

データマイニングによる次世代eラーニングシステムの開発

背景・目的

閲覧記録や操作ログ、問題の正誤、所要時間などの膨大な学習履歴データが得られる

履歴情報を分析する上でのLMSやSCORMの整備は徐々になされつつある

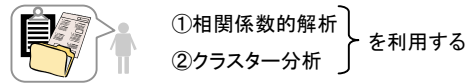


◎履歴情報の活用に関する研究はあまり進んでいないのが現状

現在、以下のような活用方法が検討されている

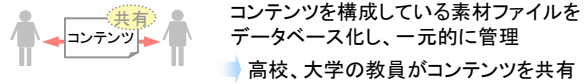
- 学習者の弱点分析の知識を獲得し、フィードバックの提供
- 学習者の異常学習プロセスの検知
- 学習者個人の学習状況に応じたカリキュラムの動的生成

アンケートや試験結果、学習履歴データの解析を行うことで、次世代eラーニングシステムに向けて改善を図る



本研究室において提案・構築した

カスタムアレンジ(CA)型eラーニングシステム*の評価を行う



シミュレータコンテンツをCA型eラーニングシステムで活用し、その学習履歴や試験結果の解析から次世代eラーニングシステムの検討を行う。

*Ref:清水賢代他 第4回日本WebCTユーザカンファレンス予稿集:85-88

学習履歴の取得

実施方法

対象: 物理学概論を受講する理学部

数物科学科・物質生物学科1・2年生 90名

高校時に物理履修者 41名
未履修者 49名

評価方法: ①各演習問題毎に関連性のあるシミュレータを表記

- ① 閲覧状況、正答率の解析
- ② 期末テストの結果
- ③ アンケートの実施

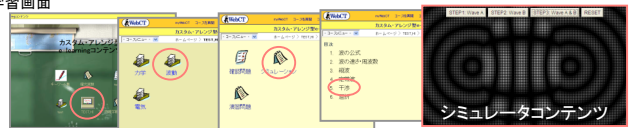
演習問題

問題にシミュレータを表記

(参考: TEST_HI > 波動 > 光 > シミュレーション > ニュートンリング) 参照した場合は、右欄にチェック

2. 屈折率1.3の水中において、平面ガラスの上に平凸レンズの凸面を下にして置き、平面に垂直に波長 $6.0 \times 10^{-4} \text{m}$ の光を当てたところ、同心円状のしま模様が見えた。ある暗線の直径は 2.0cm であったが、この装置を空気中に出すと、この暗線の直径はいくらになるか。

学習画面



相関的解析

解析方法

相関係数

データのバラツキの大きさに関する情報を除き、直線傾向への集中度を評価する指標ができる

$$r_{xy} = \frac{v_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

r_{xy} : 相関係数
 v_{xy} : 共分散
 s_x, s_y : 標準偏差

相関係数の検定

相関係数が対象としている2つの変数間の線形関係を普遍的に表しているかどうかの信頼性を確認する

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

t : 検定統計量
 r : 相関係数
 n : データ数

演習問題総得点と期末テスト結果

未履修者	期末テスト得点	演習問題総得点
閲覧無し	41.4	58.0
閲覧有り	45.9	64.9

履修者	期末テスト得点	演習問題総得点
閲覧無し	45.9	59.1
閲覧有り	45.5	64.6

(全て100点満点)

演習問題総得点とテストの相関

	未履修者	履修者	全体
相関係数	0.49*	0.54*	0.57*
閲覧無し	0.50**	0.57*	0.42*
閲覧有り	0.24	0.43	0.54*

(*1%有意, **5%有意)

シミュレータの閲覧者の方が期末テスト、演習問題総得点で優位性が確認できる。

履修者の相関係数が高いことから、基礎的な知識を有する者に対して有効なコンテンツだといえる。

予測された関連性しか見えない

▶ クラスター分析を用いて解析

クラスター分析

K-means法

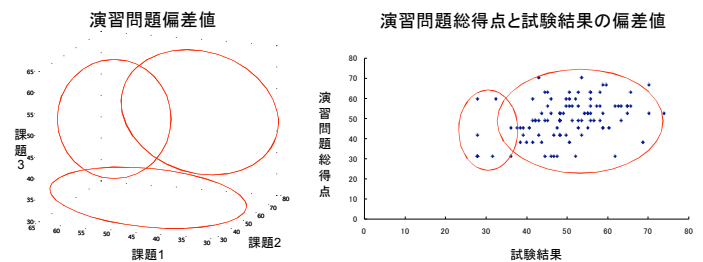
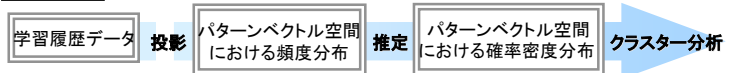
- ① クラスタ中心 $C_1(l), C_2(l), \dots, C_k(l)$ を用いて、全標本をそれに近いクラスタの中心に所属させるという形で行う
- ② 新しいクラスタ中心 $\{C_j(k+1), j=1, 2, \dots, k\}$ を、 C_j に所属した全標本の平均値によって定める
 $C_j(k+1) = \frac{1}{N_j(k)} \sum X_i$
- ③ クラスタ中心が変化しない状態になったら終了

K-means法応用例

- 健康診断データベースからの疾病リスクの算出
- 病気の治療・創薬へ繋がるタンパク質研究
- 情報検索の目的によるWeb上の文章の分割

K-means法を利用し、試験結果や学習履歴データの分布を調べ、各クラスタの特徴を観察する。

解析方法



学習履歴データや試験結果をパターンベクトル空間に投影することにより、クラスタを確認する事ができた。

Ref: 大倉恵子 確率分布に基づく医療データの解析に関する研究

まとめ・今後の課題

- シミュレータコンテンツを演習課題で活用し、その学習履歴データから、データマイニングによる次世代eラーニングの検討を行った。
- 学習履歴を相関的解析を行った結果、物理履修者に対し大変効果が高い結果が得られた。
- 試験結果や学習履歴データをK-means法でクラスター分析し、クラスタの持つ意味を解析していく。