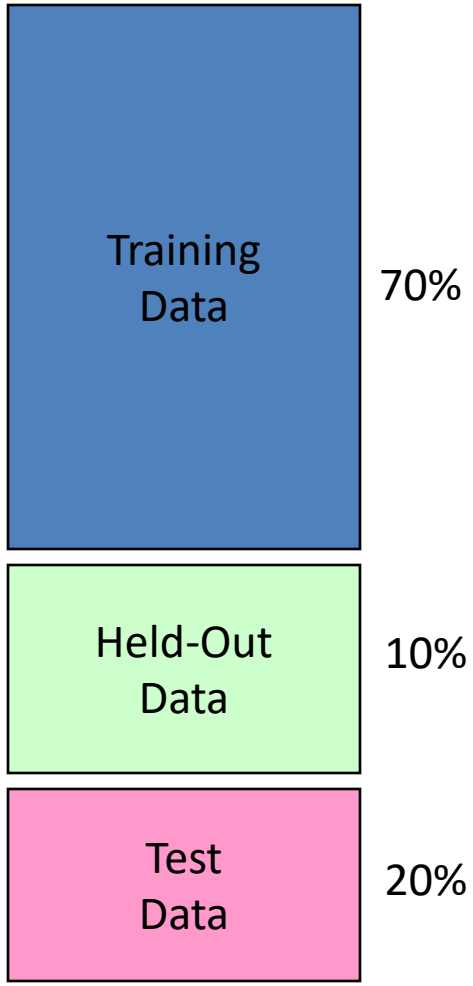
- **Core concepts:** Variables, conditionals, loops, functions, etc.
- **Key algorithms:** Mergesort, linked lists, binary search trees, breadth-first search, etc.
- **Process:** Debugging, software engineering, etc.

Machine Learning

- **Core concepts:** Classification, overfitting, underfitting, training set, etc.
- **Key algorithms:** Decision trees, nearest neighbor, linear models, etc.
- **Process:** Designing and debugging ML systems

Important Concepts

- Data: labeled instances, e.g. emails marked spam/ham
  - Training set
  - Held out set
  - Test set
- Features: attribute-value pairs which characterize each x
- Experimentation cycle
  - Learn parameters (e.g. model probabilities) on training set
  - (Tune hyperparameters on held-out set)
  - Compute accuracy of test set
  - Very important: never "peek" at the test set!
- Evaluation
  - Accuracy: fraction of instances predicted correctly
- Overfitting and generalization
  - Want a classifier which does well on *test* data
  - Overfitting: fitting the training data very closely, but not generalizing well
  - We'll investigate overfitting and generalization more formally in a few lectures



Goal: Use course attributes to predict the student's course rating (+2 = loved it, -2 = hated it).

How should we model this?

Rating	Easy?	AI?	Sys?	Thy?	Morning?
+2	y	y	n	y	n
+2	y	y	n	y	n
+2	n	y	n	n	n
+2	n	n	n	y	n
+2	n	y	y	n	y
+1	y	y	n	n	n
+1	y	y	n	y	n
+1	n	y	n	y	n
0	n	n	n	n	y
0	y	n	n	y	y
0	n	y	n	y	n
0	y	y	y	y	y
-1	y	y	y	n	y
-1	n	n	y	y	n
-1	n	n	y	n	y
-1	y	n	y	n	y
-2	n	n	y	y	n
-2	n	y	y	n	y
-2	y	n	y	n	n
-2	y	n	y	n	y

(Data is from CIML appendix.)

The Big Picture