TP1: Leveraging Complex Prior Knowledge in Learning

Neil D. Lawrence
Marc Dymetman

28th January 2008
Outline

1. Background to Thematic Programme
2. Algorithmic Issues
3. Case Studies
4. Programme of Workshops
Outline

1. Background to Thematic Programme
2. Algorithmic Issues
3. Case Studies
4. Programme of Workshops
Last year: Thematic Programme in Comp Bio

- Three workshops directly associated.
- Perhaps 100 participants.

Aims for PASCAL II:

- Five to ten workshops directly associated.
- “Most PASCAL members to participate”

This programme: March 2008 – September 2008

- Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio
  ▶ Three workshops directly associated.
  ▶ Perhaps 100 participants.

Aims for PASCAL II:
  ▶ Five to ten workshops directly associated.
  ▶ “Most PASCAL members to participate”

This programme: March 2008 – September 2008
  ▶ Extension September 2008 – March 2009
PASCAL II Thematic Programmes

- Last year: Thematic Programme in Comp Bio
  - Three workshops directly associated.
  - Perhaps 100 participants.

- Aims for PASCAL II:
  - Five to ten workshops directly associated.
  - “Most PASCAL members to participate”

- This programme: March 2008 – September 2008
  - Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio
  ▶ Three workshops directly associated.
  ▶ Perhaps 100 participants.

Aims for PASCAL II:
  ▶ Five to ten workshops directly associated.
  ▶ “Most PASCAL members to participate”

This programme: March 2008 – September 2008
  ▶ Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio
  ▶ Three workshops directly associated.
  ▶ Perhaps 100 participants.

Aims for PASCAL II:
  ▶ Five to ten workshops directly associated.
  ▶ “Most PASCAL members to participate”

This programme: March 2008 – September 2008
  ▶ Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio
  ▶ Three workshops directly associated.
  ▶ Perhaps 100 participants.

Aims for PASCAL II:
  ▶ Five to ten workshops directly associated.
  ▶ “Most PASCAL members to participate”

This programme: March 2008 – September 2008
  ▶ Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio
  
  ▶ Three workshops directly associated.
  ▶ Perhaps 100 participants.

Aims for PASCAL II:

  ▶ Five to ten workshops directly associated.
  ▶ “Most PASCAL members to participate”

This programme: March 2008 – September 2008

  ▶ Extension September 2008 – March 2009
Last year: Thematic Programme in Comp Bio

- Three workshops directly associated.
- Perhaps 100 participants.

Aims for PASCAL II:

- Five to ten workshops directly associated.
- “Most PASCAL members to participate”

This programme: March 2008 – September 2008

- Extension September 2008 – March 2009
Philosophy: Machine learning, Statistics and AI

- Statistics assumes statistician
- AI and Machine Learning replace the human.

Practice:

- Humans always in the loop.
- AI focussed on expert systems
- Machine Learning focussed Data Driven Systems
Philosophy: Machine learning, Statistics and AI

- Statistics assumes statistician
- AI and Machine Learning replace the human.

Practice:

- Humans always in the loop.
- AI focussed on expert systems
- Machine Learning focussed Data Driven Systems
Machine Learning History

- **Philosophy:** Machine learning, Statistics and AI
  - Statistics assumes statistician
  - AI and Machine Learning replace the human.

- **Practice:**
  - Humans always in the loop.
  - AI focused on expert systems
  - Machine Learning focused Data Driven Systems
Machine Learning History

- **Philosophy**: Machine learning, Statistics and AI
  - Statistics assumes statistician
  - AI and Machine Learning replace the human.

- **Practice**:
  - Humans always in the loop.
  - AI focussed on expert systems
  - Machine Learning focussed Data Driven Systems
Philosophy: Machine learning, Statistics and AI

- Statistics assumes statistician
- AI and Machine Learning replace the human.

Practice:

- Humans always in the loop.
- AI focussed on expert systems
- Machine Learning focussed Data Driven Systems
Machine Learning History

- Philosophy: Machine learning, Statistics and AI
  - Statistics assumes statistician
  - AI and Machine Learning replace the human.

- Practice:
  - Humans always in the loop.
  - AI focussed on expert systems
  - Machine Learning focussed Data Driven Systems
Machine Learning History

- Philosophy: Machine learning, Statistics and AI
  - Statistics assumes statistician
  - AI and Machine Learning replace the human.

- Practice:
  - Humans always in the loop.
  - AI focussed on expert systems
  - Machine Learning focussed on Data Driven Systems
If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
The Future

If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
The Future

If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
The Future

If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
The Future

If we are to replace the human in the loop

- Need to understand knowledge transfer.
- Data set shift.
  - Semi supervised learning
- Multitask learning and transfer learning.
- Big task! Cognitive science ... see Nello’s talk yesterday.
  - ... see Nello’s talk yesterday.
  - ... although Nello sensibly and necessarily redefined it as smaller.
What Can we do Today?

Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? Need some application focus.
  - Need to connect the applications and build a useful “toolbox”. Need a focus on tools/technologies.

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? Need some application focus.
  - Need to connect the applications and build a useful “toolbox”. Need a focus on tools/technologies.

- Cognitive science link connects us to Nello’s follow on programme.
What Can we do Today?

Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? Need some application focus.
  - Need to connect the applications and build a useful “toolbox”. Need a focus on tools/technologies.

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? *Need some application focus.*
  - Need to connect the applications and build a useful “toolbox”. *Need a focus on tools/technologies.*

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? **Need some application focus.**
  - Need to connect the applications and build a useful “toolbox”. **Need a focus on tools/technologies.**

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? Need some application focus.
  - Need to connect the applications and build a useful “toolbox”. Need a focus on tools/technologies.

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? **Need some application focus**.
  - Need to connect the applications and build a useful “toolbox”. **Need a focus on tools/technologies**.

All these subjects fall within the scope of the thematic programme.

- Cognitive science link connects us to Nello’s follow on programme.
Less Ambitiously (but still difficult): Incorporate Prior Knowledge with Data

- Mix between expert system approach and neural network “black box”.
- Need to incorporate our knowledge of the system with the data.
- How best to do this?
  - Probably the answer is application/domain dependent?? Need some application focus.
  - Need to connect the applications and build a useful “toolbox”. Need a focus on tools/technologies.

All these subjects fall within the scope of the thematic programme.

- Cognitive science link connects us to Nello’s follow on programme.
Outline

1. Background to Thematic Programme
2. Algorithmic Issues
3. Case Studies
4. Programme of Workshops
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- similarity measures
  - kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- similarity measures
  - kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - differential equations.
How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- similarity measures
  - kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - differential equations.
How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- Probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- Similarity measures
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- similarity measures
  - kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - differential equations.
How are we encoding prior knowledge?

- **Graphs**
  - Web links
  - Biological interactions
  - Social Networks

- **Probability distributions**
  - Bayesian prior distributions
  - Conditional independence structures

- **Similarity measures**
  - Kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - Differential equations.
The Methodologies

How are we encoding prior knowledge?

- Graphs
  - Web links
  - Biological interactions
  - Social Networks

- probability distributions
  - Bayesian prior distributions
  - Conditional independence structures

- similarity measures
  - kernels (kernels on strings, graphs etc.)
  - Learning kernels for knowledge transfer
  - differential equations.

How do these technologies inter-relate?
Outline

1. Background to Thematic Programme
2. Algorithmic Issues
3. Case Studies
4. Programme of Workshops
Common Themes

- Mechanistic models: built from knowledge of physics or human expert knowledge.
- Data driven models: built from data with limited assumptions (such as smoothness)

For illustration we will consider two case studies.

- Computational and Systems Biology
- Language Modelling

But there are many more!!
Mechanistic models: built from knowledge of physics or human expert knowledge.

Data driven models: built from data with limited assumptions (such as smoothness)

For illustration we will consider two case studies.

- Computational and Systems Biology
- Language Modelling

But there are many more!!
Common Themes

- Mechanistic models: built from knowledge of physics or human expert knowledge.
- Data driven models: built from data with limited assumptions (such as smoothness)

Common theme in applications is bridging the gap between these two.

- For illustration we will consider two case studies.
  - Computational and Systems Biology
  - Language Modelling

- But there are many more!!
Common Themes

- Mechanistic models: built from knowledge of physics or human expert knowledge.
- Data driven models: built from data with limited assumptions (such as smoothness)

Common theme in applications is bridging the gap between these two.

- For illustration we will consider two case studies.
  - Computational and Systems Biology
  - Language Modelling
- But there are many more!!
Mechanistic models: built from knowledge of physics or human expert knowledge.

Data driven models: built from data with limited assumptions (such as smoothness)

Common theme in applications is bridging the gap between these two.

For illustration we will consider two case studies.

- Computational and Systems Biology
- Language Modelling

But there are many more!!
Common Themes

- Mechanistic models: built from knowledge of physics or human expert knowledge.
- Data driven models: built from data with limited assumptions (such as smoothness)

Common theme in applications is bridging the gap between these two.

- For illustration we will consider two case studies.
  - Computational and Systems Biology
  - Language Modelling

- But there are many more!!
Mechanistic models: built from knowledge of physics or human expert knowledge.

Data driven models: built from data with limited assumptions (such as smoothness)

Common theme in applications is bridging the gap between these two.

For illustration we will consider two case studies.

- Computational and Systems Biology
- Language Modelling

But there are many more!!
Computational and Systems Biology

- Many different systems: some well studied (cell cycle, signalling pathways), others not.
- Large volume of data large, but sparse relative to complexity of the networks.
  - Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
  - Other quantities (protein concentration) difficult to measure in high throughput.
- Quality and quantity of data improving constantly, but biologists want answers now.
- For some systems there is (noisy) data about which constituents react with what. ChIP on chip.
- The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

- For some systems there is (noisy) data about which constituents react with what. ChIP on chip.
- The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.
Many different systems: some well studied (cell cycle, signalling pathways), others not.

Large volume of data large, but sparse relative to complexity of the networks.

- Some quantities easy to measure (mRNA expression) in high throughput. But still temporal/spatial resolution problems.
- Other quantities (protein concentration) difficult to measure in high throughput.

Quality and quantity of data improving constantly, but biologists want answers now.

For some systems there is (noisy) data about which constituents react with what. ChIP on chip.

The underlying mechanism of interactions is (somewhat) known. Chemical reaction kinetics — but parameters etc. are unknown.

**Rich playground for interaction between data and models of the interactions!**
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
- Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
  - Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
- Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
- Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
- Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Broadly speaking two large areas:

- Computational biologists have focused on constructing models (Kalman filters, “Bayesian” networks, linear models, PCA, AR models) from data. *Data driven approach.*
- Systems biologists have focused on the appropriate differential equation structures for modelling given systems. *Mechanistic approach.*

Challenge is to combine these approaches. Bridge the gap between Mechanistic and Data driven approach.

Consider the mechanistic approach as prior knowledge for the data driven approach.

- Alternative perspective is to use data to “fit” mechanistic models.
Work by Mark Girolami’s Group (Glasgow) [Vyshemirsky and Girolami, 2008]

- Multiple mechanistic models describing a pathway.

The alternative hypotheses about the topology (structure) of the ERK pathway consider single-branched and double-branched options.
Case Study

Work by Mark Girolami’s Group (Glasgow) [Vyshemirsky and Girolami, 2008]

- Multiple mechanistic models describing a pathway.

**Model 1**

The alternative hypotheses about the topology (structure) of the ERK pathway consider single-branched and double-branched options.
Work by Mark Girolami’s Group (Glasgow) [Vyshemirsky and Girolami, 2008]

- Multiple mechanistic models describing a pathway.

The alternative hypotheses about the topology (structure) of the ERK pathway consider single-branched and double-branched options.
Work by Mark Girolami’s Group (Glasgow) [Vyshemirsky and Girolami, 2008]

- Multiple mechanistic models describing a pathway.

The alternative hypotheses about the topology (structure) of the ERK pathway consider single-branched and double-branched options.
Differential Equations

Models are formally defined using systems of ordinary differential equations:

\[
\begin{align*}
\text{EGF}' &= -k_1 \cdot \text{EGF} \cdot \text{EGFR} \\
\text{Rap}^{active}_1 &= \frac{K_{cat12} \cdot \text{Rap}^{inactive}_1 \cdot \text{EPAC}}{K_{m12} + \text{Rap}^{inactive}_1} - \frac{V_{13} \cdot \text{Rap}^{active}_1}{K_{13} + \text{Rap}^{active}_1} \\
\text{MEK}' &= -\frac{K_{cat21} \cdot \text{MEK} \cdot \text{Raf-1}}{K_{m21} + \text{MEK}} - \frac{K_{cat22} \cdot \text{MEK} \cdot \text{BRaf}}{K_{m22} + \text{MEK}}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 kinetic parameters</td>
<td>55 kinetic parameters</td>
</tr>
</tbody>
</table>

SBML descriptions at: http://www.dcs.gla.ac.uk/~vvv
Bayes Factors

Hypotheses Testing: Result

1 : 1,000,000

Model1

Model2

EGF
EGFR
EGFR
Grb2
Sos
Ras
Raf-1
BRaf
MEK
ERK

EGF
EGFR
EGFR
Grb2
CRK
Sos
C3G
Ras
Rap1
Raf-1
BRaf
MEK
ERK

Neil Lawrence ()

Prior Knowledge

28th January 2008 16 / 32
Double branched model has much better support from the experimental evidence, which explains additional data sets on system robustness when inhibiting the dominant left branch and on cancerous mutations (to be published at later date).

BRaf was found to be more active than Raf-1. This is confirmed by a number of publications in biochemical journals.

Given this require further siRNA Knock-Down experiments to confirm dual-branch hypothesis
Double branched model has much better support from the experimental evidence, which explains additional data sets on system robustness when inhibiting the dominant left branch and on cancerous mutations (to be published at later date).

BRaf was found to be more active than Raf-1. This is confirmed by a number of publications in biochemical journals.

Given this require further siRNA Knock-Down experiments to confirm dual-branch hypothesis
Double branched model has much better support from the experimental evidence, which explains additional data sets on system robustness when inhibiting the dominant left branch and on cancerous mutations (to be published at later date).

BRaf was found to be more active than Raf-1. This is confirmed by a number of publications in biochemical journals.

Given this require further siRNA Knock-Down experiments to confirm dual-branch hypothesis.
Results from siRNA Knock Down Experiments

RNAi Knock down favours the dual pathway model

(Hek cells transfected with three different siRNAs (each triplicate) for 36 h, serum starving for another 12h, then treated cells with 1ng/ml EGF for 5 min and harvested cells directly in SDS loading buffer)

Knockdown GRB2 attenuates ERK activation
Knockdown CRK attenuates ERK activation
Further support for dual pathway model
Can test different prior knowledge and choose between.

These ideas transfer to many areas where differential equations encode the information.

Machine Learning has a lot to offer to the area but a lot to learn!

- Sampling
- Approximate inference
- Fast solution of differential equations
Can test different prior knowledge and choose between.

These ideas transfer to many areas where differential equations encode the information.

Machine Learning has a lot to offer to the area but a lot to learn!

- Sampling
- Approximate inference
- Fast solution of differential equations
Computational Bio Conclusions

- Can test different prior knowledge and choose between.
- These ideas transfer to many areas where differential equations encode the information.
- Machine Learning has a lot to offer to the area but a lot to learn!
  - Sampling
  - Approximate inference
  - Fast solution of differential equations
Can test different prior knowledge and choose between.

These ideas transfer to many areas where differential equations encode the information.

Machine Learning has a lot to offer to the area but a lot to learn!

- Sampling
  - Approximate inference
  - Fast solution of differential equations
Can test different prior knowledge and choose between.

These ideas transfer to many areas where differential equations encode the information.

Machine Learning has a lot to offer to the area but a lot to learn!

- Sampling
- Approximate inference
- Fast solution of differential equations
Can test different prior knowledge and choose between.

These ideas transfer to many areas where differential equations encode the information.

Machine Learning has a lot to offer to the area but a lot to learn!

- Sampling
- Approximate inference
- Fast solution of differential equations
A lot of training data.

Data is inherently complex in structure

- Risk of overfit even with very large training data.
- Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).

- Incorporation of this knowledge not always validated by data.
A lot of training data.

Data is inherently complex in structure

- Risk of overfit even with very large training data.
- Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).

- Incorporation of this knowledge not always validated by data.
A lot of training data.

Data is inherently complex in structure
  - Risk of overfit even with very large training data.
  - Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).
  - Incorporation of this knowledge not always validated by data.
A lot of training data.

Data is inherently complex in structure
  - Risk of overfit even with very large training data.
  - Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).
  - Incorporation of this knowledge not always validated by data.
A lot of training data.

- Data is inherently complex in structure
  - Risk of overfit even with very large training data.
  - Requires constraining the space of fitted models.

- Lots of existing expert knowledge (e.g. linguistic constraints).
  - Incorporation of this knowledge not always validated by data.
A lot of training data.

Data is inherently complex in structure
  ▶ Risk of overfit even with very large training data.
  ▶ Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).
  ▶ Incorporation of this knowledge not always validated by data.
A lot of training data.

Data is inherently complex in structure

- Risk of overfit even with very large training data.
- Requires constraining the space of fitted models.

Lots of existing expert knowledge (e.g. linguistic constraints).

- Incorporation of this knowledge not always validated by data.

Rich playground to study interplay between prior knowledge and data!
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) *rules* defined by experts.
  - Could be seen as an *extreme case* of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic *n*-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Suprisingly effective in many cases (widely used in speech recognition).
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) *rules* defined by experts.
  - Could be seen as an *extreme case* of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

- Opposite extreme: data only, no “prior knowledge”.
  - *e.g.* basic $n$-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Suprisingly effective in many cases (widely used in speech recognition).
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) \textit{rules} defined by experts.
  - Could be seen as an \textit{extreme case} of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic $n$-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Surprisingly effective in many cases (widely used in speech recognition).
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) *rules* defined by experts.
  - Could be seen as an *extreme case* of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic *n*-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Suprisingly effective in many cases (widely used in speech recognition).
Historic Perspective (Language)

Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) rules defined by experts.
  - Could be seen as an extreme case of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

Data Driven Systems

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic n-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Surprisingly effective in many cases (widely used in speech recognition).
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation *etc.*) *rules* defined by experts.
  - Could be seen as an *extreme case* of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

Data Driven Systems

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic *n*-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Surprisingly effective in many cases (widely used in speech recognition).
**Rule Based Systems**

- Before c. 1990 mainstream systems (for parsing translation *etc.*) *rules* defined by experts.
  - Could be seen as an *extreme case* of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

**Data Driven Systems**

- Opposite extreme: data only, no “prior knowledge”.
  - *e.g.* basic *n*-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Suprisingly effective in many cases (widely used in speech recognition).
Rule Based Systems

- Before c. 1990 mainstream systems (for parsing translation etc.) rules defined by experts.
  - Could be seen as an extreme case of “prior knowledge”. No training data to “tune” the encoded knowledge.
  - Standard linguistics is about the discovery of such rules.

Data Driven Systems

- Opposite extreme: data only, no “prior knowledge”.
  - e.g. basic n-gram language models — collect statistics over 3-grams and use these for computing probability of a test sentence.
  - Suprisingly effective in many cases (widely used in speech recognition).
Prior knowledge is encoded, but implicitly

- Even in “data driven” systems a lot of prior knowledge is *implicit* in the models/representations.
  - $n$-gram models incorporate “smoothing”. Algorithms attempt to exhibit good generalisation properties.
  - In statistical machine translation pioneer IBM systems were generative models that perform operations (e.g. fertility, distortion) that have implicit linguistic motivations. Once the generative model is defined tuning of parameters is done by looking at the data.
  - In current discriminative models for NLP, models learn to discriminate good outputs from bad on the basis of training data alone, but often the underlying feature functions are complex (e.g. in phrase-based SMT), and designed carefully to address the problem at hand.
Prior knowledge is encoded, but implicitly

- Even in “data driven” systems a lot of prior knowledge is *implicit* in the models/representations.
  - *n*-gram models incorporate “smoothing”. Algorithms attempt to exhibit good generalisation properties.
  - In statistical machine translation pioneer IBM systems were generative models that perform operations (*e.g.* fertility, distortion) that have implicit linguistic motivations. *Once* the generative model is defined tuning of parameters is done by looking at the data.
  - In current discriminative models for NLP, models learn to discriminate good outputs from bad on the basis of training data alone, but often the underlying feature functions are complex (*e.g.* in phrase-based SMT), and designed carefully to address the problem at hand.
Prior knowledge is encoded, but implicitly

- Even in “data driven” systems a lot of prior knowledge is *implicit* in the models/representations.
  - *n*-gram models incorporate “smoothing”. Algorithms attempt to exhibit good generalisation properties.
  - In statistical machine translation pioneer IBM systems were generative models that perform operations (*e.g.* fertility, distortion) that have implicit linguistic motivations. *Once* the generative model is defined tuning of parameters is done by looking at the data.
  - In current discriminative models for NLP, models learn to discriminate good outputs from bad on the basis of training data alone, but often the underlying feature functions are complex (*e.g.* in phrase-based SMT), and designed carefully to address the problem at hand.
Prior knowledge is encoded, but implicitly

- Even in “data driven” systems a lot of prior knowledge is *implicit* in the models/representations.
  - *n*-gram models incorporate “smoothing”. Algorithms attempt to exhibit good generalisation properties.
  - In statistical machine translation pioneer IBM systems were generative models that perform operations (*e.g.* fertility, distortion) that have implicit linguistic motivations. *Once* the generative model is defined tuning of parameters is done by looking at the data.
  - In current discriminative models for NLP, models learn to discriminate good outputs from bad on the basis of training data alone, but often the underlying feature functions are complex (*e.g.* in phrase-based SMT), and designed carefully to address the problem at hand.
Prior knowledge is encoded, but implicitly

- Even in “data driven” systems a lot of prior knowledge is *implicit* in the models/representations.
  - *n*-gram models incorporate “smoothing”. Algorithms attempt to exhibit good generalisation properties.
  - In statistical machine translation pioneer IBM systems were generative models that perform operations (*e.g.* fertility, distortion) that have implicit linguistic motivations. *Once* the generative model is defined tuning of parameters is done by looking at the data.
  - In current discriminative models for NLP, models learn to discriminate good outputs from bad on the basis of training data alone, but often the underlying feature functions are complex (*e.g.* in phrase-based SMT), and designed carefully to address the problem at hand.
Towards Explicit Prior Knowledge

- **Minimum Discription Length**
  - e.g. in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - e.g. in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - e.g. in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - e.g. in topic modelling: automatic determination of the number of underlying topics
    - e.g. in parsing: automatic determination of number of nonterminals in PCFG
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parametric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - e.g. in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - e.g. in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - e.g. in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parametric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - e.g. in topic modelling: automatic determination of the number of underlying topics
    - e.g. in parsing: automatic determination of number of nonterminals in PCFG

Neil Lawrence () 28th January 2008
Towards Explicit Prior Knowledge

- **Minimum Discription Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics
    - *e.g.* in parsing: automatic determination of number of nonterminals in PCFG

Neil Lawrence ()
Prior Knowledge
28th January 2008
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parametric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics.
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG.
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics.
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG.
Towards Explicit Prior Knowledge

- **Minimum Discription Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics.
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG.
Towards Explicit Prior Knowledge

- **Minimum Description Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics.
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG.
Towards Explicit Prior Knowledge

- **Minimum Discription Length**
  - *e.g.* in Grammar inference [Grünwald, A minimum description length approach to grammar inference]

- **Bayesian Parametric**
  - *e.g.* in Topic Modelling [Blei et al., 2003, Latent Dirichlet Allocation]
  - *e.g.* in Language Model Smoothing [MacKay and Peto, 1994, A Hierarchical Dirichlet Language Model]

- **Currently hot: Bayesian non-parameteric**
  - Dirichlet processes, Chinese restaurant processes, ...
  - Allow automatic selection of model complexity
    - *e.g.* in topic modelling: automatic determination of the number of underlying topics.
    - *e.g.* in parsing: automatic determination of number of non-terminals in PCFG.
Surge of Publications

From Klein and Laing 2007

- **Topic Modelling**
  - finite Bayesian model; variational [Blei et al., 2003]
  - HDP-based model; sampling [Teh et al., 2006]

- **Language Modelling**
  - Pitman-Yor $\rightarrow$ power-law; sampling [Goldwater et al. 2005]
  - Kneser-Ner $\leftrightarrow$ Pitman-Yor; sampling [Teh, 2006]

- **POS Induction** using a finite Bayesian HMM
  - Collapsed sampling [Goldwater, Griffiths, 2007]
  - Variational [Johnson, 2007]

- **Parsing** using nonparametric grammars
  - Collapsed sampling [Johnson et al., 2006]
  - Collapsed sampling [Finkel et al. 2007]
  - Variational stick-breaking representation [Liang, et al., 2007]

- **Coreference resolution**
  - Supervised clustering; collapsed sampling [Daume, Marcu, 2005]
  - HDP-Based model; sampling [Haghighi, Klein, 2007]
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- e.g. Markov logic networks [Richardson and Domingos, 2006]

  Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge).
  e.g. “If student is author of publication, professor is co-author.”

  These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.

  Good results on such tasks as link prediction. Few applications to “core” NLP so far.
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- e.g. Markov logic networks [Richardson and Domingos, 2006]
  - Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge):
    - e.g. “if student is author of publication, professor is co-author ...”
  - These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.
  - Good results on such tasks as link prediction. Few applications to “core” NLP so far.
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- e.g. Markov logic networks [Richardson and Domingos, 2006]
  - Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge):
    - e.g. “if student is author of publication, professor is co-author ...”
  - These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.
  - Good results on such tasks as link prediction. Few applications to “core” NLP so far.
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- **e.g.** Markov logic networks [Richardson and Domingos, 2006]
  - Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge):
    - e.g. “if student is author of publication, professor is co-author ...”
  - These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.
  - Good results on such tasks as link prediction. Few applications to “core” NLP so far.
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- e.g. Markov logic networks [Richardson and Domingos, 2006]
  - Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge):
    - e.g. “if student is author of publication, professor is co-author ...”
  - These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.
  - Good results on such tasks as link prediction. Few applications to “core” NLP so far.
By and large the approaches described concentrate on “generic” models of prior knowledge, not on “specific” expert linguistic knowledge.

One possible approach to that might be in the line of statistical relational learning

- e.g. Markov logic networks [Richardson and Domingos, 2006]
  - Main idea: ask experts to formulate their beliefs through first order logical formulas (a form of prior knowledge):
    - e.g. “if student is author of publication, professor is co-author ...”
  - These formulas become binary features on possible worlds. Can be false in a given world and are associated with weights that are learnt on the basis of data.
  - Good results on such tasks as link prediction. Few applications to “core” NLP so far.
Outline

1. Background to Thematic Programme
2. Algorithmic Issues
3. Case Studies
4. Programme of Workshops
Learning in Computational and Systems Biology
Glasgow, March 27th - 28th 2008, Co-located with MASAMB 2008
Co-organized by Mark Girolami and Simon Rogers

- Combining data with models in systems biology
  - how to estimate differential equation parameters
  - how to estimate difficult to measure chemical species

- Validation of model structure: which hypothesis is correct?
  - Hypothesis in the form of non-linear differential equations.
  - Sampling approaches to efficient computation of Bayes factors.
Approximate Inference in Stochastic Processes and Dynamical Systems
Co-organizers: Cedric Archambeau, Neil D. Lawrence, Andrew Stewart, John Shawe-Taylor

- Focus on combining differential equations with data.
  - mainly stochastic (applications in climate, weather, computational biology) but also some ODEs.

- Bring together Machine Learning technologies with statistics, physics, control etc.

- Focus on both applications and methodologies (day each).
Bayesian Research Kitchen
Early June 2008, Lake District, U.K.
Co-organizers: Neil Lawrence, Joaquin Quinonero Candela

- Small gathering
- Bayesian “reality check”.
  - Focus on future directions for Bayesian research.
- May lead to larger Participation Workshop in Second Half of Programme (c.f. GPRT/GPIP).
Draft call for papers:

- Prior knowledge for language modelling, parsing, translation.
- Topic modelling for document analysis and retrieval.
- Parametric and non-parametric Bayesian models in NLP.
- Graphical models embodying structural knowledge of texts.
- Complex features/kernels that incorporate linguistic knowledge; kernels built from generative models.
- Limitations of purely data-driven learning techniques for text and language applications.
- Typology of different forms of prior knowledge for NLP (knowledge embodied in generative Bayesian models, in MDL models, in ILP/logical models, in linguistic features, in representational frameworks, in grammatical rules ...).
- Formal principles for combining rule-based and data based approaches in NLP.
Other Workshops

- Mining and Learning on Graphs (July, 2008, Finland??)
- Machine Learning for Systems Biology (September, 2008, Belgium??)
- Follow up events from September 2008 - March 2009.
- Apologies for being too Bayesian!
- **Over to you!!**


