

## データ分布と予測 再試験課題

【課題】統計を現実に応用した例について調べ、その内容をレポートにまとめる。以下の2つの内どちらか1つを選んで調査せよ。

レポート:A4用紙5枚以上。Word等で作成し、PDFファイルに変換してメール添付。

提出先:メール宛先 khotta

〆切:2008/2/20(水) 18:00 必着

選択1:スパム対策で使われる「ベイズ理論」を応用したフィルタリング技術について、どのような理論を用いてどのように実行されているのか? 具体的なアルゴリズムを Web や文献などから1つ探して、そのアルゴリズムの動作について解説せよ。

また、もととなる「ベイズの定理」についても解説せよ。

選択2:Googleの検索エンジンには「Page Rank」とよばれるページ重要度を自動判定する技術が使われている。「Page Rank」は、「良質なページからリンクされているページは良質なページである」という考えのもとにハイパーリンクの張られ方をもとにしたページの重要度判定を行っている。以下の「Page Rank」に関する論文を読み、「Page Rank」とは具体的にどのような方法で求めているのか、解説せよ。論文は Web で探すこと。

``The PageRank Citation Ranking: Bringing Order to the Web'',1998