

電気通信大学 平成18年度シラバス

授業科目名	データマイニング特論		
英文授業科目名	Topics in Data Mining		
開講年度	2006年度	開講年次	
開講学期	後学期	開講コース・課程	博士前期課程
授業の方法		単位数	2
科目区分	電気通信学研究科-システム工学専攻-専門科目		
開講学科・専攻	システム工学専攻		
担当教官名	山本 渉		
居室	西5-601		

公開E-Mail	授業関連Webページ
山本	<a href="http://www.pclab.se.uec.ac.jp/">http://www.pclab.se.uec.ac.jp/</a>

<p><b>【主題および達成目標】</b></p> <p>必須項目。</p> <p>データマイニングに現れる種々の手法を、基本的な事項から始めて、手法の実装に必要な理論を解説した後、実際の問題に適用するまでの一連の過程を通じて、学んでもらいます。毎年、次のようなトピックの中から、一つないし複数を選んで、講義します。</p> <ul style="list-style-type: none"> <li>• Tree-based models and their enhancements (CART, boosting, bagging, etc.)</li> <li>• Kernel-based machines (support vector machine)</li> <li>• Network models (graphical modelling, Bayesian network, etc.)</li> </ul> <p>昨年度は、ベイズ流の統計的学習理論の中から特に、ベイジアン・ネットワークを主要なトピックとして扱いました。今年度は、カーネルマシンか、樹形モデルを扱うことを考えています。講義の準備状況に応じて、このシラバスは更新されます。</p> <p>9月17日追記：今年度(2005年度)は、前半はSubsampling、後半はBayesian Networkを扱う予定です。Subsamplingは、大量のデータがあるときにそれらを用いて、特定の確率分布を仮定せずに統計的な推測論を実施するアプローチの一つです。前半はこの技術の基本的なことを紹介します。後半のBayesian Networkは近年注目されている知識のネットワーク表現モデルの一つで、連続データか離散データの場合に限って紹介することになると思います。</p>
--

## 電気通信大学 平成18年度シラバス

### 【前もって履修しておくべき科目】

なし。

### 【前もって履修しておくことが望ましい科目】

特にありませんが、確率・統計に関する基本的な知識は得ていることを前提に、講義を進めることがあります。

### 【教科書等】

教科書は特にありません。次の二冊を参考書とします。

- ・ ``Subsampling'', Dimitris N. Politis, Joseph P. Romano, Michael Wolf. Springer, 1999  
(Springer series in statistics).
- ・ ``Learning Bayesian Network'', Richard E. Neapolitan, Pearson Prentice Hall, 2004.

### 【授業内容とその進め方】

資料を配布することもありますし、講義で使用する図表は予め用意しておくこともあります。ただ、ノートを取ることは知識を吸収するための一手段として未だに有効、と考えているので基本的には板書しながら講義を進めます。

### 【成績評価方法及び評価基準(最低達成基準を含む)】

講義時間中にプレゼンテーションを課したり、課題を示してそれに取り組んだレポートを提出してもらうなど、講義内容に関連したタスクの成果により評価します。

### 【オフィスアワー：授業相談】

質問などは随時受けつけます。講義後の休み時間中に質問されるのは歓迎です。それ以外の時間については、質問をメールで送信しながら、アポイントメントをとってください。

【学生へのメッセージ】

データマイニングは様々な分野の技術を柔軟に応用する分野です。講義ではデータを扱う作業全般の中で、特にマイニングのステージだけにスポットを当てますが、データのクレンジングのステージもデータウェアハウスにデータを流し込む併合のステージも、それぞれ興味深い技術が用いられます。

データマイニング全般をカバーはせず、トピックを掘り下げることにはなりますが、興味を持って臨んでもらえればと思います。

【その他】

私の居室は西五号館の601室です。