Predicting structures: Practical concerns

CS 6355: Structured Prediction

So far…

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

This lecture:
Using the tools: Practical concerns

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

This lecture:
Using the tools: Practical concerns

- We want to solve a task.
- Many choices ahead!

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

This lecture:
Using the tools: Practical concerns

- We want to solve a task.
- Many choices ahead!

What is the graph?

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

This lecture:
Using the tools: Practical concerns

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

- We want to solve a task.
- Many choices ahead!

What is the graph?

- Modeling our problem?
- Identifying variables?
- Identifying groups that are scored together? (factors)
- What are features?

This lecture:
Using the tools: Practical concerns

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

- We want to solve a task.
- Many choices ahead!

What is the graph?

- Modeling our problem?
- Identifying variables?
- Identifying groups that are scored together? (factors)
- What are features?

The best way to learn?

This lecture:
Using the tools: Practical concerns

- What are structures?
  - A graph
  - A collection of parts that are scored jointly
  - A collection of interconnected decisions
- Conditional models
  - We want to convert some input to an output
  - Model the conditional distribution of the output
  - Score groups of inter-connected variables
- Algorithms for learning
  - Local vs. global learning
  - Different algorithms
- Inference algorithms
  - Predicting the final output
  - Different algorithms, tradeoffs

- We want to solve a task.
- Many choices ahead!

What is the graph?

- Modeling our problem?
- Identifying variables?
- Identifying groups that are scored together? (factors)
- What are features?

The best way to learn?

What inference algorithm?

Modeling your problem

- Understand the problem: What should your program produce?
  - Is there data? Very often, the answer is no. ☹
- What are the decisions/random variables that constitute the output?
- How do they interact? Identifying factors/parts
  - Some interactions are natural, some are spurious (specific to your small collection of data)
  - Some interactions make inference impossible for computational reasons
  - What are the feature representations?
- Learning
  - What are the scoring functions?
  - Should every scoring function be jointly learned?
  - Perhaps, learn sub-sections independently and put them together with inference at the end
  - Which learning algorithm?
- Inference
  - What algorithm? How expensive is it?
  - Exact or approximate?

Example 0: Named Entity Recognition

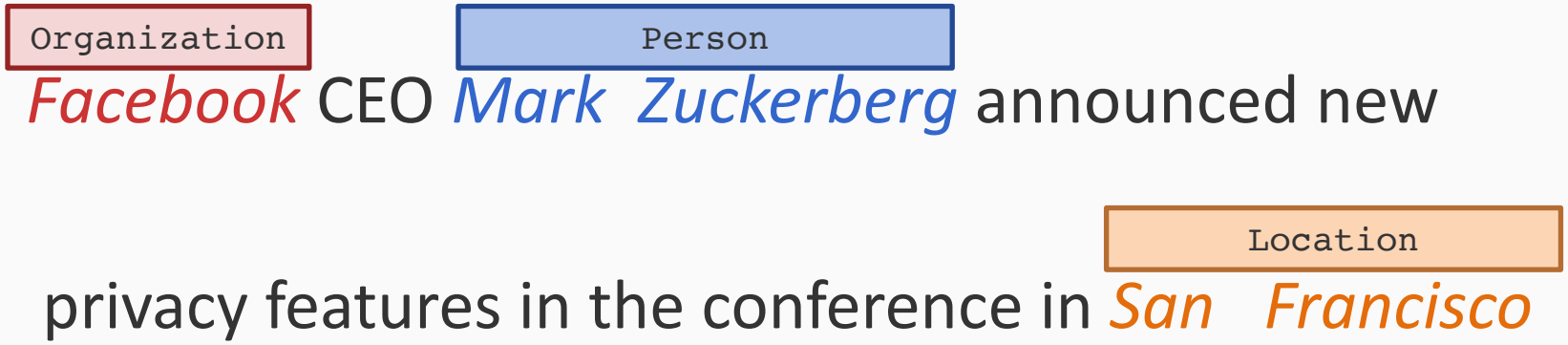
Goal: To identify persons, locations and organizations in text

Facebook CEO Mark Zuckerberg announced new

privacy features in the conference in San Francisco

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text



Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

Organization Person

Facebook CEO *Mark Zuckerberg* announced new

Location

privacy features in the conference in *San Francisco*

Design choices:

1. What are the set of decisions the predictor needs to make?
2. How do these decisions interact? Factors?
3. Features? Factor potentials/scoring functions?
4. Learning? Inference?

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

What are the set of decisions the predictor needs to make?

One option: Label spans of text

	PER	LOC	ORG	NONE
Facebook	✗	✗	✓	✗
Facebook CEO	✗	✗	✗	✓
Facebook CEO Mark	✗	✗	✗	✓
Facebook CEO Mark Zuckerberg	✗	✗	✗	✓
…				
Mark Zuckerberg	✓	✗	✗	✗
….				

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

How do the decisions interact?

A single word can have only one label

	PER	LOC	ORG	NONE
Facebook	?	?	?	?
Facebook CEO	?	?	?	?
Facebook CEO Mark	?	?	?	?
Facebook CEO Mark Zuckerberg	?	?	?	?
…				
Mark Zuckerberg	?	?	?	?
….				

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

How do the decisions interact?

A single word can have only one label

Disallowed together

	PER	LOC	ORG	NONE
Facebook	✓	?	?	?
Facebook CEO	✓	?	?	?
Facebook CEO Mark	?	?	?	?
Facebook CEO Mark Zuckerberg	?	?	?	?
…				
Mark Zuckerberg	?	?	?	?
….				

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

Features? Factor potentials/scoring functions?

Score(span, label)

- Could be linear in features
- Could be a neural network

	PER	LOC	ORG	NONE
Facebook	✓	?	?	?
Facebook CEO	✓	?	?	?
Facebook CEO Mark	?	?	?	?
Facebook CEO Mark Zuckerberg	?	?	?	?
…				
Mark Zuckerberg	?	?	?	?
….				

Disallowed together

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

Learning and inference

Various learning regimes

Various inference algorithms

	PER	LOC	ORG	NONE
Facebook	✓	?	?	?
Facebook CEO	✓	?	?	?
Facebook CEO Mark	?	?	?	?
Facebook CEO Mark Zuckerberg	?	?	?	?
…				
Mark Zuckerberg	?	?	?	?
….				

Disallowed together

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

A different modeling choice: One label per word

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

A different modeling choice: One label per word

Facebook CEO Mark Zuckerberg announced new

privacy features in the conference in San Francisco

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

A different modeling choice: One label per word

Facebook CEO Mark Zuckerberg announced new

privacy features in the conference in San Francisco

B-org = Start of organization
B-per = Start of person
I-per = In person

B-loc = Start of location
I-loc = In location
O = Not a named entity

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

A different modeling choice: One label per word

```
B-org      O    B-per I-per        O          O
Facebook CEO Mark  Zuckerberg announced new
O          O       O O    O           O B-loc I-loc
privacy features in the conference in San   Francisco
```

B-org = Start of organization
B-per = Start of person
I-per = In person

B-loc = Start of location
I-loc = In location
O = Not a named entity

Example 0: Named Entity Recognition

Goal: To identify persons, locations and organizations in text

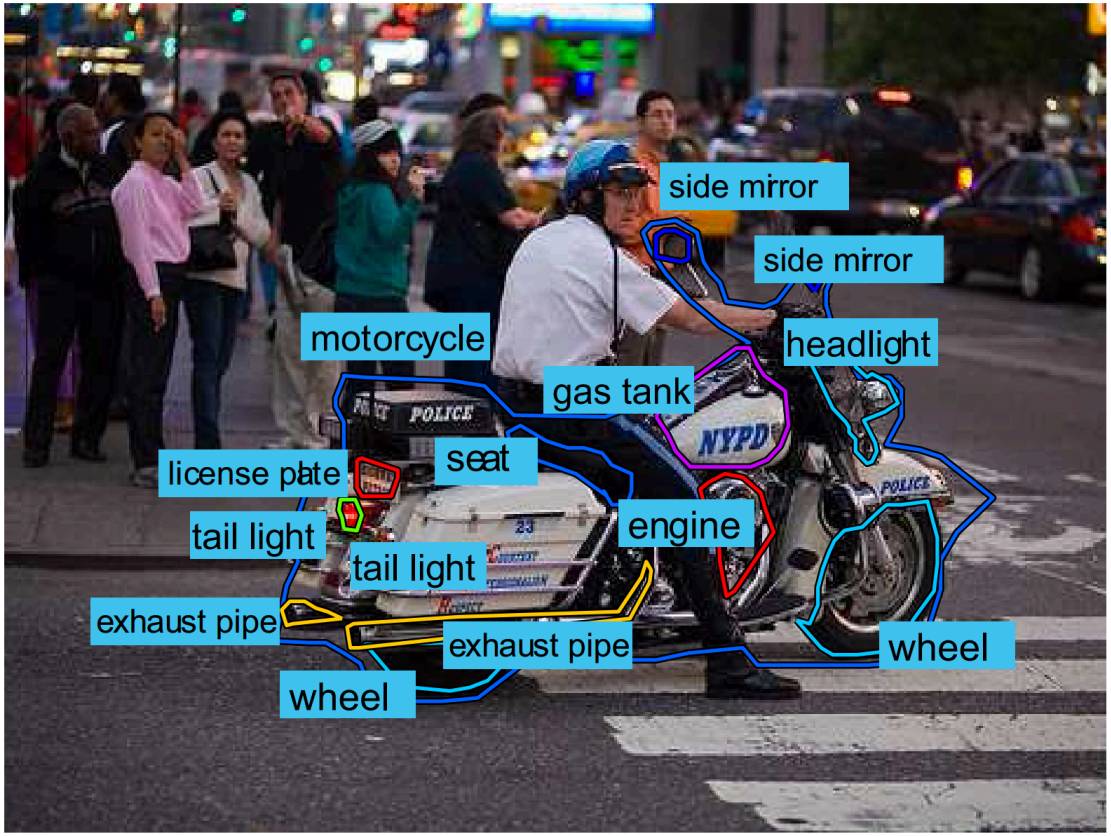
A different modeling choice: One label per word

```
B-org     O   B-per I-per       O          O
Facebook CEO Mark  Zuckerberg announced new
O         O      O O   O           O B-loc I-loc
privacy features in the conference in San   Francisco
```

This modeling choice offers its own design choices

1. How do these decisions interact? Factors?
2. Features?
3. Learning? Inference?

Example 1: Detecting objects and parts



Let's discuss the choices we have:

1. What are the set of decisions the predictor needs to make?
2. How do these decisions interact? Factors?
3. Features?
4. Learning? Inference?

[Farhadi, et al]

Example 2: Information extraction

Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.

Example 2: Information extraction

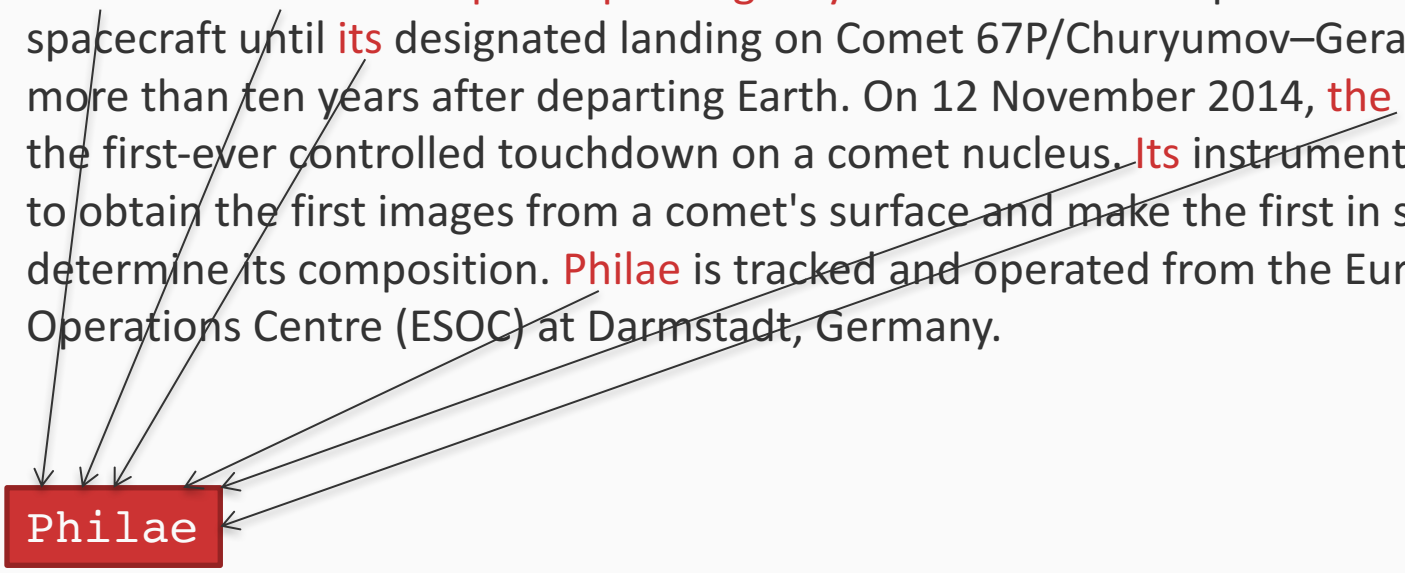
Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.

Touchdown	
`Lander`	Philae
`Destination`	Comet 67P
`When?`	12 November 2014

How do we model this problem?

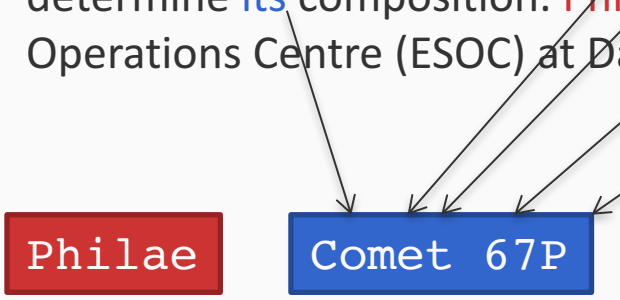
Example 2: Information extraction

Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.



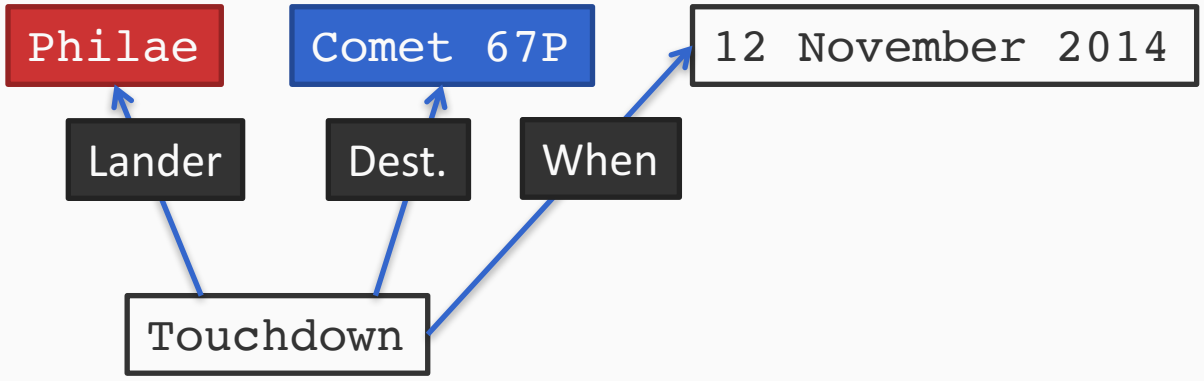
Example 2: Information extraction

Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.



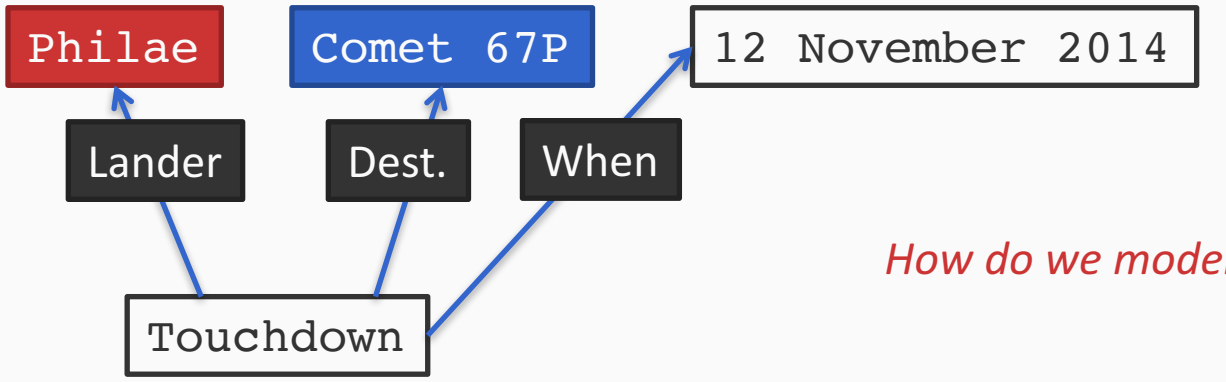
Example 2: Information extraction

Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.



Example 2: Information extraction

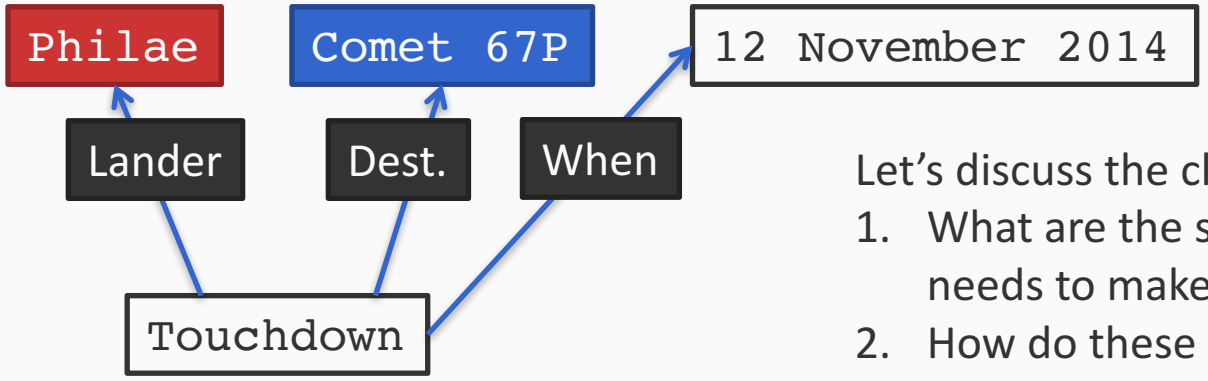
Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.



How do we model this problem?

Example 2: Information extraction

Philae is a robotic European Space Agency lander that accompanied the Rosetta spacecraft until its designated landing on Comet 67P/Churyumov–Gerasimenko (67P), more than ten years after departing Earth. On 12 November 2014, the lander achieved the first-ever controlled touchdown on a comet nucleus. Its instruments are expected to obtain the first images from a comet's surface and make the first in situ analysis to determine its composition. Philae is tracked and operated from the European Space Operations Centre (ESOC) at Darmstadt, Germany.



Let's discuss the choices we have:

1. What are the set of decisions the predictor needs to make?
2. How do these decisions interact? Factors?
3. Features?
4. Learning? Inference?

Modeling your problem

- Understand the problem: What should your program produce?
  - Is there data? Very often, the answer is no. ☹
- What are the decisions/random variables that constitute the output?
- How do they interact? Identifying factors/parts
  - Some interactions are natural, some are spurious (specific to your small collection of data)
  - Some interactions make inference impossible for computational reasons
  - What are the feature representations?
- Learning
  - What are the scoring functions?
  - Should every scoring function be jointly learned?
  - Perhaps, learn sub-sections independently and put them together with inference at the end
  - Which learning algorithm?
- Inference
  - What algorithm? How expensive is it?
  - Exact or approximate?