G2-2 知識間の語彙出現分布に基づく学習順序推定手法の提案

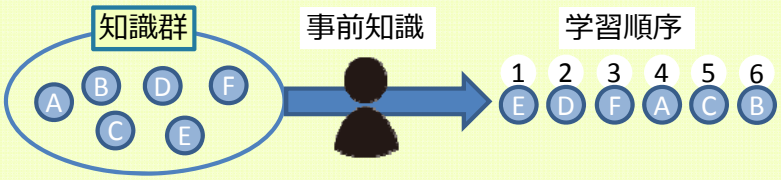
DEIM Forum2016

久保直人 †1 佐藤哲司 †2

†1筑波大学　†2筑波大学図書館情報メディア系

背景

・学習順序は対象となる知識群の関係性に基づいて人手でつくられる
・そのため事前の知識習得が必須となる



目的

・人手を必要とする以上学習者に合わせた柔軟な学習順序をつくれない
→最短学習順序、網羅的学習順序など

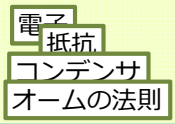
「知識間の関係性を**語彙出現分布**を用いて抽出し、事前知識を必要としない学習順序の推定手法を提案する」

提案手法

用語定義

・教科書などの章立て
→**学習順序**

・ある事柄などについて記述された文章
→**知識**



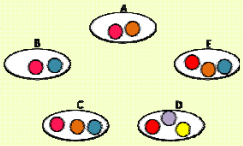
・知識に出現する語彙
→**要素知識**

知識群の関係性＋並べ替え手法

・**重要度**
知識群において**出現しやすい要素知識**を多くもつ知識は**重要性が高い**

知識群A～Eにおいて出現しやすい要素知識
基礎的な要素知識

知識群A～Eにおいて出現しにくい要素知識
専門的な要素知識

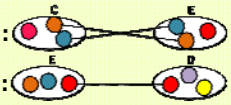


・**定式化**
知識dに出現する要素知識$s_{d,x}(1 \le x \le n_d)$とすると
重要度$I_d = \sum_{x=1}^{n_d}$ 要素知識$s_{d,x}$の全体における出現頻度　(1)

・**並べ替え手法Isort**
重要度が高い知識ほど先に学ぶべきであるとし知識を重要度が降順になるように並べ替える
→のちの知識習得に必要となる要素知識を先に学ぶことにより全体を通じてスムーズな学習順序

・**難易度**
ある知識を習得したあとに他の知識を学ぶとき共通する要素知識が多くかつその割合が多いほど難易度が低い

難易度低：
難易度高：



・**定式化**
知識Aから知識Bへ遷移する難しさ

$$難易度 D_{A,B} = 1 - \frac{|知識_Aの要素知識群 \cap 知識_Bの要素知識群|}{|知識_Bの要素知識群|} \quad (2)$$

・**並べ替え手法Dsort**
1. スタート知識を決定し、そこから難易度がもっとも低い知識を選択
2. 選択された知識から難易度が最も低い知識を選択
（全ての知識が選択されるまで2を繰り返す）
3. 全ての知識をスタート知識としたとき、得られた順序の難易度分散が最も低い順序を並べ替え結果とする
→遷移しやすい順序であり難しさが一定である学習順序

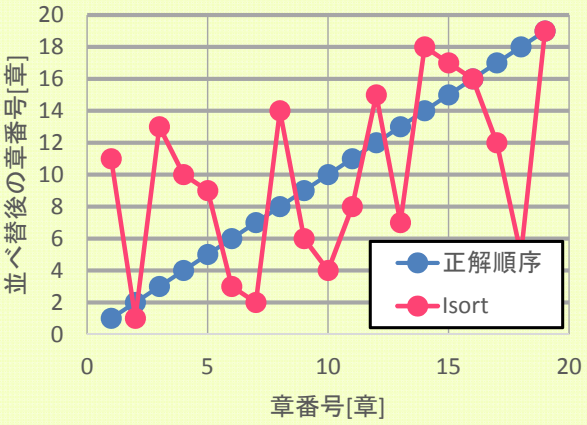
評価実験

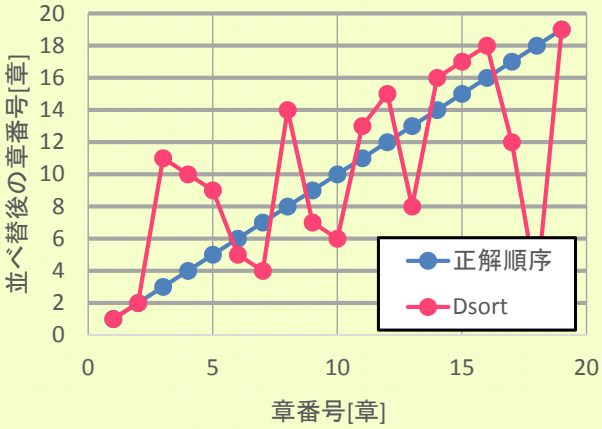
実験用データ

基礎物理学の教科書（pdf）
・章数＝知識
：19
・語彙＝要素知識
：各章に含まれる単語
Wikipedia上に記事として存在するもの

実験方法

教科書を対象に本手法を用いて元の章立ての再現性スピアマンの順位相関係数を用いて評価





考察＋課題

・正解より大きく外れる章は共通して語彙数の少ない傾向にある
正解より大きく外れた章　：3章（686）、10章（404）、18章（210）
正解に近かった章　：14章（1321）、15章（1277）、19章（1842）
・内容の関連度が高く語彙数の多い章は隣合う関係を再現できた
14章:Electricity（1321）15章:Magnetic and Electromagnetism（1277）16章:Electronics（1260）
→語彙の拡張による語彙数増加が課題
・DsortがIsortより順位相関係数が高い
→2つの章の間を評価するのでより隣り合う関係が再現できた

並べ替え手法	順位相関係数	有意確率	検定結果
Isort	0.442（弱い相関）	0.0581	0.01<P<0.05
Dsort	0.577（やや強い相関）	0.00966	P<0.01