

自己組織化マップを用いた 電子会議の議事録作成支援に関する研究

中村健二[†] 田中成典[‡] 池辺正典[†] 吉村智史[†] 大谷渉[‡]

関西大学大学院総合情報学研究科[†] 関西大学総合情報学部[‡]

1. はじめに

近年，インターネットの普及に伴い，政府や自治体が市民の意見を政策に反映するための電子会議の利用数が増加している．しかし，電子会議の文書は，ツリー形式やスレッド形式などで投稿されるため，会議内容が散在しており，把握が困難である．そのため，文書を要約して内容把握を支援する研究[1][2]や文書の分類により議論内容を整理する研究[3]がなされている．しかし，文書を要約する研究では，特徴的な文書のみを対象とするため，稀少な文書が要約結果に反映されない．稀少な文書は，重要な意見につながる可能性[4]があるため，議論の内容を把握するために不可欠な情報である．また，文書を分類する研究では，カテゴリ内に複数の話題が散在するため，カテゴリごとに最適な教師データを作成することができず，分類精度が低いという問題がある．そこで，本研究では，自己組織化マップ[5]を用いて電子会議の投稿文書を分類し，分類したグループ毎に議論の議題を抽出することで議事録の作成を支援する手法を提案する．

2. システムの概要

本研究では，電子会議の各投稿文書に重要度で重み付けを行い，自己組織化マップを用いて分類する．そして，分類結果から議事録作成を支援することを実現する．自己組織化マップを用いることにより，投稿文書の類似度を基に自動的にカテゴリを生成し，分類することが可能となる．本システムの流れを図1に示す．本システムの入力は電子会議の投稿文書，出力はグループごとに分類された投稿文書と議題となる．本システムは，1) 文書重み付け機能，2) 文書分類機能，3) 議題抽出機能の3つの機能により構成される．

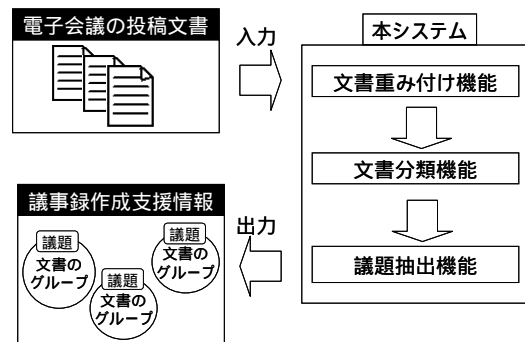


図1 システムの流れ

2. 1 文書重み付け機能

文書重み付け機能では，電子会議の各文書に重み付けを行う．まず，電子会議の文書を形態素解析器である茶笥を用いて形態素に分解し，その結果から名詞のみを抽出する．次に，電子会議の各文書で抽出した名詞の TF-IDF (Term Frequency-Inverse Document Frequency) 値を算出する．TF-IDF 値は，各文書における任意の単語の重要度を表す．最後に，各名詞の TF-IDF 値を正規化し，各名詞，TF-IDF 値をそれぞれ自己組織化マップの要素として電子会議の文書に付与する．

2. 2 文書分類機能

文書分類機能では，文書重み付け機能で算出した値を自己組織化マップにかけ，文書の分類を行う．本機能による分類では，各要素同士を比較し，文書間の類似度を算出する．そして，類似度によるグループ分けを行う．グループ分けは，文書間の類似度が，一定の閾値以内であれば1つのグループとして扱う手法を用いる．

2. 3 議題抽出機能

議題抽出機能では，文書分類機能で生成したグループから議題を抽出する．グループに含まれる単語のうち，ストップワードを除く頻出語の上位4つを抽出し，グループの議題とする．頻出語は，グループ内で数多く議論されている内容を取得するため，各名詞の DF 値で判定する．ここで，ストップワードとは，情報検索の一般的な単語であるため，検索語として不適切な単語を意味する．例えば，指示語，代名詞や

Research on Supporting to Make Proceedings of the Electronic Meeting Using Self-Organizing Maps

[†]Kenji Nakamura, Masanori Ikebe, Satoshi Yoshimura
Graduate School of Informatics, Kansai University, 2-1-1
Ryouzenji-cho Takatsuki-shi, Osaka 569-1095, Japan

[‡]Shigenori Tanaka, Wataru Otani
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-
cho Takatsuki-shi, Osaka 569-1095, Japan

数字などがストップワードである。

3. システムの実証実験と考察

本研究で考案した手法の有効性を検証するために、実証実験を行った。実証実験では、実際に分類されたグループ内の文書が、正しく分類されているかと議題が正しく抽出されているかの2点を判定するために、本システムの分類精度と議題の抽出精度を検証する。

3.1 実証実験

本実験では、神奈川県藤沢市が行っている藤沢市市民電子会議室内の文書を分類対象として実験を行った。対象とした議題は藤沢市の情報化に関する問題で、総文書数は424である。グループ分けに用いる閾値は、実験を繰り返し、偏ったグループが生成されないように最適な値を選択した。そして、自己組織化マップによる文書分類の実験では、各グループに適合率、再現率とF値を算出した。適合率とは、実際に分類された情報のうち、正しい情報の割合である。また、再現率とは、正しい情報のうち、実際に分類された情報の割合である。そして、F値とは、適合率と再現率の調和平均であり、分類精度を測るための指標として用いられる。また、議題抽出の実験では、分類された文書の内容が抽出された議題に沿っているかどうかを目視で確認することで評価を行った。

3.2 結果と考察

本システムの実験結果を表1に示す。また、実験による議題の抽出結果を表2に示す。本実験により生成されたグループの総数は78であり、そのうち、文書数の多い上位6グループを評価の対象とした。表1に示した各グループのF値を見ると、グループ1と2において0.6以上の値が出ている。F値の値は、教師データを用いない分類の場合、0.6以上で十分な精度であるため、高い精度が得られたと言える。他のグループで精度が悪かった理由として、文書数が少なかったことや、「オフラインミーティング」と「オフ会」などの表記ゆれが考えられる。また、少数意見も、類似グループに分類されていた。ストップワードを除いた各グループの議題には、パソコン、インターネット、会議や学校などといった単語が抽出された。目視による確認を行ったところ、グループ1では、情報化計画に関する議論がなされており、グループ2では、学校での情報教育に関する議論がなされていた。そのため、類似文書の分類を行い、議題を抽出する本手法の有効性を実証できた。

表1 実験結果

グループ	分類件数	適合率	再現率	F 値
グループ1	70	0.51428	0.83720	0.63716
グループ2	30	0.86666	0.63414	0.73239
グループ3	21	0.66666	0.43750	0.52830
グループ4	10	0.60000	0.37500	0.46153
グループ5	7	0.57142	0.40000	0.47058
グループ6	7	0.42857	0.42857	0.42857

表2 議題の抽出結果

グループ	議題1	議題2	議題3	議題4
グループ1	情報	計画	議論	意見
グループ2	学校	パソコン	教育	情報
グループ3	地域	市民	情報	発言
グループ4	会議	オフライン	ミーティング	情報
グループ5	インターネット	電話	接続	ネット
グループ6	議論	登録	作る	ホームページ

4. おわりに

本研究では、自己組織化マップを用いて電子会議の文書から議事録作成を支援する情報を生成するシステムを構築した。実証実験では、本システムを使用することにより、十分な分類精度が得られ、議事録作成を支援する情報の抽出が可能となった。このことから、本手法は、電子会議の議事録作成を支援する方法として有用であることが実証できた。今後の課題として、抽出した議題に対する肯定・否定の度合いを考慮した意見分類を行う必要や、議事録作成を支援する情報からの議事録の自動生成があると考えられる。

参考文献

- [1] 村越広亭, 島津明, 落水浩一郎: メーリングリストを利用した共同作業における討議構造の自動構築法, コンピュータソフトウェア, 日本ソフトウェア科学会, Vol.18, No.3, pp.19-23, 2001.9.
- [2] 古田一雄, 前原基芳, 高島亮祐, 中田圭一: 知的支援機能を備えた電子会議システム, 社会技術研究論文, 社会技術研究会, Vol.1, pp.299-306, 2003.10.
- [3] Shigeo Abe: Support Vector Machines for Pattern Classification, Springer, 2005.8.
- [4] 博報堂インタラクティブカンパニー: 図解でわかるインターネットマーケティング, 日本能率協会マネジメントセンター, 2000.2.
- [5] Teuvo Kohonen: The Self-Organizing Map, Proceedings of IEEE, IEEE, Vol.78, No.9, pp.1464-1480, 1990.9.