

グループ討議演習支援システムにおける学生のコメント分類に関する研究

吹沢将哉†, 渡辺博芳†, 高井久美子†, 鈴木和典‡

†帝京大学理工学部

‡アクリーグ株式会社

1. はじめに

我々は、学生が宿題として取り組んだ各自の解答を説明してグループでの解を考える対面でのグループ討議を支援するシステム（以下 GDS とする）を実装した。GDS はグループ討議の準備、実施、振り返りの各過程において教員と学生の活動を支援する[1]。その中で、グループ討議の結果としてグループの解とコメント(主に疑問点)を提出するが、現状では、学生のコメントを教員に提示する際、グループごとに提示している。これに対して、類似のコメントをまとめるなど、コメントの内容に着目して分類して提示することで、教員が全体を把握しやすくなると考えられる。

そこで、本研究では同様のコメントをまとめるための分類方法を検討している。これまでに、キーワードを用いる方法とクラスタリングによる方法を検討した。これら 2 つの方法について、過去の演習で提出されたコメントを用いて行った実験結果について報告する。

2. 対象とするコメント分類

本学の実習での 1 回のグループ討議では、8~10 程度のグループから合計で 10~20 のコメントが提出される。これらをグループごとではなく、同様なコメントをまとめて表示することが分類の目的である。たとえば、あるグループが「ユースケースにおける汎化の方法がよくわからなかった」というコメントを、別のグループが「ユースケース同士の汎化で結ぶとどういう意味を持つことになるのかわからなかった」というコメントを提出した場合に、「ユースケース・汎化」というカテゴリ名を付加して、まとめるといったことを目指す。

3. キーワードによる分類

キーワードによる方法では、あらかじめ設定したキーワードと自動生成したキーワードを用いコメント分類を行っている。分類手順を図 1 に示す。

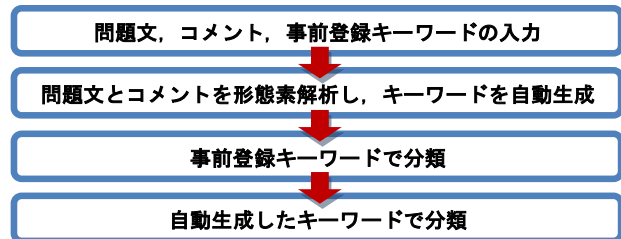


図1 キーワードによる分類

3.1. 事前登録キーワード

教員が、重要度が高く、カテゴリ名として適切であると考えられる単語を事前に設定しておき、分類に使用する。例として、UML に関する授業の際には、「アクター」「ユースケース」「シーケンス」などがあげられる。

またこの際、同じ意味を持つ単語を同じキーワードとして設定することができる。例として「包含」と「include」は語句の違いだけで、意味するものは同じであるため、同じ意味を持つキーワードとして設定でき、コメント分類を行う際にも同じものとして扱うことができる。

3.2. キーワードの自動生成

問題文・コメントを形態素解析し、両方で使用されている名詞をキーワードとして生成する。自動生成の際、名詞が文中で連続して使用されている場合、一つの名詞として認識するように設定しておく。例として、「出席登録」を形態素解析すると、「出席」「登録」という別々の名詞として認識される。しかし、上記の処理を行うことで「出席登録」という複合名詞として認識される。

3.3. コメント分類

事前登録キーワードが含まれているコメントをまとめることで分類を行う。その後、分類できなかったコメントを自動生成したキーワードで分類する。自動生成したキーワードで分類を行う際、複合名詞で結合回数が多いものから優先して分類に使用する。また、結合回数と同じ場合、問題文で使用されている回数が多いものから順に使用する。

4. クラスタリングによる分類

クラスタリングによる方法では、TFIDFを用いて単語に重み付けを行い、分類を行う。分類手順を図2に示す。



図2 クラスタリングによる分類

4.1. 同義語置換

事前に設定しておいた同義語リストを用いて各文の同義語を一つに置換する。例として、「包含」と「include」のように同様の意味を持つ単語の場合、「包含」に置換する。

4.2. 名詞の抽出

上記の処理後の各文に対して形態素解析を行い、名詞のみを抽出する。

4.3. 各名詞の TFIDF を計算

TFIDF を使用して各名詞に重み付けを行う。本研究では以下の式で求めている。

$$TFIDF = tf * idf$$

$$tf(t, d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}$$

tf(t,d) : 文書 d 内のある単語 t の TF 値
 n_{t,d} : ある単語 t の文書 d 内での出現回数
 ∑_{s∈d} n_{s,d} : 文書 d 内のすべての単語の出現回数の和

$$idf(t) = \log \frac{N}{df(t)} + 1$$

idf(t) : ある単語 t の IDF 値
 N : 全文書数
 df(t) : ある単語 t が出現する文書の数
 ※本研究ではコメント数が文書に相当する。

4.4. クラスタリング

単語の出現回数が少ない（二個以下）コメントをクラスタリング対象から除外してクラスタリングを行っている。クラスタリング手法は、

階層的手法の一つであるウォード法を用いている。また、距離法としてはコサイン距離、クラスタ数としてキーワードによる分類結果のカテゴリ数を使用している。

4.5. コメント追加処理

クラスタリング結果に対し、単語数が少ないコメントを追加する。追加方法は、TFIDF が高い単語を選択し、その単語が出現する他のコメントが属しているクラスタに追加する。どこにも一致するものがない場合、新しいグループに分類する。

5. 実験結果

本研究では、本学で 2009 年に行われた情報システム実習 2 (UML を用いたオブジェクト指向モデリング) の二つのクラス (A, B) に対する三つの課題 (ユースケース図, クラス図, シーケンス図) を対象に、キーワードによる方法と、クラスタリングによる方法の二つの手法で分類を行った。その結果に対し、2 名の教員が適切に分類できているかどうかを評価し、適切に分類できたコメントの割合を分類精度として求めた。実験結果の評価を表 1 に示す。

どちらの手法も適切に分類できているのは全体として 2/3 程度であり、ほぼ同等の結果である。だが、対象とするデータにより分類精度に優劣がある。

表 1 実験結果

データ名	コメント数	分類精度	
		キーワード分類	クラスタリング
ユースケース図A	22	71%	64%
ユースケース図B	14	64%	64%
クラス図A	18	72%	72%
クラス図B	18	78%	56%
シーケンス図A	15	33%	67%
シーケンス図B	9	78%	67%
平均	96	66%	65%

6. おわりに

学生のコメント分類を二つの方法で行った。どちらも全体として 65%程度の分類精度である。今後は、より分類精度を向上させるため、本研究で用いた二つの手法を組み合わせることを検討している。

参考文献

[1]李依朔, 渡辺博芳, 高井久美子, 佐々木茂 : グループ討議演習支援システムの設計と実装, 情報処理学会第 74 回全国大会, 3ZE-9, (2012)