

データ分布と予測 再試験課題

【課題】 データ分布・統計・確率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。対象は以下の2つの内好きな方を選んで調査せよ。

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: 2007/2/22(木) 18:00 必着

選択 1: スпам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, どのように実行されているのか? 具体的なアルゴリズムを探して, そのアルゴリズムの動作について解説せよ.

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。なお, この論文はこの課題と同ページに「選択 2 資料」として掲載している。

“The PageRank Citation Ranking: Bringing Order to the Web”, 1998