Lecture 5

Introduction to Probabilistic models

Machine Learning

Ivan Smetannikov

04.07.2019

Lecture plan

- Overview and Motivation
- Distributions

Lecture plan

- Overview and Motivation
- Distributions

Problem

An illness, which is spread among 1% of population. This illness test returns true answers in 95% of cases. Someone receives a positive result.

What is the probability, he actually suffers the illness?

PGM: Motivation and Overview

PGM: Motivation and Overview

predisposing
symptoms
test results
diseases
treatment outcomes

PGM: Motivation and Overview

predisposing
symptoms
test results
diseases
treatment outcomes

millions of pixels or
thousands of superpixels

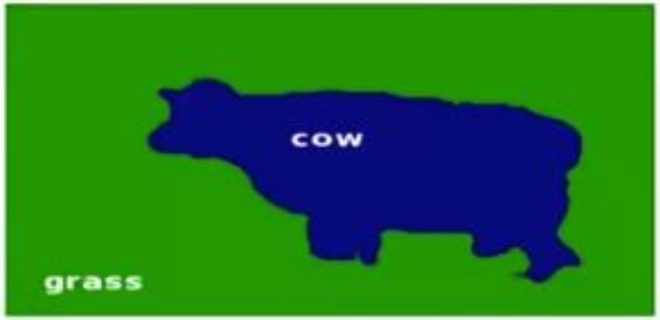
PGM: Motivation and Overview

predisposing
symptoms
test results
diseases
treatment outcomes

millions of pixels or
thousands of superpixels

Each, needs to be labeled
{grass, sky, water, cow, horse, …}

PGM: Motivation and Overview



predisposing
symptoms
test results
diseases
treatment outcomes

millions of pixels or
thousands of superpixels

Each, needs to be labeled
{grass, sky, water, cow, horse, …}

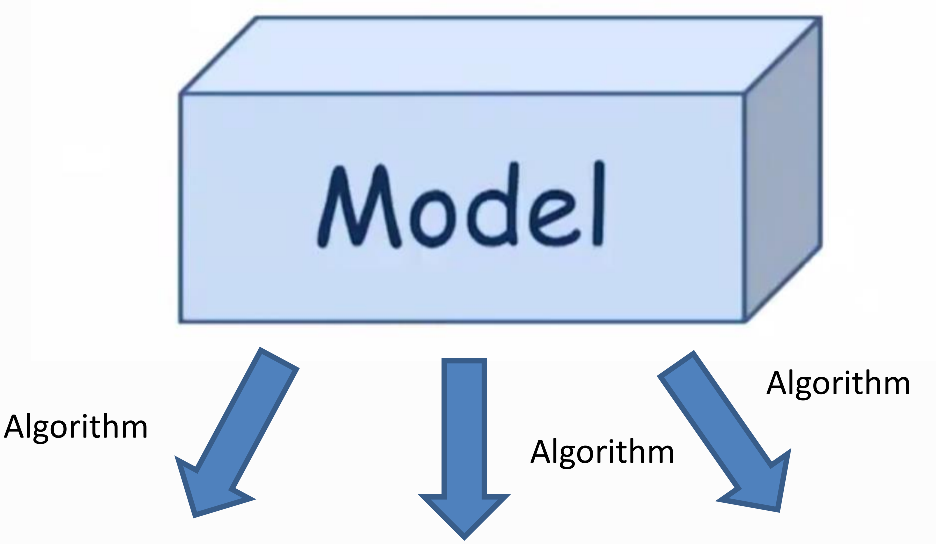
Probabilistic Graphical Models

Models

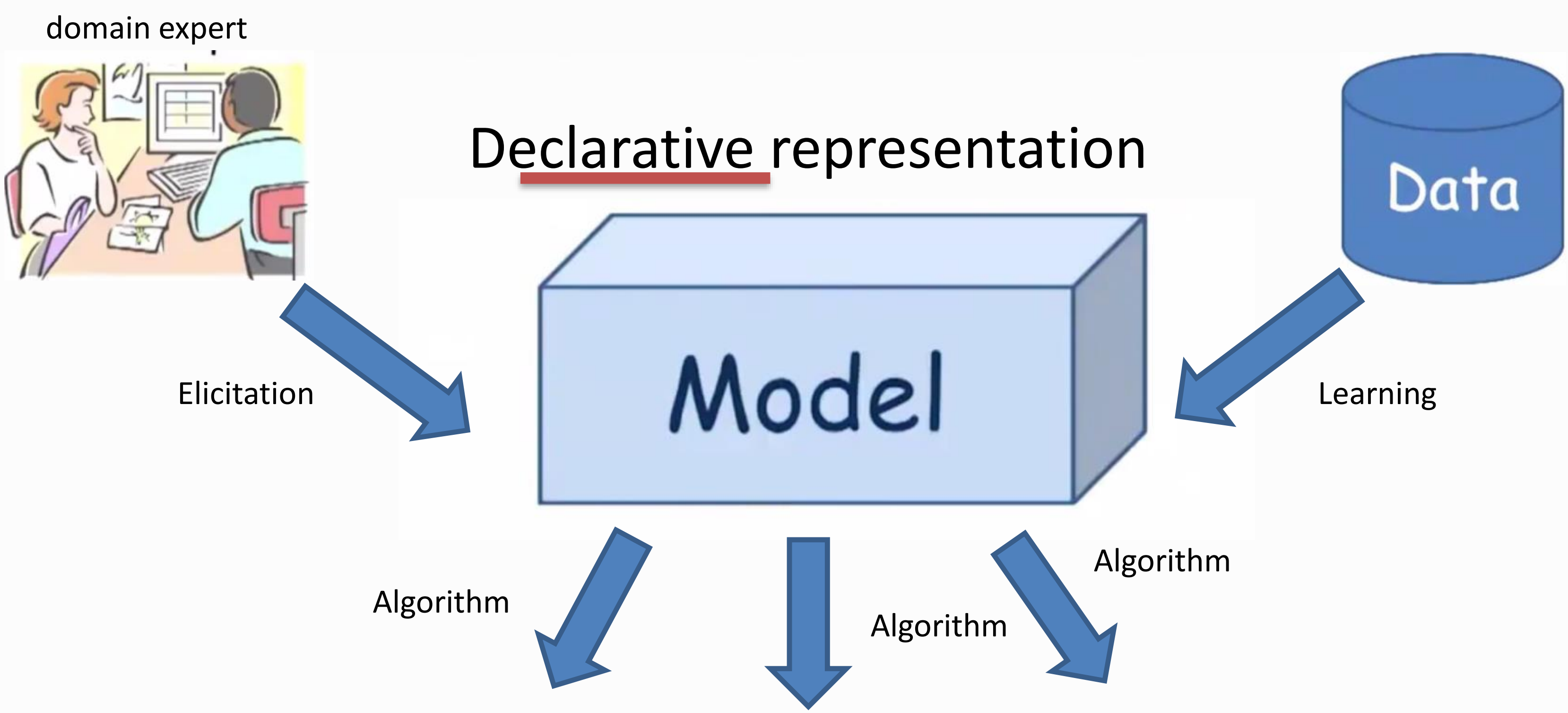
Declarative representation

Models

Declarative representation



Models


domain expert
Declarative representation
Data
Model
Elicitation
Learning
Algorithm
Algorithm
Algorithm

Uncertainty

- Partial knowledge of state of the world

Uncertainty

- Partial knowledge of state of the world
- Noisy observations

Uncertainty

- Partial knowledge of state of the world
- Noisy observations
- Phenomena not covered by our model

Uncertainty

- Partial knowledge of state of the world
- Noisy observations
- Phenomena not covered by our model
- Inherent stochasticity

Probability Theory

- Declarative representation with clear semantics

Probability Theory

- Declarative representation with clear semantics

Probability Theory

- Declarative representation with clear semantics
- Powerful reasoning patterns

Conditioning
Decision making

Probability Theory

- Declarative representation with clear semantics
- Powerful reasoning patterns

Conditioning
Decision making

- Established learning methods

Complex Systems

predisposing
symptoms
test results
diseases
treatment outcomes

class labels for
thousands of superpixels

Complex Systems

predisposing
symptoms
test results
diseases
treatment outcomes

class labels for
thousands of superpixels

Random variables $\boxed{X_1, \ldots, X}_n$

Complex Systems

predisposing
symptoms
test results
diseases
treatment outcomes

class labels for
thousands of superpixels

Random variables $\boxed{X_1, \ldots, X}_n$

Joint distribution $P(X_1, \ldots, X_n)$

Complex Systems

predisposing
symptoms
test results
diseases
treatment outcomes

class labels for
thousands of superpixels

Random variables $\boxed{X_1, \ldots, X}_n$

$\boxed{\text{Joint}}$ distribution $P(\underline{X_1, \ldots, X_n})$

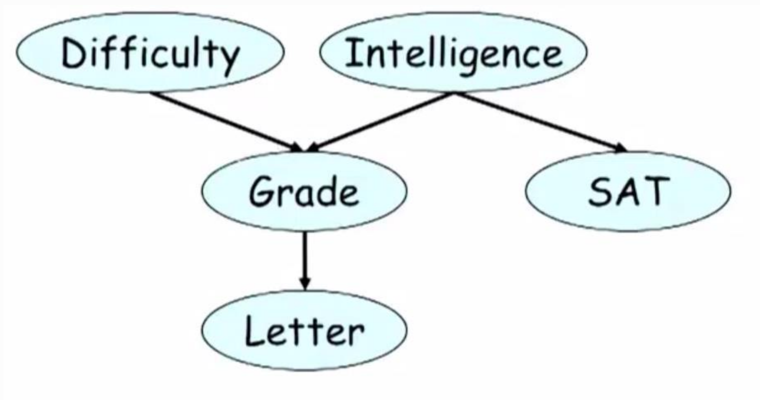
Binary valued distibrution over
2^n possible states

Graphical Models

Bayesian networks

$X_1, \ldots, X_n$ — nodes

Directed graph



Graphical Models

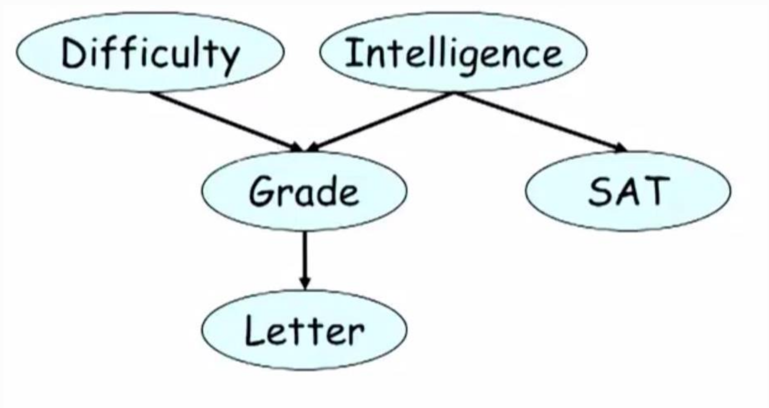
Bayesian networks

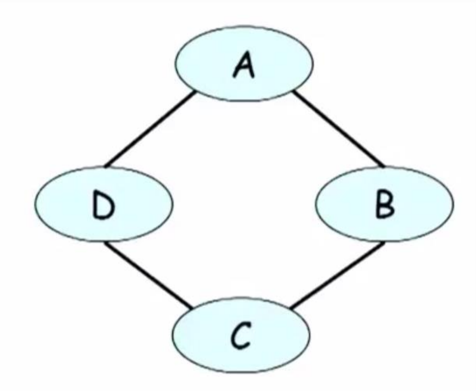
$X_1, \ldots, X_n$ — nodes

Directed graph

Markov networks

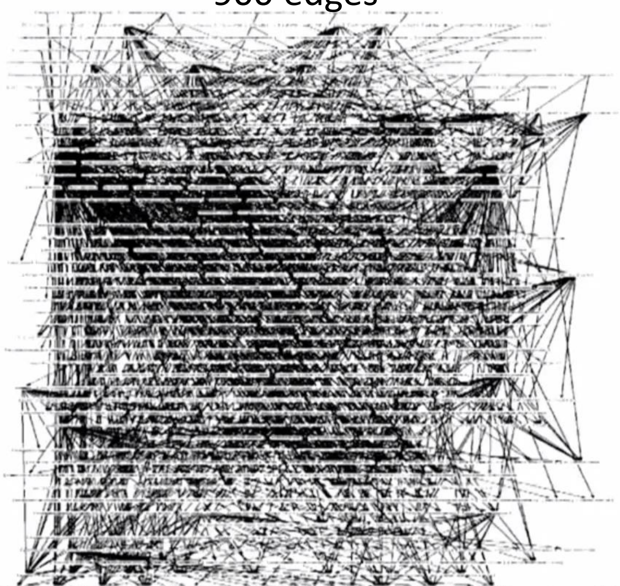
Undirected graph





Graphical Models

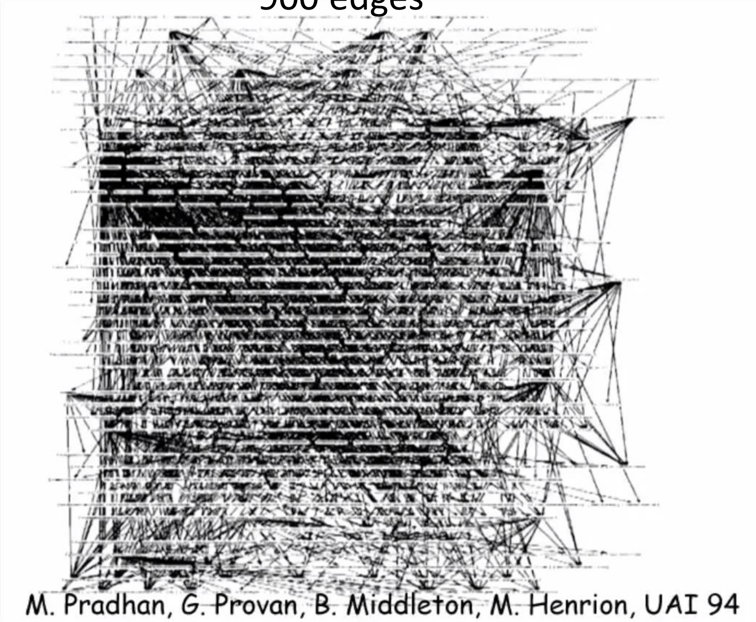
CPCS diagnosis
~480 nodes
~900 edges



M. Pradhan, G. Provan, B. Middleton, M. Henrion, UAI 94

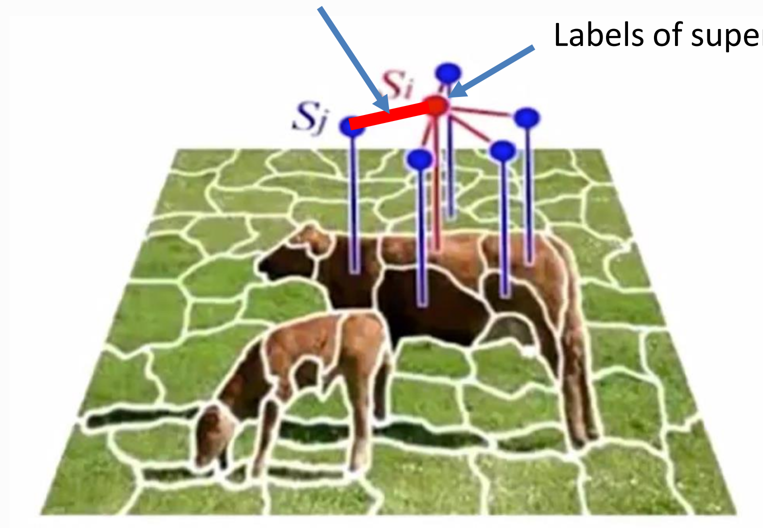
Graphical Models

CPCS diagnosis
~480 nodes
~900 edges



Probabilistic relationships

Labels of superpixels



Graphical Representation

- Intuitive and compact data structure
- Efficient reasoning using general–purpose algorithms
- Sparse parameterization
  - feasible elicitation
  - learning from data

Graphical Representation

- Intuitive and compact data structure
- Efficient reasoning using general–purpose algorithms
- Sparse parameterization
  - feasible elicitation ← by hand
  - learning from data ← automatically

Many Applications

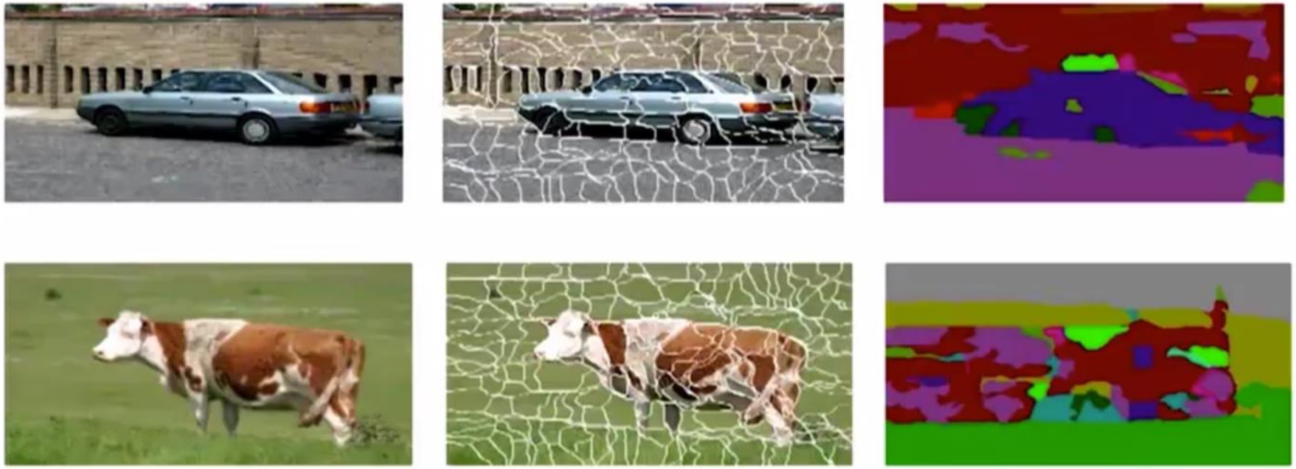
- Medial diagnosis
- Fault diagnosis
- Natural language processing
- Traffic analysis
- Social network models
- Message decoding
- Computer vision
  - Image segmentation
  - 3D reconstruction
  - Holistic scene analysis
- Speech recognition
- Robot localization and mapping

Image Segmentation

Image Segmentation

superpixels

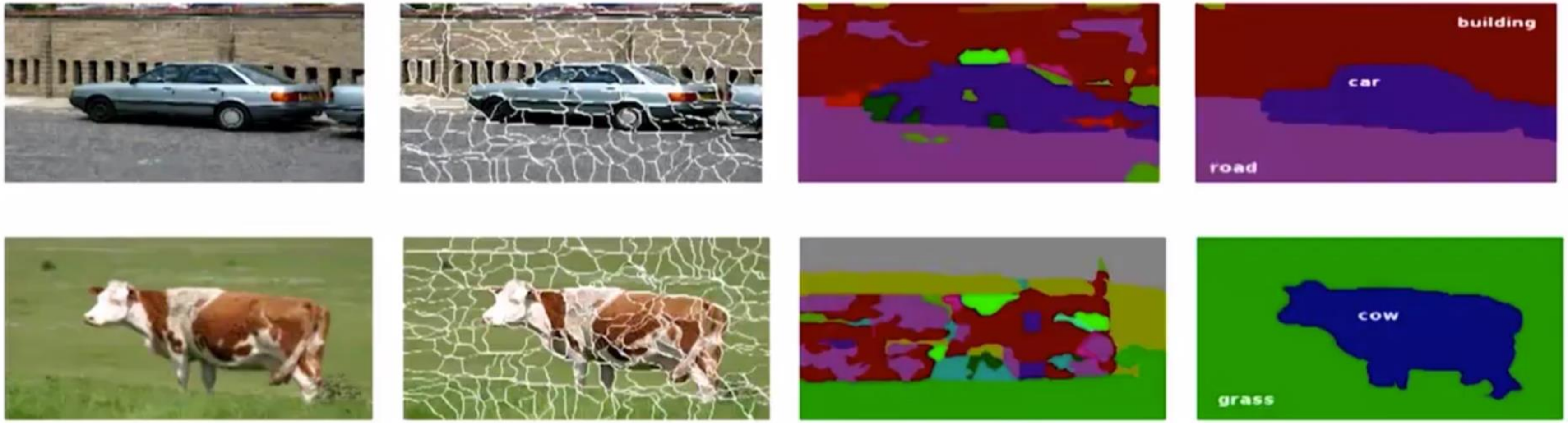
Image Segmentation



superpixels

machine learning
to separate superpixels

Image Segmentation



superpixels

machine learning
to separate superpixels

Textual Information Extracion

Mrs. Green spoke today in New York. Green chairs the finance committee.

person

Textual Information Extracion

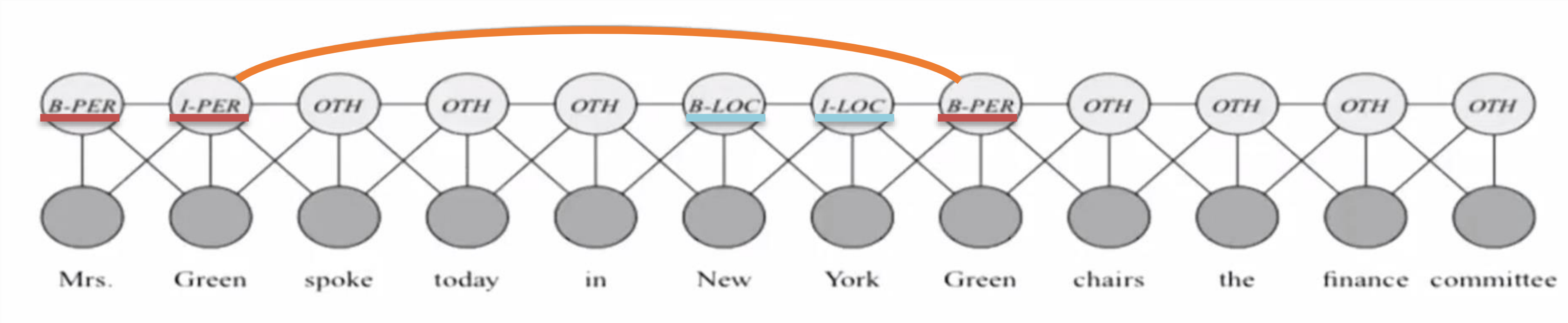
Mrs. Green spoke today in New York. Green chairs the finance committee.

person location person organization

Textual Information Extracion

Mrs. Green spoke today in New York. Green chairs the finance committee.

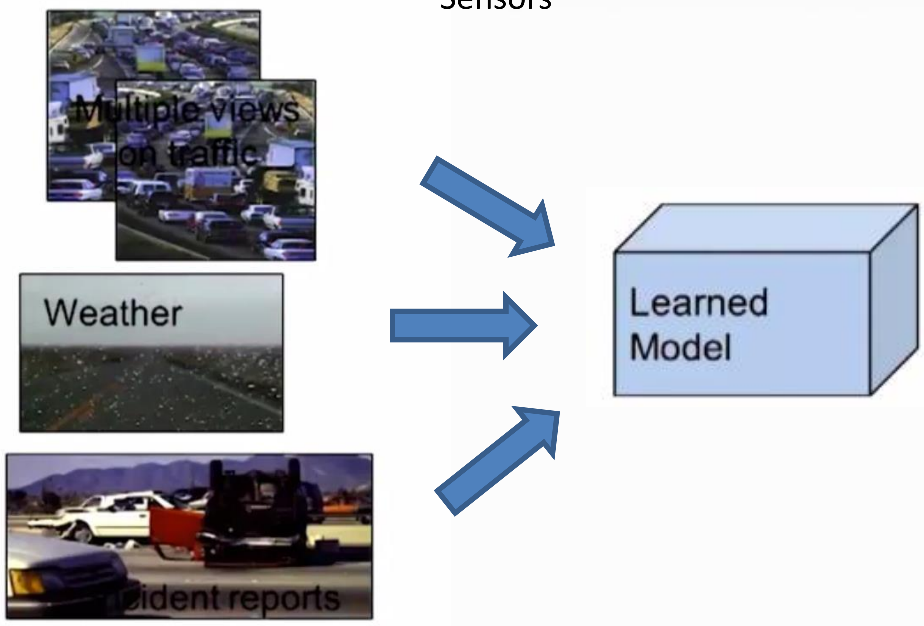
person location person organization



Multi-Sensor Integration

Multi-Sensor Integration

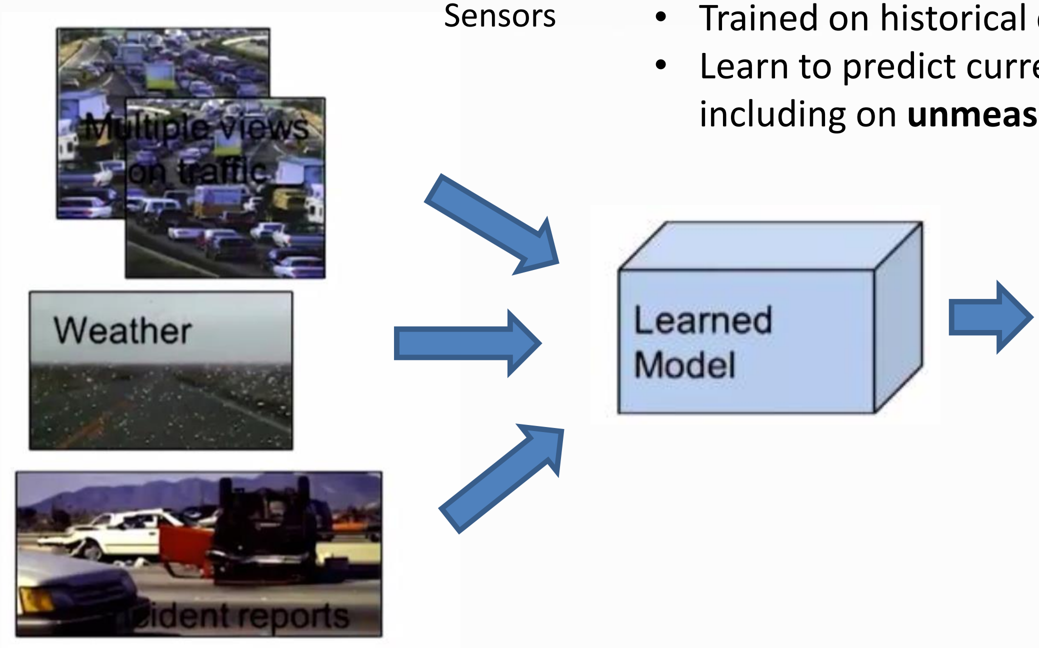
Sensors



Multi-Sensor Integration

Sensors

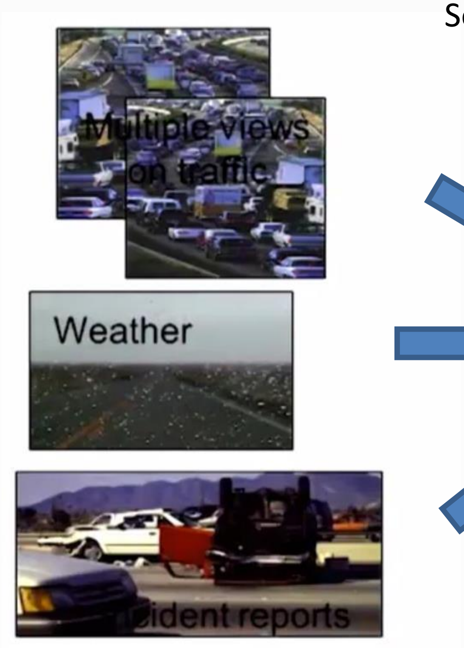
- Trained on historical data
- Learn to predict current and **future** road speed, including on **unmeasured** roads

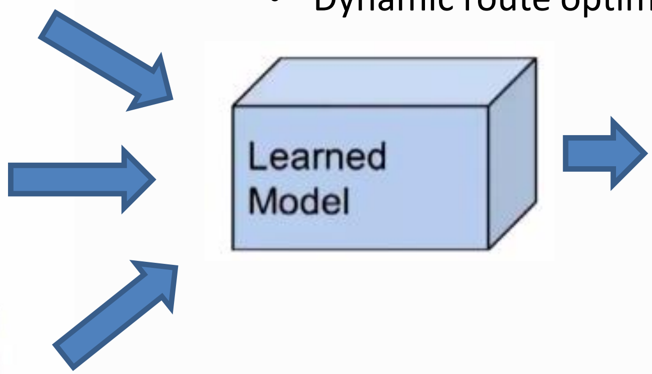


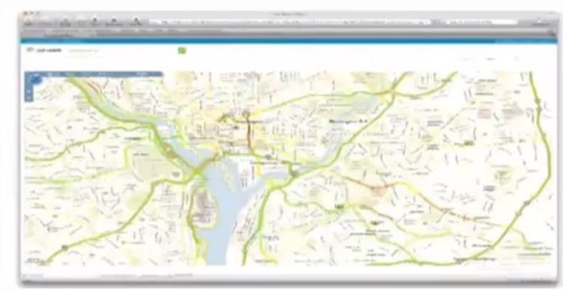
Multi-Sensor Integration

Sensors

- Trained on historical data
- Learn to predict current and **future** road speed, including on **unmeasured** roads
- Dynamic route optimization

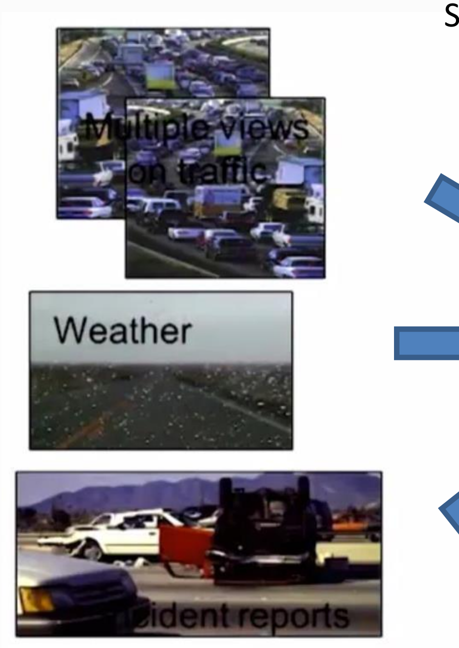




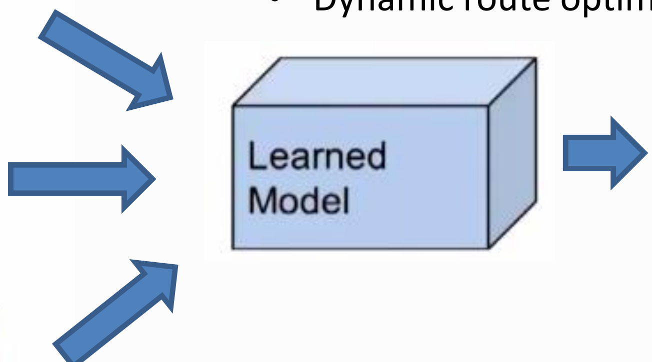


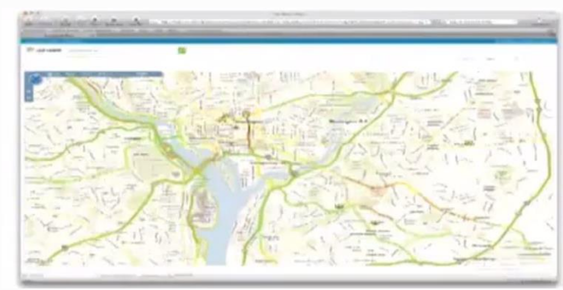
Multi-Sensor Integration

Sensors



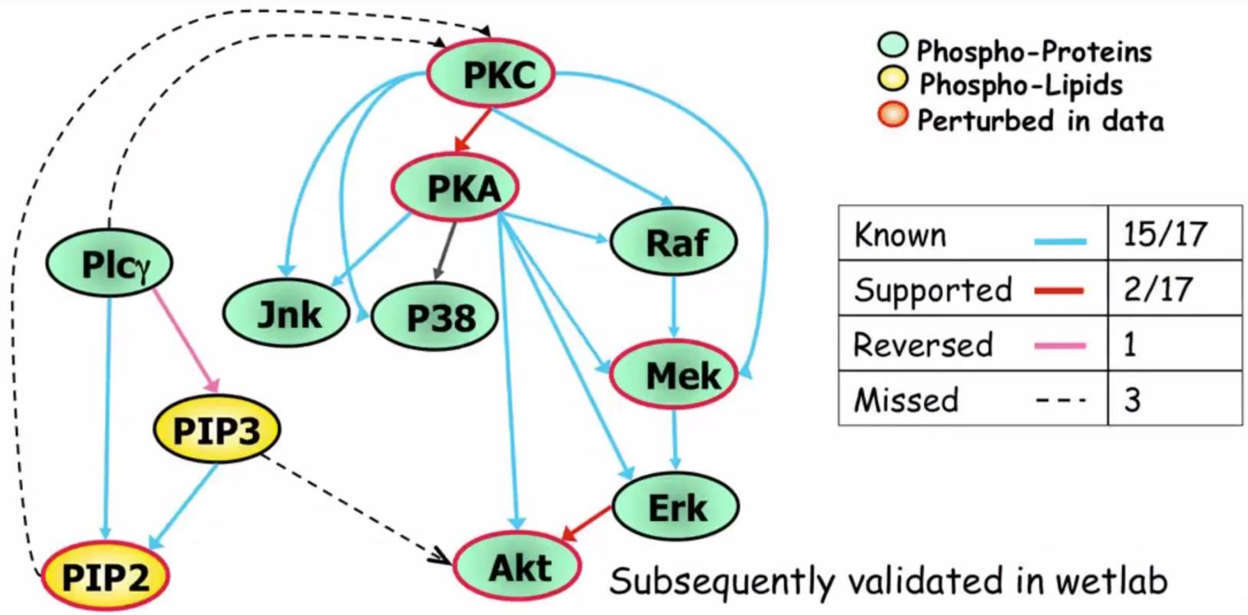
- Trained on historical data
- Learn to predict current and **future** road speed, including on **unmeasured** roads
- Dynamic route optimization





- I95 corridor experiment: accurate to ±5 MPH in 85% of cases
- Fielded in 72 cities

Multi-Sensor Integration



Causal protein-signaling networks derived from multiparameter single-cell data

Multi-Sensor Integration

- Representation
  - Directed and undirected
  - Temporal and plate models
- Inference
  - Exact and approximate
  - Decision making
- Learning
  - Parameters and structure
  - With and without complete data

Lecture Plan

- Overview and Motivation
- Distributions

Joint Distribution

- Intelligence (I)

  i^0 (low), i^1 (high)
- Difficulty (D)

  d^0 (easy), d^1 (hard)
- Grade (G)

  g^1 (A), g^2 (B), g^3 (C)

Joint Distribution

- Intelligence (I) $\leftarrow$ 2

  i^0 (low), i^1 (high)
- Difficulty (D) $\leftarrow$ 2

  d^0 (easy), d^1 (hard)
- Grade (G) $\leftarrow$ 3

  g^1 (A), g^2 (B), g^3 (C)

Joint Distribution

- Intelligence (I) ← **2**
  i^0 (low), i^1 (high)
- Difficulty (D) ← **2**
  d^0 (easy), d^1 (hard)
- Grade (G) ← **3**
  g^1 (A), g^2 (B), g^3 (C)

Parameters:
2x2x3=12

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Joint Distribution

- Intelligence (I) ← 2
  i^0 (low), i^1 (high)
- Difficulty (D) ← 2
  d^0 (easy), d^1 (hard)
- Grade (G) ← 3
  g^1 (A), g^2 (B), g^3 (C)

Parameters: 2x2x3=12

Independent parameters: 11

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

sum=1

Conditioning

condition on g^1

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Conditioning

condition on g^1

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
~~i^0~~	~~d^0~~	~~g^2~~	~~0.168~~
~~i^0~~	~~d^0~~	~~g^3~~	~~0.126~~
i^0	d^1	g^1	0.009
~~i^0~~	~~d^1~~	~~g^2~~	~~0.045~~
~~i^0~~	~~d^1~~	~~g^3~~	~~0.126~~
i^1	d^0	g^1	0.252
~~i^1~~	~~d^0~~	~~g^2~~	~~0.0224~~
~~i^1~~	~~d^0~~	~~g^3~~	~~0.0056~~
i^1	d^1	g^1	0.06
~~i^1~~	~~d^1~~	~~g^2~~	~~0.036~~
~~i^1~~	~~d^1~~	~~g^3~~	~~0.024~~

Conditioning: Reduction

condition on g^1

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

Conditioning: Renormalization

I	D	G	P(I,D,G)
i^0	d^0	g^1	0.126
i^0	d^1	g^1	0.009
i^1	d^0	g^1	0.252
i^1	d^1	g^1	0.06

sum=0.447

P(I,D, g^1)

unnormalized measure

Conditioning: Renormalization

I	D	G	P(I,D, g^1)	
i^0	d^0	g^1	0.126	/0.447
i^0	d^1	g^1	0.009	/0.447
i^1	d^0	g^1	0.252	/0.447
i^1	d^1	g^1	0.06	/0.447

sum=0.447

I	D	P(I,D|g^1)
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

P(I,D, g^1)

unnormalized measure

Conditioning: Marginalization

Marginalize I

I	D	P(I,D)
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

Conditioning: Marginalization

Marginalize I

I	D	P(I,D)
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

D	P(D)
d^0	0.846
d^1	0.154