

2018年度 芝浦工業大学大学院

修 士 論 文

題目：混合多項分布と q -divergence に基づく

ファジィクラスタリング及び

ファジィクラスタリングの協調フィルタリングへの応用

専 攻 理工学研究科（修士課程）電気電子情報工学専攻

学籍番号 MA17049

ふりがな こんどう ただふみ

氏 名 近藤 唯史

指導教員 神澤 雄智

目次

| | | |
|--------------|--|----------|
| 第 1 章 | 序論 | 1 |
| 1.1 | 背景 | 1 |
| 1.1.1 | ファジィクラスタリング | 1 |
| 1.1.2 | q -divergence と球面クラスタリング | 2 |
| 1.1.3 | 混合多項分布 | 3 |
| 1.1.4 | 協調フィルタリング | 4 |
| 1.2 | 目的 | 4 |
| 1.3 | 構成 | 5 |
| 第 2 章 | ファジィクラスタリング | 6 |
| 2.1 | はじめに | 6 |
| 2.2 | ハードクラスタリング | 6 |
| 2.3 | k -means++ | 8 |
| 2.4 | ファジィクラスタリング | 8 |
| 2.4.1 | Bezdek によるファジィ c -平均法 | 9 |
| 2.4.2 | Entropy と Divergence | 11 |
| 2.4.3 | 混合正規分布とクラスタリング | 12 |
| 2.4.4 | KL-divergence 正則化ファジィ c -平均法 | 13 |
| 2.4.5 | q -divergence 正則化ファジィ c -平均法 | 15 |
| 2.5 | 球面データのためのファジィクラスタリング | 18 |
| 2.5.1 | 球面データのためのクラスタリング | 18 |
| 2.5.2 | BFCS | 18 |
| 2.5.3 | KLFCS | 20 |
| 2.5.4 | QFCS | 21 |
| 2.6 | 混合多項分布に基づくファジィクラスタリング | 23 |
| 2.6.1 | 混合多項分布に基づくクラスタリング | 23 |
| 2.6.2 | KLFCCM | 24 |
| 2.6.3 | QFCCM | 25 |
| 2.6.4 | BFCCM | 27 |
| 2.6.5 | 混合多項分布に基づくファジィクラスタリングの変形 | 29 |
| 2.6.6 | KLFCCMM | 29 |
| 2.6.7 | BFCCMM | 31 |
| 2.7 | 提案法 (QFCCMM) | 32 |

| | | |
|------------|---|------------|
| 2.8 | おわりに | 35 |
| 第3章 | ファジィクラスタリングにおける数値実験 | 36 |
| 3.1 | はじめに | 36 |
| 3.2 | 文書データ | 36 |
| 3.3 | tf.idf | 37 |
| 3.4 | 精度評価方法 | 38 |
| 3.4.1 | Contingency Table | 38 |
| 3.4.2 | Adjusted Rand Index (ARI) | 39 |
| 3.5 | 数値実験 1 | 39 |
| 3.5.1 | 数値実験 1 の結果 | 93 |
| 3.6 | 数値実験 2 | 93 |
| 3.6.1 | 数値実験 2 の結果 | 141 |
| 3.7 | 数値実験 3 | 141 |
| 3.7.1 | 数値実験 3 の結果 | 152 |
| 3.8 | 考察 | 152 |
| 3.9 | おわりに | 154 |
| 第4章 | 協調フィルタリング | 155 |
| 4.1 | はじめに | 155 |
| 4.2 | 従来の協調フィルタリング | 155 |
| 4.3 | 提案法 | 157 |
| 4.4 | 精度評価方法 | 157 |
| 4.4.1 | Mean Absolute Error (MAE) | 157 |
| 4.4.2 | F-measure | 157 |
| 4.4.3 | Receiver Operating Characteristic 曲線 (ROC 曲線) | 158 |
| 4.5 | おわりに | 158 |
| 第5章 | 協調フィルタリングにおける数値実験 | 160 |
| 5.1 | はじめに | 160 |
| 5.2 | 人工データ数値実験 | 160 |
| 5.2.1 | 設定 | 160 |
| 5.2.2 | 結果と考察 | 163 |
| 5.3 | 実データ数値実験 | 169 |
| 5.3.1 | 実データ | 169 |
| 5.3.2 | 設定 | 170 |
| 5.3.3 | 結果と考察 | 174 |
| 5.4 | まとめ | 176 |
| 5.5 | おわりに | 178 |
| 第6章 | 結論 | 181 |
| | 参考文献 | 183 |

| | |
|----------|-----|
| 代表的な研究業績 | 186 |
| 感想 | 187 |
| 謝辞 | 188 |
| 付録 | 189 |

図目次

| | | |
|------|---|-----|
| 1.1 | 機械学習の種類 | 1 |
| 1.2 | クラスタリング例 | 3 |
| 3.1 | 10 通りの初期値による最高平均 ARI と標準偏差 tf.idf なし | 138 |
| 3.2 | 10 通りの初期値による最高平均 ARI と標準偏差 tf.idfA | 139 |
| 3.3 | 10 通りの初期値による最高平均 ARI と標準偏差 tf.idfB | 140 |
| 3.4 | 100 通りの初期値による最高平均 ARI と標準偏差 tf.idf なし | 149 |
| 3.5 | 100 通りの初期値による最高平均 ARI と標準偏差 tf.idfA | 150 |
| 3.6 | 100 通りの初期値による最高平均 ARI と標準偏差 tf.idfB | 151 |
| 4.1 | ROC 曲線の例 | 159 |
| 5.1 | KLFCS($\lambda = 35$) における 10 通り欠損パターンの平均 MAE | 164 |
| 5.2 | BFCS($m = 1.5$) における 10 通り欠損パターンの平均 MAE | 164 |
| 5.3 | QFCS($\lambda = 35, m = 1.01$) における 10 通り欠損パターンの平均 MAE | 165 |
| 5.4 | KLFCCM($\lambda = 14$) における 10 通り欠損パターンの平均 MAE | 165 |
| 5.5 | BFCCM($m = 1.01$) における 10 通り欠損パターンの平均 MAE | 166 |
| 5.6 | QFCCM($\lambda = 100, m = 1.01$) における 10 通り欠損パターンの平均 MAE | 166 |
| 5.7 | KLFCCMM($\lambda = 10, t = 0.001$) における 10 通り欠損パターンの平均 MAE | 167 |
| 5.8 | BFCCMM($m = 1.01, t = 0.1$) における 10 通り欠損パターンの平均 MAE | 167 |
| 5.9 | QFCCMM($\lambda = 100, m = 1.01, t = 0.01$) における 10 通り欠損パターンの平均 MAE | 168 |
| 5.10 | 欠損数 5,500 におけるクラスタ数を変化させたときの 10 通りの欠損パターンの平均 MAE | 168 |
| 5.11 | libimseti の第 3 欠損パターンにおける QFCS($C = 20$),QFCCMM($C = 15$) を適用 した提案法の TP/FP | 178 |
| 5.12 | libimseti の第 0 欠損パターンにおける Q{FCS,FCCMM}($C = 20$) を適用した提案 法の TP/FP | 179 |
| 5.13 | libimseti の第 0 欠損パターンにおける Q{FCS,FCCMM}($C = 20$) の第 1 クラスタ に着目した提案法の TP/FP | 180 |
| 6.1 | クラス図 | 190 |

| | | |
|------|---|----|
| 3.1 | 数値実験に使用する文書データ | 38 |
| 3.2 | Contingency Table | 39 |
| 3.3 | 文書データ数値実験におけるファジィクラスタリングで採用したパラメータ範囲 | 40 |
| 3.4 | cacmcisi dataset の最高 ARI とパラメータ tf.idf なし | 41 |
| 3.5 | citeseer dataset の最高 ARI とパラメータ tf.idf なし | 41 |
| 3.6 | classic dataset の最高 ARI とパラメータ tf.idf なし | 42 |
| 3.7 | classic4 dataset の最高 ARI とパラメータ tf.idf なし | 42 |
| 3.8 | cora dataset の最高 ARI とパラメータ tf.idf なし | 43 |
| 3.9 | cranmed dataset の最高 ARI とパラメータ tf.idf なし | 43 |
| 3.10 | fbis dataset の最高 ARI とパラメータ tf.idf なし | 44 |
| 3.11 | hitech dataset の最高 ARI とパラメータ tf.idf なし | 44 |
| 3.12 | k1a dataset の最高 ARI とパラメータ tf.idf なし | 45 |
| 3.13 | k1b dataset の最高 ARI とパラメータ tf.idf なし | 45 |
| 3.14 | la1 dataset の最高 ARI とパラメータ tf.idf なし | 46 |
| 3.15 | la2 dataset の最高 ARI とパラメータ tf.idf なし | 47 |
| 3.16 | la12 dataset の最高 ARI とパラメータ tf.idf なし | 47 |
| 3.17 | mm dataset の最高 ARI とパラメータ tf.idf なし | 48 |
| 3.18 | new3 dataset の最高 ARI とパラメータ tf.idf なし | 48 |
| 3.19 | ohscal dataset の最高 ARI とパラメータ tf.idf なし | 49 |
| 3.20 | re0 dataset の最高 ARI とパラメータ tf.idf なし | 49 |
| 3.21 | re1 dataset の最高 ARI とパラメータ tf.idf なし | 50 |
| 3.22 | reviews dataset の最高 ARI とパラメータ tf.idf なし | 50 |
| 3.23 | sports dataset の最高 ARI とパラメータ tf.idf なし | 51 |
| 3.24 | tr11 dataset の最高 ARI とパラメータ tf.idf なし | 51 |
| 3.25 | tr12 dataset の最高 ARI とパラメータ tf.idf なし | 52 |
| 3.26 | tr23 dataset の最高 ARI とパラメータ tf.idf なし | 53 |
| 3.27 | tr31 dataset の最高 ARI とパラメータ tf.idf なし | 53 |
| 3.28 | tr41 dataset の最高 ARI とパラメータ tf.idf なし | 54 |
| 3.29 | tr45 dataset の最高 ARI とパラメータ tf.idf なし | 54 |
| 3.30 | wap dataset の最高 ARI とパラメータ tf.idf なし | 55 |
| 3.31 | webkb dataset の最高 ARI とパラメータ tf.idf なし | 55 |
| 3.32 | cacmcisi dataset の最高 ARI とパラメータ tf.idfA | 56 |
| 3.33 | citeseer dataset の最高 ARI とパラメータ tf.idfA | 56 |

| | | |
|------|---|----|
| 3.34 | classic dataset の最高 ARI とパラメータ tf.idfA | 57 |
| 3.35 | classic4 dataset の最高 ARI とパラメータ tf.idfA | 57 |
| 3.36 | cora dataset の最高 ARI とパラメータ tf.idfA | 58 |
| 3.37 | cranmed dataset の最高 ARI とパラメータ tf.idfA | 58 |
| 3.38 | fbis dataset の最高 ARI とパラメータ tf.idfA | 59 |
| 3.39 | hitech dataset の最高 ARI とパラメータ tf.idfA | 59 |
| 3.40 | k1a dataset の最高 ARI とパラメータ tf.idfA | 60 |
| 3.41 | k1b dataset の最高 ARI とパラメータ tf.idfA | 61 |
| 3.42 | la1 dataset の最高 ARI とパラメータ tf.idfA | 61 |
| 3.43 | la2 dataset の最高 ARI とパラメータ tf.idfA | 62 |
| 3.44 | la12 dataset の最高 ARI とパラメータ tf.idfA | 62 |
| 3.45 | mm dataset の最高 ARI とパラメータ tf.idfA | 63 |
| 3.46 | new3 dataset の最高 ARI とパラメータ tf.idfA | 63 |
| 3.47 | ohscal dataset の最高 ARI とパラメータ tf.idfA | 64 |
| 3.48 | re0 dataset の最高 ARI とパラメータ tf.idfA | 64 |
| 3.49 | re1 dataset の最高 ARI とパラメータ tf.idfA | 65 |
| 3.50 | reviews dataset の最高 ARI とパラメータ tf.idfA | 65 |
| 3.51 | sports dataset の最高 ARI とパラメータ tf.idfA | 66 |
| 3.52 | tr11 dataset の最高 ARI とパラメータ tf.idfA | 66 |
| 3.53 | tr12 dataset の最高 ARI とパラメータ tf.idfA | 67 |
| 3.54 | tr23 dataset の最高 ARI とパラメータ tf.idfA | 68 |
| 3.55 | tr31 dataset の最高 ARI とパラメータ tf.idfA | 68 |
| 3.56 | tr41 dataset の最高 ARI とパラメータ tf.idfA | 69 |
| 3.57 | tr45 dataset の最高 ARI とパラメータ tf.idfA | 70 |
| 3.58 | wap dataset の最高 ARI とパラメータ tf.idfA | 70 |
| 3.59 | webkb dataset の最高 ARI とパラメータ tf.idfA | 71 |
| 3.60 | cacmcisi dataset の最高 ARI とパラメータ tf.idfB | 71 |
| 3.61 | citeseer dataset の最高 ARI とパラメータ tf.idfB | 72 |
| 3.62 | classic dataset の最高 ARI とパラメータ tf.idfB | 72 |
| 3.63 | classic4 dataset の最高 ARI とパラメータ tf.idfB | 73 |
| 3.64 | cora dataset の最高 ARI とパラメータ tf.idfB | 73 |
| 3.65 | cranmed dataset の最高 ARI とパラメータ tf.idfB | 74 |
| 3.66 | fbis dataset の最高 ARI とパラメータ tf.idfB | 74 |
| 3.67 | hitech dataset の最高 ARI とパラメータ tf.idfB | 75 |
| 3.68 | k1a dataset の最高 ARI とパラメータ tf.idfB | 76 |
| 3.69 | k1b dataset の最高 ARI とパラメータ tf.idfB | 77 |
| 3.70 | la1 dataset の最高 ARI とパラメータ tf.idfB | 78 |
| 3.71 | la2 dataset の最高 ARI とパラメータ tf.idfB | 79 |
| 3.72 | la12 dataset の最高 ARI とパラメータ tf.idfB | 79 |
| 3.73 | mm dataset の最高 ARI とパラメータ tf.idfB | 80 |
| 3.74 | new3 dataset の最高 ARI とパラメータ tf.idfB | 80 |

| | | |
|-------|--|-----|
| 3.75 | ohscal dataset の最高 ARI とパラメータ tf.idfB | 81 |
| 3.76 | re0 dataset の最高 ARI とパラメータ tf.idfB | 81 |
| 3.77 | re1 dataset の最高 ARI とパラメータ tf.idfB | 82 |
| 3.78 | reviews dataset の最高 ARI とパラメータ tf.idfB | 83 |
| 3.79 | sports dataset の最高 ARI とパラメータ tf.idfB | 84 |
| 3.80 | tr11 dataset の最高 ARI とパラメータ tf.idfB | 84 |
| 3.81 | tr12 dataset の最高 ARI とパラメータ tf.idfB | 85 |
| 3.82 | tr23 dataset の最高 ARI とパラメータ tf.idfB | 85 |
| 3.83 | tr31 dataset の最高 ARI とパラメータ tf.idfB | 86 |
| 3.84 | tr41 dataset の最高 ARI とパラメータ tf.idfB | 87 |
| 3.85 | tr45 dataset の最高 ARI とパラメータ tf.idfB | 88 |
| 3.86 | wap dataset の最高 ARI とパラメータ tf.idfB | 89 |
| 3.87 | webkb dataset の最高 ARI とパラメータ tf.idfB | 89 |
| 3.88 | 初期値に正解のファジィな帰属度を与えた最高 ARI tf.idf なし | 90 |
| 3.89 | 初期値に正解のファジィな帰属度を与えた最高 ARI tf.idfA | 91 |
| 3.90 | 初期値に正解のファジィな帰属度を与えた最高 ARI tf.idfB | 92 |
| 3.91 | 最高 ARI を達成したデータ数 | 93 |
| 3.92 | cacmcisi dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 95 |
| 3.93 | citeseer dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 95 |
| 3.94 | classic dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 96 |
| 3.95 | classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 96 |
| 3.96 | cora dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 97 |
| 3.97 | cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 97 |
| 3.98 | fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 98 |
| 3.99 | hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 98 |
| 3.100 | k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 99 |
| 3.101 | k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 99 |
| 3.102 | la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 100 |
| 3.103 | la2 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 100 |
| 3.104 | la12 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 101 |
| 3.105 | mm dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 101 |
| 3.106 | ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 102 |
| 3.107 | re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 102 |
| 3.108 | re1 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 103 |
| 3.109 | reviews dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 103 |
| 3.110 | sports dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 104 |
| 3.111 | tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 104 |
| 3.112 | tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 105 |
| 3.113 | tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 105 |
| 3.114 | tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 106 |
| 3.115 | tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 106 |

| | | |
|-------|--|-----|
| 3.116 | tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 107 |
| 3.117 | wap dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 107 |
| 3.118 | webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし | 108 |
| 3.119 | cacmcisi dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 108 |
| 3.120 | citeseer dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 109 |
| 3.121 | classic dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 109 |
| 3.122 | classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 110 |
| 3.123 | cora dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 110 |
| 3.124 | cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 111 |
| 3.125 | fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 111 |
| 3.126 | hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 112 |
| 3.127 | k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 112 |
| 3.128 | k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 113 |
| 3.129 | la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 113 |
| 3.130 | la2 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 114 |
| 3.131 | la12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 114 |
| 3.132 | mm dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 115 |
| 3.133 | ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 115 |
| 3.134 | re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 116 |
| 3.135 | re1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 116 |
| 3.136 | reviews dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 117 |
| 3.137 | sports dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 117 |
| 3.138 | tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 118 |
| 3.139 | tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 118 |
| 3.140 | tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 119 |
| 3.141 | tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 119 |
| 3.142 | tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 120 |
| 3.143 | tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 120 |
| 3.144 | wap dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 121 |
| 3.145 | webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idfA | 121 |
| 3.146 | cacmcisi dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 121 |
| 3.147 | citeseer dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 122 |
| 3.148 | classic dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 122 |
| 3.149 | classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 123 |
| 3.150 | cora dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 123 |
| 3.151 | cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 124 |
| 3.152 | fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 124 |
| 3.153 | hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 125 |
| 3.154 | k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 125 |
| 3.155 | k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 126 |
| 3.156 | la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 126 |

| | | |
|-------|---|-----|
| 3.157 | la2 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 127 |
| 3.158 | la12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 127 |
| 3.159 | mm dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 128 |
| 3.160 | ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 128 |
| 3.161 | re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 129 |
| 3.162 | re1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 129 |
| 3.163 | reviews dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 130 |
| 3.164 | sports dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 130 |
| 3.165 | tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 131 |
| 3.166 | tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 131 |
| 3.167 | tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 132 |
| 3.168 | tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 132 |
| 3.169 | tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 133 |
| 3.170 | tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 133 |
| 3.171 | wap dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 134 |
| 3.172 | webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idfB | 134 |
| 3.173 | 10 通りの初期値による最高平均 ARI tf.idf なし | 135 |
| 3.174 | 10 通りの初期値による最高平均 ARI tf.idfA | 136 |
| 3.175 | 10 通りの初期値による最高平均 ARI tf.idfB | 137 |
| 3.176 | 10 通りの初期設定による平均 ARI の最高値を達成したデータ数 | 141 |
| 3.177 | 100 通りの初期値による最高平均 ARI tf.idf なし | 143 |
| 3.178 | 100 通りの初期値による最高平均 ARI tf.idfA | 144 |
| 3.179 | 100 通りの初期値による最高平均 ARI tf.idfB | 145 |
| 3.180 | 100 通りの初期値による目的関数値最小 ARI tf.idf なし | 146 |
| 3.181 | 100 通りの初期値による目的関数値最小 ARI tf.idfA | 147 |
| 3.182 | 100 通りの初期値による目的関数値最小 ARI tf.idfB | 148 |
| 3.183 | 100 通りの初期設定による平均 ARI の最高値を達成したデータ数 | 152 |
| 3.184 | 目的関数値最小の ARI の最高値を達成したデータ数 | 152 |
| 4.1 | スレットスコア | 158 |
| 5.1 | 人工データ | 161 |
| 5.2 | 式 (5.1) において $C = 5$ としたときの初期帰属度 $u_{i,k}^{\text{Init}}$ | 161 |
| 5.3 | 人工データを使用した数値実験で採用したパラメータ範囲 | 162 |
| 5.4 | $C = 5$ としたときの 10 通りの欠損パターンにおける平均 MAE | 169 |
| 5.5 | 実データ数値実験に使用する整形済みの評価データ | 170 |
| 5.6 | 実データ数値実験で採用したパラメータ範囲 $\{\text{KL}, \text{B}, \text{Q}\} \text{FC} \{\text{CM}, \text{CS}\}$ | 172 |
| 5.7 | 実データ数値実験で採用したパラメータ範囲 $\{\text{KL}, \text{B}, \text{Q}\} \text{FCCMM}$ | 173 |
| 5.8 | 5 通りの欠損パターンにおける最高平均 AUC を達成したパラメータ | 175 |
| 5.9 | 5 通りの欠損パターンにおける最高平均 MAE | 176 |
| 5.10 | 5 通りの欠損パターンにおける最高平均 F-measure | 176 |
| 5.11 | 5 通りの欠損パターンにおける最高平均 AUC | 177 |

| | | |
|------|---|-----|
| 5.12 | $C = 5$ としたときの 10 通りの欠損パターンにおける平均 ARI (人工データ) | 177 |
| 5.13 | libimseti の第 0 欠損パターンにおける $Q\{\text{FCS}, \text{FCCMM}\}(C = 20)$ の Contingency Table | 179 |
| 5.14 | libimseti の第 0 欠損パターンにおける $Q\{\text{FCS}, \text{FCCMM}\}(C = 20)$ を適用した AUC の差 | 180 |

第1章

序論

1.1 背景

情報通信技術の発達に伴い、日々多様なデータがコンピュータに蓄積されている。この大量のデータから有益な情報を抽出し活用する技術の注目度は大きい。例えば、ECサイトのようなインターネットサービスでは、高精度な推薦システムが期待される。消費者のレビューからは、商品の改善や開発の役立つ情報を得ることもできる。しかし、日々増え続ける膨大なデータを人手で分析するのは、コストがかかり非効率である。そこで本研究では、前半にデータ分析手法であるファジィクラスタリングについて、混合多項分布に基づく q -divergence 正則化ファジィクラスタリングを提案し、文書データを用いて他手法との比較を行う (代表的な研究業績 (1))。後半では、ファジィクラスタリングを協調フィルタリングに応用した推薦手法を提案し、その精度を評価する (代表的な研究業績 (2))。

1.1.1 ファジィクラスタリング

クラスタリングとはデータが与えられたとき、そのデータに属す個体間の類似性に基づいて、個体をいくつかのクラスと呼ばれるグループに分割する機械学習手法である。機械学習とは、データから有用な規則やパターンを抽出し、人間が自然に行っている学習をコンピュータで実現する技術であり、大きく教師あり学習と教師なし学習に分かれる (図 1.1)。教師あり学習のタスクの一つに「分類」があ

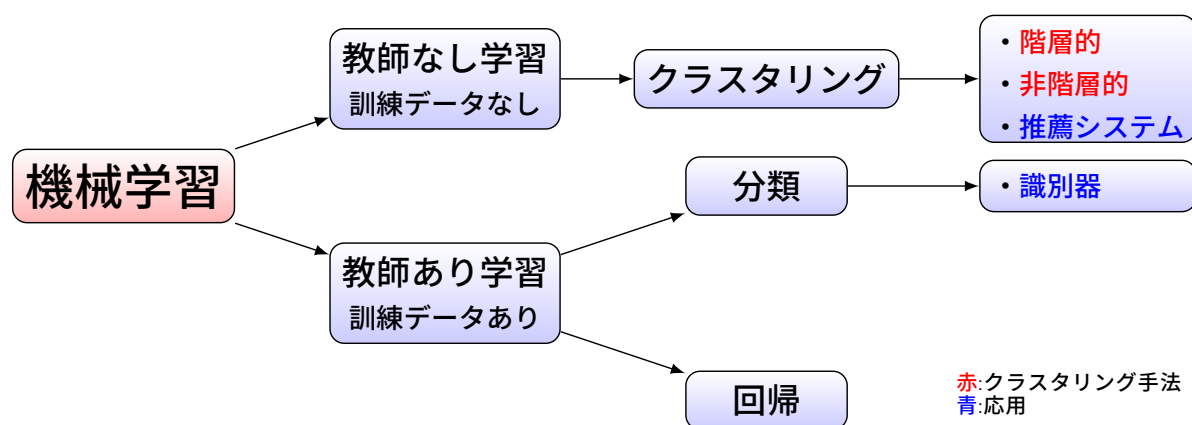


図 1.1: 機械学習の種類

る。予め定められたクラスに従ってデータを学習し、新たに観測した個体を学習したクラスの特徴と照らし合わせ自動的に最適なクラスに振り分ける手法である。ニュース記事をデータとして例えると、記

事が個体にあたり政治、経済、スポーツといったジャンルがクラスにあたる。ここで新たな記事がどのジャンルに属す記事なのか判定するシステムを識別器と呼ぶ。予め学習するデータは、訓練データや学習データとも呼ばれ分析者が事前に作成しておく必要があり、この訓練データの出来によって分類結果が大きく左右する。また「分類」の他に株価のような連続するデータの予測をする「回帰」がある。対して、訓練データを用いないクラスタリングは教師なし学習と呼ばれる。文献 [1] では 43,861 人のサッカーファンの Facebook ユーザをユーザの特徴に基づいてクラスタリングすることで四つのグループに分けている。各グループの特徴を分析することでユーザの嗜好に合う最適な広告をグループごとに提示できる。このような顧客セグメンテーションで活用される教師なし学習は一般的に、分析者が分割結果から各クラスに属す個体がどのような特徴を持つか考察する必要があるが、後述する推薦システムの協調フィルタリングに応用された場合はこの限りではない。

クラスタリングには、個体を逐次的に結合することでクラスを形成する階層的クラスタリングと尤もらしいクラスを形成する関数を定義し、その関数を最適化する非階層的クラスタリングがある。図 1.2 にニュース記事のクラスタリング例を示す。階層的クラスタリングは、図 1.2a のように階層的に個体を分ける特徴があり、分割結果の系統図から任意にクラス数を決定することができる。一方短所として、データの規模が大きくなるほど計算量が多くなることや系統図が巨大になり分析結果が不明瞭になることが挙げられる。そのため膨大なデータに対して、非階層的クラスタリングを採用するべきである。

一般的な非階層クラスタリングに図 1.2b に示すようなハードクラスタリングが知られているが、単純に個体を分けることしかできず、本質的にはっきり分割できない複雑なデータに対応していない。そこで必ずしも属すか属さないかで分けるのではなく、属す度合いに従って分割するファジィクラスタリングの期待が高まっている。このクラスターへ属す度合いを表す変数を帰属度と呼ぶ。図 1.2c は、帰属度を色合いで表現した。ファジィクラスタリングは、分析者が帰属度を調整するパラメータを設定する必要がある、最適なパラメータを推定するのに計算コストがかかるが、ハードクラスタリングで問題とされる初期値依存性を緩和できるという点で有利である。これまでにファジィクラスタリングは、多岐にわたるファジィ化の研究がされてきた。そこで本研究では、大規模データに適しかつ、発展性に富んだファジィクラスタリングに着目した。

1.1.2 q -divergence と球面クラスタリング

自然や社会のあらゆる分布が冪分布に従うことが分かっている。例えば、地震の大きさと発生回数や都市人口の規模と順位などがあたる。近年では人が選択する行動の履歴データから統計をとると、多く場合で冪分布を示すことが報告されており、データ分析手法の研究においても冪分布に基づくモデルが着目されている。ファジィクラスタリングでは、確率分布の尺度である KL-divergence を最適化問題の目的関数に導入した手法が開発されている。KL-divergence を q -対数関数によって一般化した q -divergence は、冪分布と関連があることが知られている [2, 3]。そこで q -divergence を導入したファジィクラスタリング手法に注目した。

一般的な非階層的クラスタリングは、平方ユークリッド距離によって各クラスターへの割り当てを決定する。一方、球面化されたデータを対象とした手法も考案されている。これは、クラスター中心が球面上に分布する制約のもとデータとクラスター中心の内積を非類似度化した手法であり、球面データのためのクラスタリングと呼ばれる。この手法は、高次元データや疎なデータに対して有効性が確認されている。 q -divergence を球面データのためのファジィクラスタリングへ適用した手法は、球面データに対して高

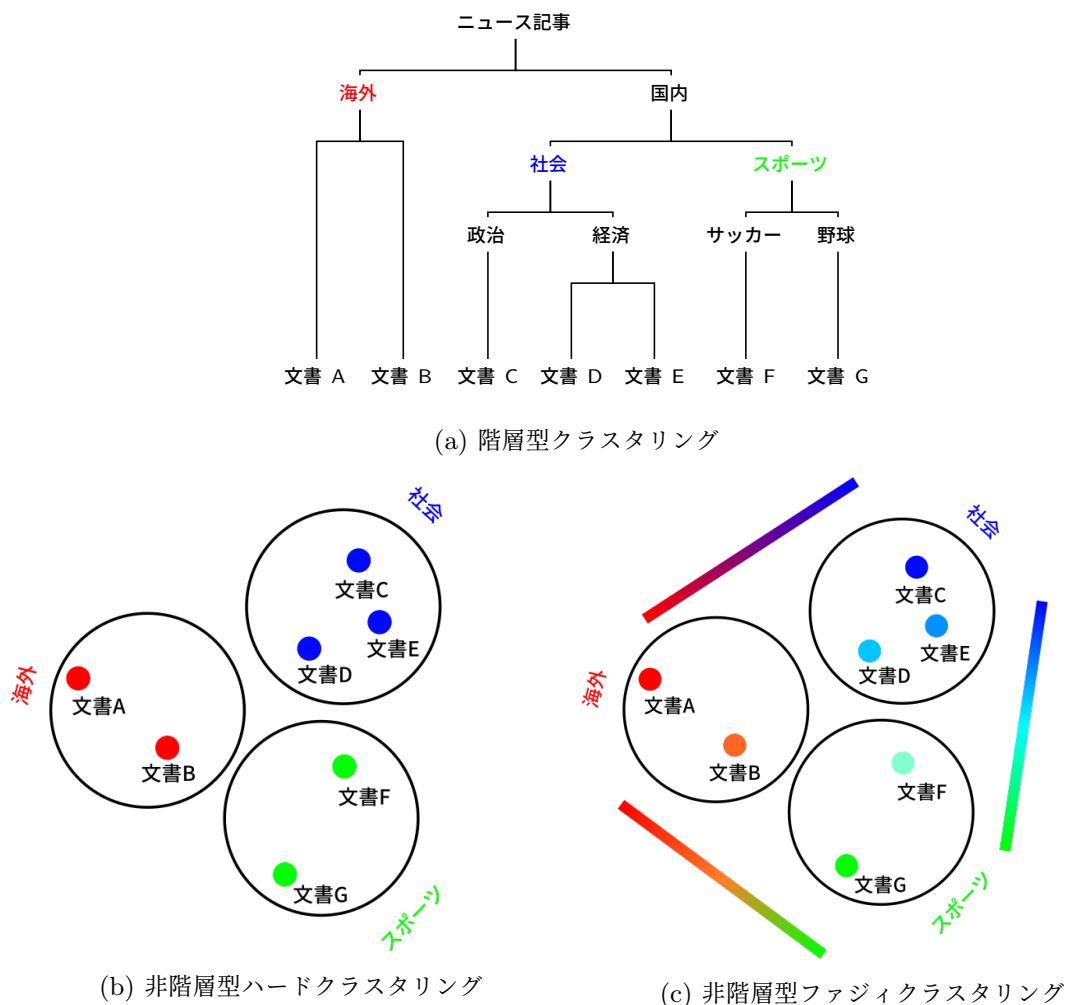


図 1.2: クラスタリング例

い精度であることが明らかにされた [4].

1.1.3 混合多項分布

多項分布とは二項分布に対して確率変数を確率ベクトルに一般化した分布である．混合多項分布に基づいて開発されたファジィクラスタリング手法は，分書と単語や顧客と購買履歴など異種で関係のあるデータに対して用いられ，文書分類や顧客セグメンテーションに応用される [5, 6]．混合多項分布を対数尤度最大化によって解くことで得られる目的関数をファジィ化した手法を混合多項分布に基づくファジィクラスタリングと呼ぶ [7]．また，得られた目的関数の非類似度にあたる因子の対数関数を新たなパラメータを導入し，一般化することで新たな最適化問題を導出できる [8]．非類似度をファジィ化したこの手法は，より柔軟な分割を行うことができると期待される．多項分布の確率がファジィクラスタリングにおけるクラスタ中心にあたり，総和が 1 となる制約からクラスタ中心が単体上に分布される．このことから混合多項分布に基づくファジィクラスタリングは，単体データのためのファジィクラスタリングとも呼ばれる．混合多項分布に基づくファジィクラスタリングの目的関数を q -divergence によって正規化することでより高精度な手法の開発を見込める．

1.1.4 協調フィルタリング

クラスタリングの応用先に推薦システムがある。推薦システムとは、ユーザの嗜好や行動履歴などの特徴に基づいてそのユーザが求める情報を提示するシステムである。あらゆる情報がデータベース化された情報過多な現代では、意思決定を促す推薦システムは重要視される。以下に推薦システムの応用例を示す：

- EC サイトで顧客が好む商品を提示することで販売促進できる。
- SNS でおすすめのユーザを提示することで利用促進できる。
- レシピ検索サイトで毎日の献立を提案することで主婦や料理人の負担を軽減できる。
- 時間帯に応じて営業先候補を優先付けすることで営業支援できる。

推薦システムは、主に協調フィルタリングと内容ベースフィルタリングに分かれる。協調フィルタリングは、商品やサービスといったアイテムの購買履歴や評価履歴に基づいてユーザを特徴ベクトルで表現する [9, 10]。そして、推薦システムの利用者に対して、似た特徴を持つその他のユーザが評価したアイテムでまだ利用者が未評価であるアイテムを推薦する手法である [11]。内容ベース型フィルタリングは、利用者がアイテムのジャンル、価格、製造元といった情報を検索し、より特徴が一致するアイテムを推薦する手法である。協調フィルタリングの長所として、他のユーザの特徴に依存するため利用者自身が未知である情報を推薦することができる。しかし、ユーザが少ない状況において協調フィルタリングは精度が落ち、新しいアイテムに対して推薦することができないという短所がある。また、アイテムの特徴が定まっているデータを扱う内容ベースフィルタリングと異なり、不定なユーザ情報を用いる協調フィルタリングは見当違いな推薦をすることもある。本来の推薦システムであれば両手法を使い分けたり使い合わせることが重要であるが本研究では、より発展性の高い協調フィルタリングのみを扱う。ファジィクラスタリングによってフィルタリング効果を向上させ、潜在的に似たユーザ情報しか扱わないことで高精度な推薦を期待できる。

1.2 目的

本論文では、混合多項分布に基づくファジィクラスタリングの高精度化および協調フィルタリングの高精度化を目的とする。

前半に混合多項分布から得られる目的関数の非類似度をファジィ化し、さらに q -divergence によって正則化するファジィクラスタリング手法を提案する。これを混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (q -divergence-regularized Fuzzy Clustering for Categorical Multivariate data induced by Multinomial mixture model: QFCCMM) と呼ぶ。Bezdek によるファジィ化、KL-divergence, q -divergence 正則化によるファジィ化三手法と非類似度が異なる三つのファジィクラスタリングを組み合わせ、提案した QFCCMM を含め計九つのファジィクラスタリング手法を数値実験に適用する。文書データを使用した数値実験では、二つの初期設定によってファジィクラスタリング手法を比較し、QFCCMM の精度を評価する。

後半では、ファジィクラスタリングを協調フィルタリングへ応用した推薦手法を提案する。一種類の評価データに見立てた人工データと六種類の実データを適用して、二つの従来法と九つのファジィクラスタリングを応用した提案法の精度比較を行う。

なお、クラスタリングについて、行だけでなく列も同時に分割する共クラスタリングや、ファジィク

ラスタリングにおける帰属度の制約をなくした可能性クラスタリングや、データを高次元に写像するカーネル法を応用したクラスタリングなど重要であるが本論文では扱わない。また、協調フィルタリングについて、Matrix Factorization やトピックモデルを応用した Collaborative Topic Regression が知られているが本論文では扱わない [12, 13].

1.3 構成

本文書の構成を次に示す。第 2 章では、ファジィクラスタリングについて説明する。提案法である QFCCMM を含めた九つの最適化問題やアルゴリズムを示す。第 3 章では、文書データを使用したファジィクラスタリングにおける数値実験について記述する。詳細な数値実験条件や実験結果を示し、考察する。第 4 章では、ユーザ間類似度として用いられるピアソン相関係数や、従来の協調フィルタリングである FireFly アルゴリズムと GroupLens アルゴリズム、提案法であるファジィクラスタリング手法と GroupLens を組み合わせた手法について説明する。第 5 章では、評価データを使用した協調フィルタリングにおける数値実験について記述する。第 6 章では、結論としてまとめる。

第2章

ファジィクラスタリング

2.1 はじめに

本章では、ファジィクラスタリングについて、関連研究を述べた上で、混合多項分布に基づく q -divergence 正則化ファジィクラスタリングを提案する。まず第 2.2 節でハードクラスタリングについて記述する。代表的な手法である HCM について二通りのアルゴリズムを示す。次に第 2.3 節で初期値クラスタ中心の与え方を改良した k -means++ のアルゴリズムについて記述する。次に第 2.4 節で Bezdek によるファジィ化について記述し、Entropy と Divergence について書き、混合正規分布について触れ、その後 KL-divergence と q -divergence 正則化によるファジィ化について述べる。次に第 2.5 節で球面データのためのファジィクラスタリング三手法について記述する。次に第 2.6 節で混合多項分布に基づくファジィクラスタリングについて記述する。混合多項分布を対数尤度最大化によって最適化問題を導出し、それをもとにして三つのファジィ化によるファジィクラスタリング手法を記述する。また、非類似度のファジィ化による混合多項分布に基づくファジィクラスタリングの変形について述べる。最後に第 2.7 節で混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (QFCCMM) を提案し、最適化問題とその導出、アルゴリズムを記述する。

2.2 ハードクラスタリング

最も有名な非階層的クラスタリングに k -平均法 (k -means) がある [14]。本文書では、後述するファジィクラスタリングと対比させるため k -means をハード c -平均法 (Hard c -means: HCM) と呼ぶ。HCM は、 N をオブジェクト数、 M を次元数としたとき、データ $X = \{x_k \in \mathbb{R}^M \mid k \in \{1, \dots, N\}\}$ におけるオブジェクト x_k を C 個のクラスタに分ける手法である。クラスタの重心であるクラスタ中心は、 $v = \{v_i \in \mathbb{R}^M \mid i \in \{1, \dots, C\}\}$ と表現する。 $\#k$ 番目のオブジェクトと $\#i$ 番目のクラスタ中心の非類似度 $d_{i,k}$ の総和を最小にする最適化問題を解き、得られたクラスタ中心に基づき各オブジェクトを最適なグループに割り当てることでクラスタが形成される。類似度に平方ユークリッド距離を適用し、オブジェクトがどのグループに属するかを表す集合 $G = \{G_1, \dots, G_i, \dots, G_C\}$ を導入すると、HCM の最適化問題は式 (2.1) のように記述できる：

$$\underset{G, v}{\text{minimize}} \sum_{i=1}^C \sum_{x_k \in G_i} \|x_k - v_i\|_2^2. \quad (2.1)$$

HCM は、次のアルゴリズムに従う：

アルゴリズム 1 (HCM).

STEP 1. 初期クラスタ中心 $\{v_i\}_{i=1}^C$ を設定する.

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$d_{i,k} = \|x_k - v_i\|_2^2. \quad (2.2)$$

STEP 3. オブジェクト集合 G_i を次のように設定する:

$$G_i = \{x_k \in X | d_{i,k} \leq d_{j,k}, j \in \{1, \dots, C\}\}. \quad (2.3)$$

STEP 4. クラスタ中心 $\{v_i\}_{i=1}^C$ を次のように設定する:

$$v_i = \frac{1}{|G_i|} \sum_{x_k \in G_i} x_k. \quad (2.4)$$

STEP 5. (G, v) が収束すれば終了, そうなければ STEP 2. へ ■

アルゴリズム 1 では, クラスタ中心 v_i を求める際, オブジェクト x_k が集合 G_i に属するかを確かめる必要がある. そこで, 単純な加算計算を行うだけで済むように集合 G の代わりに潜在変数である帰属度 $u = \{u_{i,k} | i \in \{1, \dots, C\}, k \in \{1, \dots, N\}\}$ を導入する. これは, 一対 K 符号化法 (1 of K 表現) として知られており, 以下の制約を満たす:

$$\sum_{i=1}^C u_{i,k} = 1, \quad (2.5)$$

$$u_{i,k} \in \{0, 1\}. \quad (2.6)$$

したがって帰属度を導入した HCM の最適化問題は, 式 (2.7) のように記述できる:

$$\underset{u, v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 \text{ subject to Eqs. (2.5), (2.6)}. \quad (2.7)$$

帰属度を導入した HCM は, 次のアルゴリズムに従う:

アルゴリズム 2 (帰属度を導入した HCM).

STEP 1. 初期クラスタ中心 $\{v_i\}_{i=1}^C$ を設定する.

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$d_{i,k} = \|x_k - v_i\|_2^2. \quad (2.8)$$

STEP 3. 帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$u_{i,k} = \begin{cases} 1 & (i = \arg \min_{1 \leq j \leq C} \{d_{j,k}\}), \\ 0 & (\text{それ以外のとき}). \end{cases} \quad (2.9)$$

STEP 4. クラスタ中心 $\{v_i\}_{i=1}^C$ を次のように設定する:

$$v_i = \frac{\sum_{k=1}^N u_{i,k} x_k}{\sum_{k=1}^N u_{i,k}}. \quad (2.10)$$

STEP 5. (u, v) が収束すれば終了, そうなければ STEP 2. へ. ■

本アルゴリズムでは初期設定としてクラスタ中心を与えているが, 初期帰属度を設定して, クラスタ中心, オブジェクト-クラスタ間非類似度, 帰属度の順に設定することもできる.

2.3 k -means++

一般的なクラスタリングアルゴリズムでは, まず初期値として帰属度もしくはクラスタ中心を与える必要がある. k -means の名付け親である MacQueen は, データからクラスタ数分のオブジェクトをランダムに選択し, それらを初期クラスタ中心としてクラスタリングを実行している [14]. その後, 新たに KKZ と呼ばれる初期クラスタ中心決定アルゴリズムが考案された [15]. KKZ では, 最初に選択するオブジェクトだけランダムに決定し, 二つ目以降はそれまで選ばれたオブジェクトから最長距離であるオブジェクトを初期クラスタ中心として与える. しかし, この初期クラスタ中心の設定方法にはデータに外れ値が存在するとき, 外れ値であるオブジェクトが必ずクラスタ中心として選択されてしまう問題がある. そこで考案されたのが k -means++ である [16]. k -means++ では, KKZ 同様まずランダムにオブジェクトの一つを選択し, それを第一初期クラスタ中心とする. その後, それまで選択された初期クラスタ中心から最長距離であるオブジェクトが選択されやすいよう確率的にオブジェクトを決定する. k -means++ による初期クラスタ中心の設定は, 全ての初期クラスタ中心をランダム選択する MacQueen の方法より初期値依存性は低い, KKZ より初期値依存性が高い. 一方 KKZ の外れ値問題を緩和しており, 初期値による試行回数を増やし, その中で目的関数が最小となるクラスタリング結果を採用することで最適な分割を決定することができる. 以下に k -means++ の初期クラスタ中心決定アルゴリズムを示す.

アルゴリズム 3 (k -means++).

STEP 1. クラスタ数 C を設定する.

STEP 2. データ x からランダムにオブジェクトインデックス k_1 を選び, そのオブジェクト x_{k_1} をクラスタ中心 v_1 に与える.

STEP 3. 非類似度 $d_{1,1}, \dots, d_{1,N}$ を計算し, 非類似度がより大きいオブジェクトインデックス k_2 を確率的に選び, オブジェクト x_{k_2} をクラスタ中心 v_2 に与える.

STEP 4. 既に決まったクラスタ中心のうち, オブジェクトと最も近傍のクラスタ中心の非類似度がより大きいオブジェクトインデックス k_i を確率的に選び, オブジェクト x_{k_i} をクラスタ中心 v_i に与える. STEP 4. を $i = C$ まで行う.

2.4 ファジィクラスタリング

ハードクラスタリングでは, データを構成する個々のオブジェクトがクラスタに属するか属さないか明確に判別する必要がある. しかし, 現実のデータは複雑であるため, オブジェクトすべてを明確に判別できるとは限らない. そこで, 曖昧さを表現するファジィ集合 [17] の考え方を取り入れたファジィクラスタリングに着目する. ファジィクラスタリングとは, 個々のオブジェクトがクラスタに属するか属さないかによってデータを分割するのではなく, どの程度属するかを考慮するクラスタリング手法である. 具体的には, 帰属度が 1 か 0 かにしかなかったのを式 (2.11) のように 0 から 1 の範囲で連続性をもた

せる必要がある：

$$u_{i,k} \in [0, 1]. \quad (2.11)$$

2.4.1 Bezdek によるファジィ c -平均法

HCM は、帰属度の制約を $\{0, 1\}$ から $[0, 1]$ に緩和してもなお、最適解は $\{0, 1\}$ のいずれかになる。Dunn (1974) は、帰属度の制約 $u_{i,k} \in \{0, 1\}$ を緩和するためファジィ分割の考え方を導入し、最適化問題 (2.12) のように拡張した [18]：

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^2 \|x_k - v_i\|_2^2 \text{ subject to Eqs. (2.5), (2.11).} \quad (2.12)$$

つづいて Bezdek (1981) は、最適化問題 (2.12) をファジィ c -平均法として一般化し、最適化問題 (2.13) のように定義した [19]：

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m \|x_k - v_i\|_2^2 \text{ subject to Eqs. (2.5), (2.11).} \quad (2.13)$$

ここで $m > 1$ は、帰属度のファジィさを調整するファジィ化パラメータである。文献によっては、この最適化問題を解くことによって得られる手法を標準的ファジィ c -平均法 (Standard Fuzzy c -Means: SFCM) や、単純にファジィ c -平均法 (Fuzzy c -Means: FCM) としているが、本文書では、他のファジィクラスタリング手法との混同を回避するため Bezdek によるファジィ c -平均法 (Bezdek-type Fuzzy c -Means: BFCM) と名付ける。

Miyamoto ら (2004) は、BFCM の最適化問題に係数 $\pi_i = \{\pi_i \in \mathbb{R} \mid i \in \{1, \dots, C\}\}$ を導入し、クラスタの大きさを考慮した手法を提案した [20]。この係数は、クラスタサイズ調整変数や混合比と呼ばれ、以下の制約を満たす：

$$\sum_{i=1}^C \pi_i = 1, \pi_i \geq 0. \quad (2.14)$$

クラスタサイズ調整変数つき BFCM の最適化問題を以下に示す：

$$\underset{u,v,\pi}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \|x_k - v_i\|_2^2 \text{ subject to Eqs. (2.5), (2.11), (2.14).} \quad (2.15)$$

最適化問題 (2.15) に対応する Lagrange 関数 $L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)$ は次のように表される：

$$\begin{aligned} L_{\text{BFCM}}(u, v, \pi, \gamma, \beta) = & \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \|x_k - v_i\|_2^2 \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right). \end{aligned} \quad (2.16)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$ は Lagrange 乗数である。最適性の必要条件は次のように

表される：

$$\frac{\partial L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.17)$$

$$\frac{\partial L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)}{\partial v_i} = 0, \quad (2.18)$$

$$\frac{\partial L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.19)$$

$$\frac{\partial L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.20)$$

$$\frac{\partial L_{\text{BFCM}}(u, v, \pi, \gamma, \beta)}{\partial \beta_k} = 0. \quad (2.21)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.17) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1}(\pi_i)^{1-m}d_{i,k} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \left(\frac{\gamma_k}{m d_{i,k}} \right)^{1/(m-1)}. \end{aligned} \quad (2.22)$$

ここで、 $d_{i,k} = \|x_k - v_i\|_2^2$ である．条件式 (2.20) を考慮して式 (2.22) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{1/(1-m)}}. \quad (2.23)$$

．最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.18) によって次のように表される：

$$\begin{aligned} 2 \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{-m} (x_k - v_i) &= 0 \\ \Leftrightarrow v_i &= \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\sum_{k=1}^N (u_{i,k})^m}. \end{aligned} \quad (2.24)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.19) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} \sum_{k=1}^N (1-m)(u_{i,k})^m (\pi_i)^{-m} d_{i,k} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.25)$$

条件式 (2.21) を考慮して式 (2.25) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m d_{j,k}}{(u_{i,k})^m d_{i,k}} \right)^{1/m}}. \quad (2.26)$$

以上より、クラスタサイズ調整変数つき BFCM は次のアルゴリズムにまとめられる：

アルゴリズム 4 (クラスタサイズ調整変数つき BFCM).

STEP 1. ファジィ化パラメータ $m > 1$, 初期クラスタ中心 $\{v_i\}_{i=1}^C$ と初期クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を設定する.

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$d_{i,k} = \|x_k - v_i\|_2^2. \quad (2.27)$$

STEP 3. 帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{1/(1-m)}}. \quad (2.28)$$

STEP 4. クラスタ中心 $\{v_i\}_{i=1}^C$ を次のように設定する:

$$v_i = \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\sum_{k=1}^N (u_{i,k})^m}. \quad (2.29)$$

STEP 5. クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を次のように設定する:

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m d_{j,k}}{(u_{i,k})^m d_{i,k}} \right)^{1/m}}. \quad (2.30)$$

STEP 6. (u, v, π) が収束すれば終了, そうなければ STEP 2. へ. ■

2.4.2 Entropy と Divergence

後述するファジィクラスタリング手法で導入される Entropy と Divergence について紹介する. 確率分布 P が与えられ, 事象 $A = \{a_i \in \mathbb{R} | i \in \{1, \dots, n\}\}$ が互いに排反で, その生起確率 $P(a_i)$ の総和を 1 とするとき, A の情報量の期待値 $H_{\text{Shannon}}(P)$ を Shanonn Entropy と呼ぶ:

$$\sum_{i=1}^n P(a_i) = 1, \quad (2.31)$$

$$H_{\text{Shannon}}(P) = - \sum_{i=1}^n P(a_i) \log(P(a_i)). \quad (2.32)$$

二つの確率分布 P と Q が与えられたとき, P の Q に対する Kullback-Leibler-divergence (KL-divergence) を以下に示す:

$$D_{\text{KL}}(P||Q) = \sum_{i=1}^n P(a_i) \log \left(\frac{P(a_i)}{Q(a_i)} \right). \quad (2.33)$$

Shannon Entropy と KL-divergence を式 (2.34) の q -対数関数によって一般化すると、それぞれ Tsallis Entropy $H_{\text{Tsallis}}(P)$ と q -divergence $D_q(P||Q)$ を導くことができる [2, 3] :

$$\log_q(x) = \frac{1}{1-q}(x^{1-q} - 1) \quad (\text{for } x > 0), \quad (2.34)$$

$$\begin{aligned} H_{\text{Tsallis}}(P) &= - \sum_{i=1}^n P(a_i) \log_q(P(a_i)), \\ &= - \sum_{i=1}^n P(a_i) \left(\frac{1}{1-q}(P(a_i)^{1-q} - 1) \right), \\ &= - \frac{1}{1-q} \left(\sum_{i=1}^n P(a_i)^{2-q} - \sum_{i=1}^n P(a_i) \right), \\ &= - \frac{1}{q'-1} \left(\sum_{i=1}^n (a_i)^{q'} - 1 \right), \end{aligned} \quad (2.35)$$

ここで, $q = 2 - q'$ とおいた.

$$\begin{aligned} D_q(P||Q) &= \sum_{i=1}^n P(a_i) \log_q \left(\frac{P(a_i)}{Q(a_i)} \right), \\ &= \sum_{i=1}^n P(a_i) \left(\frac{1}{1-q}(P(a_i)^{1-q} Q(a_i)^{q-1} - 1) \right), \\ &= \frac{1}{1-q} \left(\sum_{i=1}^n P(a_i)^{2-q} Q(a_i)^{q-1} - \sum_{i=1}^n P(a_i) \right), \\ &= \frac{1}{q'-1} \left(\sum_{i=1}^n P(a_i)^{q'} Q(a_i)^{1-q'} - 1 \right), \end{aligned} \quad (2.36)$$

ここで, $q = 2 - q'$ とおいた. $q' \rightarrow 1$ とすることで, Tsallis Entropy は Shannon Entropy に, q -divergence は KL-divergence になり, Tsallis Entropy, q -divergence は Shannon Entropy, q -divergence を包括している.

2.4.3 混合正規分布とクラスタリング

HCM と関連のある混合正規分布 (Gaussian Mixture Model: GMM) の最適化問題を先に導出する. C 個の分布が与えられ, 各分布がそれぞれ M 次元正規分布に従うものとしたとき, 混合された分布は, 正規分布の重ね合わせで記述できる. $\mu = \{\mu_i \in \mathbb{R}^M \mid i \in \{1, \dots, C\}\}$ は平均, $\Sigma = \{\Sigma_i \in \mathbb{R}^{M \times M} \mid i \in \{1, \dots, C\}\}$ は分散共分散行列を表す. また, $\pi_i = \{\pi_i \in \mathbb{R} \mid i \in \{1, \dots, C\}\}$ は分布の相対的な大きさであり, クラスタサイズ調整変数の制約式 (2.14) を満たす. 以下に混合正規分布を示す:

$$P_{\text{GMM}}(x_k; \mu, \Sigma, \pi) = \sum_{i=1}^C \pi_i \frac{1}{\sqrt{(2\pi)^M |\Sigma_i|}} \exp \left(-\frac{1}{2} (x_k - \mu_i)^T \Sigma_i^{-1} (x_k - \mu_i) \right), \quad (2.37)$$

ここで, 分散共分散行列が単位行列である特殊な場合を考える:

$$P_{\text{GMM}}(x_k; \mu, \pi) = \sum_{i=1}^C \pi_i \frac{1}{\sqrt{(2\pi)^M}} \exp \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right). \quad (2.38)$$

次に, $\{x_k\}_{k=1}^N$ が独立のもと対数尤度が最大となる (μ, π) を求める. 以下に対数尤度を示す:

$$\begin{aligned} \log \left(\prod_{k=1}^N P_{\text{GMM}}(x_k; \mu, \pi) \right) &= \log \left(\prod_{k=1}^N \sum_{i=1}^C \pi_i \frac{1}{\sqrt{(2\pi)^M}} \exp \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right) \right), \\ &= \sum_{k=1}^N \log \left(\sum_{i=1}^C \pi_i \frac{1}{\sqrt{(2\pi)^M}} \exp \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right) \right). \end{aligned} \quad (2.39)$$

つづいて, 以下のように帰属度を導入し, イェンセンの不等式より (u, μ, π) を求める最適化問題を導出する:

$$\begin{aligned} &= \sum_{k=1}^N \log \left(\sum_{i=1}^C u_{i,k} \frac{\pi_i}{u_{i,k}} \frac{1}{\sqrt{(2\pi)^M}} \exp \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right) \right), \\ &\geq \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \frac{1}{\sqrt{(2\pi)^M}} \exp \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right) \right), \\ &= \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \right) + \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \frac{1}{\sqrt{(2\pi)^M}} + \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \left(-\frac{1}{2} \|x_k - \mu_i\|_2^2 \right), \\ &= \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \right) + N \log \frac{1}{\sqrt{(2\pi)^M}} - \frac{1}{2} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - \mu_i\|_2^2. \end{aligned} \quad (2.40)$$

したがって, 最適化問題は以下のように求まる:

$$\begin{aligned} &\underset{u, \mu, \pi}{\text{maximize}} \quad -\frac{1}{2} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - \mu_i\|_2^2 + \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \right) \\ &\Leftrightarrow \underset{u, \mu, \pi}{\text{minimize}} \quad \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - \mu_i\|_2^2 + 2 \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right). \end{aligned} \quad (2.41)$$

これは, HCM の最適化問題 (2.7) に帰属度とクラスタサイズ調整変数に関する KL-divergence による正則化項を加えたソフトクラスタリング手法である. この正則化の度合いを調整することで, 確率論におけるソフトクラスタリングからファジィ理論におけるファジィクラスタリングに拡張されたと考える.

2.4.4 KL-divergence 正則化ファジィ c -平均法

Miyamoto ら (1997) は HCM を特異な状態であるとして正則化することで帰属度の制約 $u_{i,k} \in \{0, 1\}$ を緩和した [21]. 正則化項を帰属度に関する負の Shannon Entropy とすると以下の最適化問題が得られる:

$$\begin{aligned} &\underset{u, v}{\text{minimize}} \quad \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}) \\ &\quad \text{subject to Eqs. (2.5), (2.11)}. \end{aligned} \quad (2.42)$$

これは, Entropy 正則化ファジィ c -平均法 (Entropy-regularized Fuzzy c -Means: EFCM) と呼ばれる. ここで, $\lambda > 0$ は正則化の度合いを調整するためのファジィ化パラメータである. また, クラスタサイ

ズ調整変数を導入することで次の最適化問題を得られる [20] :

$$\begin{aligned} & \underset{u, v, \pi}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ & \text{subject to Eqs. (2.5), (2.11), (2.14).} \end{aligned} \quad (2.43)$$

これは, HCM を帰属度 u とクラスタサイズ調整変数 π の KL-divergence 項により正則化されたと解釈され, ファジィ化パラメータ $\lambda = \frac{1}{2}$ とすることで 2.4.3 小節で述べた特異な混合正規分布と同義になる. この最適化問題を解くことによって得られる手法を KL-divergence 正則化ファジィ c -平均法 (KL-divergence-regularized Fuzzy c -Means: KLFCM) と呼ぶ. 最適化問題 (2.43) に対応する Lagrange 関数 $L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)$ は次のように表される :

$$\begin{aligned} L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta) = & \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \|x_k - v_i\|_2^2 + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right). \end{aligned} \quad (2.44)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$ は Lagrange 乗数である. 最適性の必要条件は次のように表される :

$$\frac{\partial L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.45)$$

$$\frac{\partial L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)}{\partial v_i} = 0, \quad (2.46)$$

$$\frac{\partial L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.47)$$

$$\frac{\partial L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.48)$$

$$\frac{\partial L_{\text{KLFCM}}(u, v, \pi, \gamma, \beta)}{\partial \beta_k} = 0. \quad (2.49)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.45) によって Lagrange 乗数 γ_k を用いて次のように表される :

$$\begin{aligned} d_{i,k} + \lambda^{-1} \left(\log \left(\frac{u_{i,k}}{\pi_i} \right) + 1 \right) - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \exp(\gamma_k - 1 - \lambda d_{i,k}). \end{aligned} \quad (2.50)$$

ここで, $d_{i,k} = \|x_k - v_i\|_2^2$ である. 条件式 (2.48) を考慮して式 (2.50) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る :

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \exp(-\lambda(d_{j,k} - d_{i,k}))}. \quad (2.51)$$

最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.46) によって次のように表される :

$$\begin{aligned} 2 \sum_{k=1}^N u_{i,k} (x_k - v_i) &= 0 \\ \Leftrightarrow v_i &= \frac{\sum_{k=1}^N u_{i,k} x_k}{\sum_{k=1}^N u_{i,k}}. \end{aligned} \quad (2.52)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.47) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} -\lambda^{-1} \sum_{k=1}^N \frac{u_{i,k}}{\pi_i} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= -\frac{\lambda^{-1} \sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \beta_k}. \end{aligned} \quad (2.53)$$

条件式 (2.49) を考慮して式 (2.53) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\begin{aligned} \pi_i &= \frac{\sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \sum_{j=1}^C u_{j,k}}, \\ &= \frac{\sum_{k=1}^N u_{i,k}}{N}. \end{aligned} \quad (2.54)$$

以上より、KLFCM は次のアルゴリズムにまとめられる：

アルゴリズム 5 (KLFCM).

STEP 1. ファジィ化パラメータ $\lambda > 0$, 初期クラスタ中心 $\{v_i\}_{i=1}^C$ と初期クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を設定する.

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように計算する：

$$d_{i,k} = \|x_k - v_i\|_2^2. \quad (2.55)$$

STEP 3. 帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \exp(-\lambda(d_{j,k} - d_{i,k}))}. \quad (2.56)$$

STEP 4. クラスタ中心 $\{v_i\}_{i=1}^C$ を次のように設定する：

$$v_i = \frac{\sum_{k=1}^N u_{i,k} x_k}{\sum_{k=1}^N u_{i,k}}. \quad (2.57)$$

STEP 5. クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を次のように設定する：

$$\pi_i = \frac{\sum_{k=1}^N u_{i,k}}{N}. \quad (2.58)$$

STEP 6. (u, v, π) が収束すれば終了, そうなければ STEP 2. へ. ■

2.4.5 q -divergence 正則化ファジィ c -平均法

Tsallis Entropy を用いたファジィクラスタリングは、Ménard (2003) らによって提案され、Yasuda (2012) によって自由エネルギーと関連付けられて深く研究されている [22, 23]. 以下に Tsallis

Entropy 正則化ファジィ c -平均法 (Tsallis entropy-regularized Fuzzy c -Means: TFCM) の最適化問題を示す：

$$\begin{aligned} & \underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m \|x_k - v_i\|_2^2 + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m \\ & \text{subject to Eqs. (2.5), (2.11).} \end{aligned} \quad (2.59)$$

ここで, $\lambda > 0$, $m > 1$ は正則化の度合いを調整するためのファジィ化パラメータである. また, クラスサイズ調整変数を導入することで次の最適化問題を得られる [24]：

$$\begin{aligned} & \underset{u,v,\pi}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \|x_k - v_i\|_2^2 + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m}. \\ & \text{subject to Eqs. (2.5), (2.11), (2.14)).} \end{aligned} \quad (2.60)$$

これは, HCM を帰属度 u とクラスサイズ調整変数 π の q -divergence で正則化したと解釈され, ファジィ化パラメータ $\lambda \rightarrow 0$ とすることで BFCM に, ファジィ化パラメータ $m \rightarrow 1$ とすることで KLFCM と同義になる. この最適化問題を解くことによって得られる手法を q -divergence 正則化ファジィ c -平均法 (q -divergence-regularized Fuzzy c -Means: QFCM) と呼ぶ. 最適化問題 (2.60) に対応する Lagrange 関数 $L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)$ は次のように表される：

$$\begin{aligned} L_{\text{QFCM}}(u, v, \pi, \gamma, \beta) = & \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \|x_k - v_i\|_2^2 + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{i=1}^C \beta_i \left(1 - \sum_{k=1}^N \pi_i \right). \end{aligned} \quad (2.61)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$ は Lagrange 乗数である. 最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.62)$$

$$\frac{\partial L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)}{\partial v_i} = 0, \quad (2.63)$$

$$\frac{\partial L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.64)$$

$$\frac{\partial L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.65)$$

$$\frac{\partial L_{\text{QFCM}}(u, v, \pi, \gamma, \beta)}{\partial \beta_k} = 0. \quad (2.66)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.62) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} & m(u_{i,k})^{m-1} (\pi_i)^{1-m} d_{i,k} + \frac{\lambda^{-1}m}{m-1} (u_{i,k})^{m-1} (\pi_i)^{1-m} - \gamma_k = 0 \\ \Leftrightarrow u_{i,k} = & \left(\frac{\gamma_k}{m(\pi_i)^{1-m} d_{i,k} + m/\lambda(m-1)(\pi_i)^{1-m}} \right)^{1/(m-1)}. \end{aligned} \quad (2.67)$$

ここで, $d_{i,k} = \|x_k - v_i\|_2^2$ である. 条件式 (2.65) を考慮して式 (2.67) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る:

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.68)$$

最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.63) によって次のように表される:

$$\begin{aligned} 2 \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (x_k - v_i) &= 0 \\ \Leftrightarrow v_i &= \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\sum_{k=1}^N (u_{i,k})^m}. \end{aligned} \quad (2.69)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.64) によって Lagrange 乗数 β_k を用いて次のように表される:

$$\begin{aligned} \sum_{k=1}^N (1-m)(\pi_i)^{-m} (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{-m} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N u_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.70)$$

条件式 (2.66) を考慮して式 (2.70) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る:

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.71)$$

以上より, QFCM は次のアルゴリズムにまとめられる:

アルゴリズム 6 (QFCM).

STEP 1. ファジィ化パラメータ $\lambda > 0, m > 1$, 初期クラスタ中心 $\{v_i\}_{i=1}^C$ と初期クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を設定する.

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように計算する:

$$d_{i,k} = \|x_k - v_i\|_2^2. \quad (2.72)$$

STEP 3. 帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する:

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.73)$$

STEP 4. クラスタ中心 $\{v_i\}_{i=1}^C$ を次のように設定する:

$$v_i = \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\sum_{k=1}^N (u_{i,k})^m}. \quad (2.74)$$

STEP 5. クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を次のように設定する：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.75)$$

STEP 6. (u, v, π) が収束すれば終了， そうなければ STEP 2. へ. ■

2.5 球面データのためのファジィクラスタリング

2.5.1 球面データのためのクラスタリング

球面化されたデータを対象としたクラスタリング手法は，高次元データや疎なデータに対して高精度なクラスタリングを行うことができる [25]. クラスタ中心が球面上に分布する制約のもとデータとクラスタ中心の内積を非類似度化した手法であり，球面クラスタリングや球面データのためのクラスタリングと呼ぶ (Hard Clustering for Spherical Data: HCS). HCS は，球面上に分布するようにデータを整形する必要があり以下の制約を満たす：

$$\|x_k\|_2 = 1, \quad (2.76)$$

$$\|v_i\|_2 = 1. \quad (2.77)$$

また，HCM の非類似度にあたる平方ユークリッド距離の代わりにコサイン類似度を用いることでデータクラスタ間非類似度を表現する：

$$d_{i,k} = 1 - x_k^\top v_i. \quad (2.78)$$

この節では，HCS をファジィクラスタリングへ拡張した三手法を紹介する。

2.5.2 Bezdek-type 球面ファジィクラスタリング

BFCM を球面ファジィクラスタリングへ変更した Bezdek-type 球面ファジィクラスタリング (Bezdek-type Fuzzy Clustering for Spherical Data: BFCS) の最適化問題を以下に示す [26]：

$$\begin{aligned} & \underset{u, v, \pi}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (1 - x_k^\top v_i) \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.77).} \end{aligned} \quad (2.79)$$

最適化問題 (2.79) に対応する Lagrange 関数 $L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma) &= \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (1 - x_k^\top v_i) \\ &+ \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i (1 - \|v_i\|_2). \end{aligned} \quad (2.80)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial u_{i,k}} = 0, \quad (2.81)$$

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial v_i} = 0, \quad (2.82)$$

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \pi_i} = 0, \quad (2.83)$$

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \gamma_k} = 0, \quad (2.84)$$

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \beta_k} = 0, \quad (2.85)$$

$$\frac{\partial L_{\text{BFCS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.86)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.81) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1}(\pi_i)^{1-m}d_{i,k} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \left(\frac{\gamma_k}{m d_{i,k}} \right)^{1/(m-1)}. \end{aligned} \quad (2.87)$$

ここで, $d_{i,k} = 1 - x_k^\top v_i$ である．条件式 (2.84) を考慮して式 (2.22) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{1/(1-m)}}. \quad (2.88)$$

最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.82) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} - \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} x_k - \sigma_i \frac{v_i}{\|v_i\|_2} &= 0 \\ \Leftrightarrow v_i &= \frac{-\|v_i\|_2 (\pi_i)^{1-m}}{\sigma_i} \sum_{k=1}^N (u_{i,k})^m x_k. \end{aligned} \quad (2.89)$$

条件式 (2.86) を考慮をするとクラスタ中心の更新式を次のように得る：

$$v_i = \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\|\sum_{k=1}^N (u_{i,k})^m x_k\|_2}. \quad (2.90)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.83) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} \sum_{k=1}^N (1-m)(u_{i,k})^m (\pi_i)^{-m} d_{i,k} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.91)$$

条件式 (2.85) を考慮して式 (2.91) をクラスティンデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m d_{j,k}}{(u_{i,k})^m d_{i,k}} \right)^{1/m}}. \quad (2.92)$$

2.5.3 KL-divergence 正則化に基づく球面ファジィクラスタリング

KLFCM を球面クラスタリングへ変更した KL-divergence 正則化に基づく球面ファジィクラスタリング (KL-divergence-regularized Fuzzy Clustering for Spherical Data: KLFCFS) の最適化問題を以下に示す [26]：

$$\begin{aligned} \underset{u, v, \pi}{\text{minimize}} \quad & \sum_{i=1}^C \sum_{k=1}^N u_{i,k} (1 - x_k^\top v_i) + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ \text{subject to Eqs. } & (2.5), (2.11), (2.14), (2.77). \end{aligned} \quad (2.93)$$

最適化問題 (2.93) に対応する Lagrange 関数 $L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta, \sigma) = & \sum_{i=1}^C \sum_{k=1}^N u_{i,k} (1 - x_k^\top v_i) + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i (1 - \|v_i\|_2). \end{aligned} \quad (2.94)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.95)$$

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta)}{\partial v_i} = 0, \quad (2.96)$$

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.97)$$

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.98)$$

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.99)$$

$$\frac{\partial L_{\text{KLFCFS}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.100)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.95) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} d_{i,k} + \lambda^{-1} \left(\log \left(\frac{u_{i,k}}{\pi_i} \right) + 1 \right) - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \exp(\gamma_k - 1 - \lambda d_{i,k}). \end{aligned} \quad (2.101)$$

ここで, $d_{i,k} = 1 - x_k^\top v_i$ である．条件式 (2.98) を考慮して式 (2.101) をクラスティンデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \exp(-\lambda(d_{j,k} - d_{i,k}))}. \quad (2.102)$$

最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.96) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} -\sum_{k=1}^N u_{i,k} x_k - \sigma_i \frac{v_i}{\|v_i\|_2} &= 0 \\ \Leftrightarrow v_i &= \frac{-\|v_i\|_2}{\sigma_i} \sum_{k=1}^N u_{i,k} x_k. \end{aligned} \quad (2.103)$$

条件式 (2.100) を考慮をするとクラスタ中心の更新式を次のように得る：

$$v_i = \frac{\sum_{k=1}^N u_{i,k} x_k}{\left\| \sum_{k=1}^N u_{i,k} x_k \right\|_2}. \quad (2.104)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.97) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} -\lambda^{-1} \sum_{k=1}^N \frac{u_{i,k}}{\pi_i} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= -\frac{\lambda^{-1} \sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \beta_k}. \end{aligned} \quad (2.105)$$

条件式 (2.99) を考慮して式 (2.105) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\begin{aligned} \pi_i &= \frac{\sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \sum_{j=1}^C u_{j,k}}, \\ &= \frac{\sum_{k=1}^N u_{i,k}}{N}. \end{aligned} \quad (2.106)$$

2.5.4 q -divergence 正則化に基づく球面ファジィクラスタリング

QFCM を球面クラスタリングへ変更した q -divergence 正則化に基づく球面クラスタリング (q -divergence-regularized Fuzzy Clustering for Spherical Data: QFCS) の最適化問題を以下に示す [4]：

$$\begin{aligned} \underset{u,v,\pi}{\text{minimize}} \quad & \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (1 - x_k^\top v_i) + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \\ \text{subject to Eqs.} \quad & (2.5), (2.11), (2.14), (2.77). \end{aligned} \quad (2.107)$$

最適化問題 (2.107) に対応する Lagrange 関数 $L_{\text{QFCS}}(u, v, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{QFCS}}(u, v, \pi, \gamma, \beta, \sigma) &= \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (1 - x_k^\top v_i) + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \\ &+ \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i (1 - \|v_i\|_2). \end{aligned} \quad (2.108)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{QFCS}}(u, v, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.109)$$

$$\frac{\partial L_{\text{QFCS}}(u, v, \pi, \gamma, \beta)}{\partial v_i} = 0, \quad (2.110)$$

$$\frac{\partial L_{\text{QFCS}}(u, v, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.111)$$

$$\frac{\partial L_{\text{QFCS}}(u, v, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.112)$$

$$\frac{\partial L_{\text{QFCS}}(u, v, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.113)$$

$$\frac{\partial L_{\text{KLFC}}(u, v, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.114)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.109) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1}(\pi_i)^{1-m}d_{i,k} + \frac{\lambda^{-1}m}{m-1}(u_{i,k})^{m-1}(\pi_i)^{1-m} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \left(\frac{\gamma_k}{m(\pi_i)^{1-m}d_{i,k} + m/\lambda(m-1)(\pi_i)^{1-m}} \right)^{1/(m-1)}. \end{aligned} \quad (2.115)$$

ここで, $d_{i,k} = 1 - x_k^\top v_i$ である．条件式 (2.112) を考慮して式 (2.115) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.116)$$

最適なクラスタ中心 $\{v_i\}_{i=1}^C$ は条件式 (2.110) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} -\sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} x_k - \sigma_i \frac{v_i}{\|v_i\|_2} &= 0 \\ \Leftrightarrow v_i &= \frac{-\|v_i\|_2 (\pi_i)^{1-m}}{\sigma_i} \sum_{k=1}^N (u_{i,k})^m x_k. \end{aligned} \quad (2.117)$$

条件式 (2.114) を考慮をするとクラスタ中心の更新式を次のように得る：

$$v_i = \frac{\sum_{k=1}^N (u_{i,k})^m x_k}{\|\sum_{k=1}^N (u_{i,k})^m x_k\|_2}. \quad (2.118)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.111) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} \sum_{k=1}^N (1-m)(\pi_i)^{-m}(u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{-m} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N u_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.119)$$

条件式 (2.113) を考慮して式 (2.119) をクラスインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.120)$$

2.6 混合多項分布に基づくファジィクラスタリング

2.6.1 混合多項分布に基づくクラスタリング

多項分布とは、二項分布に対して確率変数を確率ベクトルに一般化した分布である。 C 個の分布が与えられ、各分布がそれぞれ M 次元多項分布に従うものとしたとき、混合された分布は、多項分布の重ね合わせで記述できる。ここでは、文書を例として変数を説明する。 L_k は $\#k$ 番目の文書における単語の総頻出回数を、 $x_{k,\ell}$ は $\#k$ 番目の文書における $\#\ell$ 番目の単語の頻出回数を表し、以下を満たす：

$$x = \left\{ \{x_{1,1}, \dots, x_{1,M}\}, \dots, \{x_{N,1}, \dots, x_{N,M}\} \mid x_{k,\ell} \in \{0, 1, \dots, L_k\}, \sum_{\ell=1}^M x_{k,\ell} = L_k \right\}. \quad (2.121)$$

$w_{i,\ell}$ は $\#i$ 番目の分布において 1 回の試行で $\#\ell$ 番目の単語が発生する確率を表し、以下を満たす。

$$w = \left\{ \{w_{1,1}, \dots, w_{1,M}\}, \dots, \{w_{C,1}, \dots, w_{C,M}\} \mid 0 \leq w_{i,\ell} \leq 1, \sum_{\ell=1}^M w_{i,\ell} = 1 \right\}. \quad (2.122)$$

$\pi_i = \{\pi_i \in \mathbb{R} \mid i \in \{1, \dots, C\}\}$ は分布の相対的な大きさであり、クラスサイズ調整変数の制約式 (2.14) を満たす。以下に混合多項分布を示す：

$$P_{\text{MMM}}(x_k; w, \pi) = \sum_{i=1}^C \pi_i \Gamma(L_k + 1) \prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)}, \quad (M \in \mathbb{N}, M \geq 2). \quad (2.123)$$

このとき $\{x_k\}_{k=1}^N$ が独立の下で $P_{\text{MMM}}(x_{k,\ell}; w_{i,\ell})$ の尤度を最大化するような (w, π) を求める。以下に対数尤度を示す：

$$\begin{aligned} \log \left(\prod_{k=1}^N P_{\text{MMM}}(x_k; w, \pi) \right) &= \log \left(\prod_{k=1}^N \sum_{i=1}^C \pi_i \Gamma(L_k + 1) \prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)} \right), \\ &= \sum_{k=1}^N \log \left(\sum_{i=1}^C \pi_i \Gamma(L_k + 1) \prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)} \right). \end{aligned} \quad (2.124)$$

つづいて、以下のように帰属度を導入し、イエンセン不等より (u, w, π) を求める最適化問題を導出する：

$$\begin{aligned} &= \sum_{k=1}^N \log \left(\sum_{i=1}^C u_{i,k} \frac{\pi_i}{u_{i,k}} \Gamma(L_k + 1) \prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)} \right), \\ &\geq \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \Gamma(L_k + 1) \prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)} \right), \\ &= \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \left(\log \left(\frac{\pi_i}{u_{i,k}} \right) + \log(\Gamma(L_k + 1)) + \log \left(\prod_{\ell=1}^M \frac{(w_{i,\ell})^{x_{k,\ell}}}{\Gamma(x_{k,\ell} + 1)} \right) \right). \end{aligned} \quad (2.125)$$

したがって、最適化問題は以下のように求まる：

$$\underset{u, w, \pi}{\text{maximize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} + \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{\pi_i}{u_{i,k}} \right) \quad (2.126)$$

$$\Leftrightarrow \underset{u, w, \pi}{\text{minimize}} - \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} + \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right). \quad (2.127)$$

クラスタリングにおけるオブジェクト-クラス間非類似度は以下にあたる：

$$d_{i,k} = - \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell}. \quad (2.128)$$

2.6.2 混合多項分布に基づく KL-divergence 正則化ファジィクラスタリング

混合多項分布に基づくクラスタリングの目的関数における KL-divergence 正則化項にパラメータ $\lambda > 0$ を導入することでファジィ化を行う [7]. この手法を混合多項分布に基づく KL-divergence 正則化ファジィクラスタリング (KL-divergence-regularized Fuzzy Clustering for Categorical Multivariate Data: KLFCCM) と呼ぶ. KLFCCM の最適化問題を以下に示す：

$$\underset{u, w, \pi}{\text{minimize}} - \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \quad (2.129)$$

subject to Eqs. (2.5), (2.11), (2.14), (2.122).

最適化問題 (2.129) に対応する Lagrange 関数 $L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta, \sigma) = & - \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ & + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell} \right). \end{aligned} \quad (2.130)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である. 最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.131)$$

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.132)$$

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.133)$$

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.134)$$

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.135)$$

$$\frac{\partial L_{\text{KLFCCM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.136)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.131) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} d_{i,k} + \lambda^{-1} \left(\log \left(\frac{u_{i,k}}{\pi_i} \right) + 1 \right) - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \exp(\gamma_k - 1 - \lambda d_{i,k}). \end{aligned} \quad (2.137)$$

ここで, $d_{i,k} = -\sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell}$ である. 条件式 (2.134) を考慮して式 (2.137) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \exp(-\lambda(d_{j,k} - d_{i,k}))}. \quad (2.138)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.132) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} -\sum_{k=1}^N u_{i,k} \frac{x_{k,\ell}}{w_{i,\ell}} - \sigma_i &= 0 \\ \Leftrightarrow w_{i,\ell} &= \frac{-1}{\sigma_i} \sum_{k=1}^N u_{i,k} x_{k,\ell}. \end{aligned} \quad (2.139)$$

条件式 (2.136) を考慮をするとクラスタ中心の更新式を次のように得る：

$$w_{i,\ell} = \frac{\sum_{k=1}^N u_{i,k} x_{k,\ell}}{\sum_{k=1}^N \sum_{r=1}^M u_{i,k} x_{k,r}}. \quad (2.140)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.133) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} -\lambda^{-1} \sum_{k=1}^N \frac{u_{i,k}}{\pi_i} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= -\frac{\lambda^{-1} \sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \beta_k}. \end{aligned} \quad (2.141)$$

条件式 (2.135) を考慮して式 (2.141) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\begin{aligned} \pi_i &= \frac{\sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \sum_{j=1}^C u_{j,k}}, \\ &= \frac{\sum_{k=1}^N u_{i,k}}{N}. \end{aligned} \quad (2.142)$$

2.6.3 混合多項分布に基づく q -divergence 正則化ファジィクラスタリング

KLFCCM の目的関数における KL-divergence の代わりに q -divergence を適用することで混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (q -divergence-regularized Fuzzy Clustering for Categorical Multivariate Data: QFCCM) の目的関数を得る. QFCCM の最適化問題を以下に

示す：

$$\begin{aligned} \underset{u, w, \pi}{\text{minimize}} & - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.122).} \end{aligned} \quad (2.143)$$

最適化問題 (2.143) に対応する Lagrange 関数 $L_{\text{QFCCM}}(u, v, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta, \sigma) = & - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} \\ & + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) \\ & + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell} \right). \end{aligned} \quad (2.144)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.145)$$

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.146)$$

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.147)$$

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.148)$$

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.149)$$

$$\frac{\partial L_{\text{QFCCM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.150)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.145) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1} (\pi_i)^{1-m} d_{i,k} + \frac{\lambda^{-1} m}{m-1} (u_{i,k})^{m-1} (\pi_i)^{1-m} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} = \left(\frac{\gamma_k}{m(\pi_i)^{1-m} d_{i,k} + m/\lambda(m-1)(\pi_i)^{1-m}} \right)^{1/(m-1)}. \end{aligned} \quad (2.151)$$

ここで, $d_{i,k} = -\sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell}$ である．条件式 (2.148) を考慮して式 (2.151) をクラスターインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.152)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.146) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} & - \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{x_{k,\ell}}{w_{i,\ell}} - \sigma_i = 0 \\ \Leftrightarrow w_{i,\ell} &= \frac{-(\pi_i)^{1-m}}{\sigma_i} \sum_{k=1}^N (u_{i,k})^m x_{k,\ell}. \end{aligned} \quad (2.153)$$

条件式 (2.150) を考慮をするとクラスタ中心の更新式を次のように得る：

$$w_{i,\ell} = \frac{\sum_{k=1}^N (u_{i,k})^m x_{k,\ell}}{\sum_{k=1}^N \sum_{r=1}^M (u_{i,k})^m x_{k,r}}. \quad (2.154)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.147) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} & \sum_{k=1}^N (1-m)(\pi_i)^{-m} (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{-m} - \sum_{k=1}^N \beta_k = 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N (u_{i,k})^m} \right)^{-1/m}. \end{aligned} \quad (2.155)$$

条件式 (2.149) を考慮して式 (2.155) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.156)$$

2.6.4 混合多項分布に基づく Bezdek-type ファジィクラスタリング

QFCCM の目的関数における q -divergence 正則化項のパラメータを $\lambda \rightarrow \infty$ とすることで新たな目的関数を得る．これを混合多項分布に基づく Bezdek-type ファジィクラスタリング (Bezdek-type Fuzzy Clustering for Categorical Multivariate Data: BFCCM) と呼ぶ．BFCCM の最適化問題を以下に示す：

$$\begin{aligned} & \underset{u, w, \pi}{\text{minimize}} - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.122)}. \end{aligned} \quad (2.157)$$

最適化問題 (2.157) に対応する Lagrange 関数 $L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta, \sigma) &= - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell} \\ &+ \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell} \right). \end{aligned} \quad (2.158)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.159)$$

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.160)$$

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.161)$$

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.162)$$

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.163)$$

$$\frac{\partial L_{\text{BFCCM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.164)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.159) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1}(\pi_i)^{1-m}d_{i,k} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \left(\frac{\gamma_k}{md_{i,k}} \right)^{1/(m-1)}. \end{aligned} \quad (2.165)$$

ここで, $d_{i,k} = -\sum_{\ell=1}^M \log(w_{i,\ell})x_{k,\ell}$ である．条件式 (2.162) を考慮して式 (2.165) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{1/(1-m)}}. \quad (2.166)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.160) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} -\sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{x_{k,\ell}}{w_{i,\ell}} - \sigma_i &= 0 \\ \Leftrightarrow w_{i,\ell} &= \frac{-(\pi_i)^{1-m}}{\sigma_i} \sum_{k=1}^N (u_{i,k})^m x_{k,\ell}. \end{aligned} \quad (2.167)$$

条件式 (2.164) を考慮をするとクラスタ中心の更新式を次のように得る：

$$w_{i,\ell} = \frac{\sum_{k=1}^N (u_{i,k})^m x_{k,\ell}}{\sum_{k=1}^N \sum_{r=1}^M (u_{i,k})^m x_{k,r}}. \quad (2.168)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.161) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} \sum_{k=1}^N (1-m)(u_{i,k})^m (\pi_i)^{-m} d_{i,k} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.169)$$

条件式 (2.163) を考慮して式 (2.169) をクラスインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m d_{j,k}}{(u_{i,k})^m d_{i,k}} \right)^{1/m}}. \quad (2.170)$$

2.6.5 混合多項分布に基づくファジィクラスタリングの変形

混合多項分布に基づくファジィクラスタリング {KL,Q,B}FCCM における非類似度の対数関数を一般化することで新たな目的関数を得る：

$$\log a = \lim_{t \rightarrow 0} \frac{1}{t} (a^t - 1) \text{ subject to } (a > 0), \quad (2.171)$$

式 (2.171) を用いて式 (2.128) に代入すると以下のように求まる：

$$d_{i,k} = -\frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell}. \quad (2.172)$$

非類似度に式 (2.172) を適用したファジィクラスタリング手法を Fuzzy Clustering for Categorical Multivariate data induced by Multinomial mixture model (FCCMM) と呼び、 $t \rightarrow 0$ とすることで FCCM と理論的に同義となる [8].

2.6.6 混合多項分布に基づく KL-divergence 正則化ファジィクラスタリング

KLFCMM の非類似度を式 (2.172) に置き換えた手法を混合多項分布に基づく KL-divergence 正則化ファジィクラスタリング (KL-divergence-regularized Fuzzy Clustering for Categorical Multivariate data induced by Multinomial mixture model: KLFCMM) と呼ぶ。KLFCMM の最適化問題を以下に示す：

$$\begin{aligned} \underset{u, w, \pi}{\text{minimize}} & - \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.122)}. \end{aligned} \quad (2.173)$$

最適化問題 (2.173) に対応する Lagrange 関数 $L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta, \sigma) = & - \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} \\ & + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\pi_i} \right) + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) \\ & + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell} \right). \end{aligned} \quad (2.174)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.175)$$

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.176)$$

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.177)$$

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.178)$$

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.179)$$

$$\frac{\partial L_{\text{KLFCMM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.180)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.175) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} d_{i,k} + \lambda^{-1} \left(\log \left(\frac{u_{i,k}}{\pi_i} \right) + 1 \right) - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \exp(\gamma_k - 1 - \lambda d_{i,k}). \end{aligned} \quad (2.181)$$

ここで, $d_{i,k} = -\frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell}$ である．条件式 (2.178) を考慮して式 (2.181) をクラスティンデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \exp(-\lambda(d_{j,k} - d_{i,k}))}. \quad (2.182)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.176) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} -\sum_{k=1}^N u_{i,k} (w_{i,\ell})^{t-1} x_{k,\ell} - \sigma_i &= 0 \\ \Leftrightarrow w_{i,\ell} &= \left(-\sigma_i \sum_{k=1}^N u_{i,k} x_{k,\ell} \right)^{1/(1-t)}. \end{aligned} \quad (2.183)$$

条件式 (2.180) を考慮をするとクラスタ中心の更新式を次のように得る：

$$w_{i,\ell} = \frac{1}{\sum_{r=1}^M \left(\sum_{k=1}^N \frac{u_{i,k} x_{k,r}}{u_{i,k} x_{k,\ell}} \right)^{1/(1-t)}}. \quad (2.184)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.177) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} -\lambda^{-1} \sum_{k=1}^N \frac{u_{i,k}}{\pi_i} - \sum_{k=1}^N \beta_k &= 0 \\ \Leftrightarrow \pi_i &= -\frac{\lambda^{-1} \sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \beta_k}. \end{aligned} \quad (2.185)$$

条件式 (2.179) を考慮して式 (2.185) をクラスインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスサイズ調整変数の更新式を次のように得る：

$$\begin{aligned}\pi_i &= \frac{\sum_{k=1}^N u_{i,k}}{\sum_{k=1}^N \sum_{j=1}^C u_{j,k}}, \\ &= \frac{\sum_{k=1}^N u_{i,k}}{N}.\end{aligned}\quad (2.186)$$

2.6.7 混合多項分布に基づく Bezdek-type ファジィクラスタリング

BFCCM の非類似度を式 (2.172) に置き換えた手法を混合多項分布に基づく Bezdek-type ファジィクラスタリング (Bezdek-type Fuzzy Clustering for Categorical Multivariate data induced by Multinomial mixture model: BFCCMM) と呼ぶ。BFCCMM の最適化問題を以下に示す：

$$\begin{aligned}\underset{u, w, \pi}{\text{minimize}} & - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.122).}\end{aligned}\quad (2.187)$$

最適化問題 (2.187) に対応する Lagrange 関数 $L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned}L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta, \sigma) &= - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} \\ &+ \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k}\right) + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i\right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell}\right).\end{aligned}\quad (2.188)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である。最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.189)$$

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.190)$$

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.191)$$

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.192)$$

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.193)$$

$$\frac{\partial L_{\text{BFCCMM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.194)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.189) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned}m(u_{i,k})^{m-1} (\pi_i)^{1-m} d_{i,k} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} &= \pi_i \left(\frac{\gamma_k}{m d_{i,k}} \right)^{1/(m-1)}.\end{aligned}\quad (2.195)$$

ここで, $d_{i,k} = -\frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell}$ である. 条件式 (2.192) を考慮して式 (2.195) をクラスティンデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る:

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{d_{j,k}}{d_{i,k}} \right)^{1/(1-m)}}. \quad (2.196)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.190) によって Lagrange 乗数 σ_i を用いて次のように表される:

$$\begin{aligned} & - \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (w_{i,\ell})^{t-1} x_{k,\ell} - \sigma_i = 0 \\ \Leftrightarrow w_{i,\ell} &= \left(-\sigma_i \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} x_{k,\ell} \right)^{1/(1-t)}. \end{aligned} \quad (2.197)$$

条件式 (2.194) を考慮をするとクラスタ中心の更新式を次のように得る:

$$w_{i,\ell} = \frac{1}{\sum_{r=1}^M \left(\sum_{k=1}^N \frac{(u_{i,k})^m x_{k,r}}{(u_{i,k})^m x_{k,\ell}} \right)^{1/(1-t)}}. \quad (2.198)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.191) によって Lagrange 乗数 β_k を用いて次のように表される:

$$\begin{aligned} & \sum_{k=1}^N (1-m)(u_{i,k})^m (\pi_i)^{-m} d_{i,k} - \sum_{k=1}^N \beta_k = 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.199)$$

条件式 (2.193) を考慮して式 (2.199) をクラスティンデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る:

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m d_{j,k}}{(u_{i,k})^m d_{i,k}} \right)^{1/m}}. \quad (2.200)$$

2.7 提案法 混合多項分布に基づく q -divergence 正則化ファジィクラスタリング

QFCCM の非類似度を式 (2.172) に置き換えた手法を混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (q -divergence-regularized Fuzzy Clustering for Categorical Multivariate data induced by Multinomial mixture model: QFCCMM) を提案する. QFCCMM の最適化問題を以下に示す:

$$\begin{aligned} & \underset{u, w, \pi}{\text{minimize}} - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \\ & \text{subject to Eqs. (2.5), (2.11), (2.14), (2.122)}. \end{aligned} \quad (2.201)$$

最適化問題 (2.201) に対応する Lagrange 関数 $L_{\text{QFCCMM}}(u, v, \pi, \gamma, \beta, \sigma)$ は次のように表される：

$$\begin{aligned} L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta, \sigma) = & - \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} \frac{1}{t} \sum_{\ell=1}^M ((w_{i,\ell})^t - 1) x_{k,\ell} \\ & + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} + \sum_{k=1}^N \gamma_k \left(1 - \sum_{i=1}^C u_{i,k} \right) \\ & + \sum_{k=1}^N \beta_k \left(1 - \sum_{i=1}^C \pi_i \right) + \sum_{i=1}^C \sigma_i \left(1 - \sum_{\ell=1}^M w_{i,\ell} \right). \end{aligned} \quad (2.202)$$

ここで $\gamma = (\gamma_1, \dots, \gamma_N)$, $\beta = (\beta_1, \dots, \beta_N)$, $\sigma = (\sigma_1, \dots, \sigma_C)$ は Lagrange 乗数である．最適性の必要条件は次のように表される：

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial u_{i,k}} = 0, \quad (2.203)$$

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial w_{i,\ell}} = 0, \quad (2.204)$$

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \pi_i} = 0, \quad (2.205)$$

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \gamma_k} = 0, \quad (2.206)$$

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta)}{\partial \beta_k} = 0, \quad (2.207)$$

$$\frac{\partial L_{\text{QFCCMM}}(u, w, \pi, \gamma, \beta, \sigma)}{\partial \sigma_i} = 0. \quad (2.208)$$

最適な帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ は条件式 (2.203) によって Lagrange 乗数 γ_k を用いて次のように表される：

$$\begin{aligned} m(u_{i,k})^{m-1} (\pi_i)^{1-m} d_{i,k} + \frac{\lambda^{-1}m}{m-1} (u_{i,k})^{m-1} (\pi_i)^{1-m} - \gamma_k &= 0 \\ \Leftrightarrow u_{i,k} = \left(\frac{\gamma_k}{m(\pi_i)^{1-m} d_{i,k} + m/\lambda(m-1)(\pi_i)^{1-m}} \right)^{1/(m-1)}. \end{aligned} \quad (2.209)$$

ここで, $d_{i,k} = -\sum_{\ell=1}^M \log(w_{i,\ell}) x_{k,\ell}$ である．条件式 (2.206) を考慮して式 (2.209) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとると帰属度の更新式を次のように得る：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.210)$$

最適なクラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,M)}$ は条件式 (2.204) によって Lagrange 乗数 σ_i を用いて次のように表される：

$$\begin{aligned} - \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} (w_{i,\ell})^{t-1} x_{k,\ell} - \sigma_i &= 0 \\ \Leftrightarrow w_{i,\ell} = \left(-\sigma_i \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{1-m} x_{k,\ell} \right)^{1/(1-t)}. \end{aligned} \quad (2.211)$$

条件式 (2.208) を考慮をするとクラスタ中心の更新式を次のように得る：

$$w_{i,\ell} = \frac{1}{\sum_{r=1}^M \left(\sum_{k=1}^N \frac{(u_{i,k})^m x_{k,r}}{(u_{i,k})^m x_{k,\ell}} \right)^{1/(1-t)}}. \quad (2.212)$$

最適なクラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ は条件式 (2.205) によって Lagrange 乗数 β_k を用いて次のように表される：

$$\begin{aligned} & \sum_{k=1}^N (1-m)(\pi_i)^{-m} (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N (u_{i,k})^m (\pi_i)^{-m} - \sum_{k=1}^N \beta_k = 0 \\ \Leftrightarrow \pi_i &= \left(\frac{\sum_{k=1}^N \beta_k}{(1-m) \sum_{k=1}^N (u_{i,k})^m d_{i,k} - \lambda^{-1} \sum_{k=1}^N u_{i,k}} \right)^{-1/m}. \end{aligned} \quad (2.213)$$

条件式 (2.207) を考慮して式 (2.213) をクラスタインデックス $j \in \{1, \dots, C\}$ について総和をとるとクラスタサイズ調整変数の更新式を次のように得る：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.214)$$

以上より，QFCCMM は次のアルゴリズムにまとめられる：

アルゴリズム 7 (QFCCMM).

STEP 1. ファジィ化パラメータ $\lambda > 0$ ，初期クラスタ中心 $\{w_{i,\ell}\}_{(i,\ell)=(1,1)}^{(C,N)}$ と初期クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を設定する．

STEP 2. オブジェクト-クラスタ間非類似度 $\{d_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する：

$$d_{i,k} = -\frac{1}{t} \sum_{\ell=1}^M \left((w_{i,\ell})^t - 1 \right) x_{k,\ell}. \quad (2.215)$$

STEP 3. 帰属度 $\{u_{i,k}\}_{(i,k)=(1,1)}^{(C,N)}$ を次のように設定する：

$$u_{i,k} = \frac{1}{\sum_{j=1}^C \frac{\pi_j}{\pi_i} \left(\frac{1-\lambda(1-m)d_{j,k}}{1-\lambda(1-m)d_{i,k}} \right)^{1/(1-m)}}. \quad (2.216)$$

STEP 4. クラスタ中心 $\{w_{i,\ell}\}_{i=1}^C$ を次のように設定する：

$$w_{i,\ell} = \frac{1}{\sum_{r=1}^M \left(\sum_{k=1}^N \frac{(u_{i,k})^m x_{k,r}}{(u_{i,k})^m x_{k,\ell}} \right)^{\frac{1}{1-t}}}. \quad (2.217)$$

STEP 5. クラスタサイズ調整変数 $\{\pi_i\}_{i=1}^C$ を次のように設定する：

$$\pi_i = \frac{1}{\sum_{j=1}^C \left(\sum_{k=1}^N \frac{(u_{j,k})^m (1-\lambda(1-m)d_{j,k})}{(u_{i,k})^m (1-\lambda(1-m)d_{i,k})} \right)^{1/m}}. \quad (2.218)$$

STEP 6. (u, v, π) が収束すれば終了，そうなければ STEP 2. へ。 ■

2.8 おわりに

本章では、ファジィクラスタリングについて、関連研究を述べた上で、混合多項分布に基づく q -divergence 正則化ファジィクラスタリングを提案した。まず、代表的なクラスタリング手法である HCM について最適化問題からクラスタ中心と帰属度の更新式を導出し、アルゴリズムを記述した。次に初期クラスタ中心を決定する指針として代表的なアルゴリズムである k -means++ について記述した。次に三つのファジィ化によるファジィクラスタリング $\{B, KL, Q\}$ FCM について記述した。次に HCM における非類似度をデータとクラスタ中心の内積によって表現し、球面上に分布したデータを分析するのに適した球面データのためのファジィクラスタリング $\{B, KL, Q\}$ FCS について述べた。次に混合多項分布に基づくクラスタリングの最適化問題を導出し、それから派生したファジィクラスタリング $\{KL, Q, B\}$ FCCM と、非類似度を変形した $\{KL, B\}$ FCCMM について記述した。最後に提案法である混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (QFCCMM) を提案し、その最適化問題とクラスタ中心、帰属度、クラスタサイズ調整変数の更新式を導出し、アルゴリズムを示した。

第 3 章

ファジィクラスタリングにおける数値実験

3.1 はじめに

本章では、文書データを使用した九つのファジィクラスタリング手法を適用した数値実験について記述し、その精度を比較する。まず第 3.2 節でベンチマークデータである文書データについて記述する。次に第 3.3 節で文書中の単語の重みに基づいてデータを変換する二通りの Term Frequency Inverse Document Frequency (tf.idf) について示す。次に第 3.4 節でクラスタリング精度を評価する指標について記述する。次に第 3.5 節で正解の帰属度を初期値として設定する数値実験 1 について記述し、その結果を示し、精度比較する。次に第 3.6 節でクラスタ中心を初期値として設定し、10 通り与える数値実験 2 について記述し、その結果を示し、精度比較する。次に第 3.7 節で数値実験 2 で適用したパラメータと同様の設定で 100 通りの初期値を与える数値実験 3 について記述し、その結果を示し、精度比較する。最後に第 3.8 節で、数値実験の考察を行う。

3.2 文書データ

クラスタリングの性能評価のため、CLUTO や LINQS から 27 の文書データを得た [27, 28, 29]。cacmcisi は、4663 の文書から構成され、“cacm” と “cisi” の 2 クラスへ分けられたデータである。citeseer は、3312 の学術雑誌から構成され、“Agents”, “AI”, “DB”, “IR”, “ML”, “HCI” とラベル付けされた 6 クラスに分けられたデータである。classic と classic4 は、7094 の文書から構成され、“cran”, “med”, “cacm”, “cisi” の 4 クラスへ分けられたデータである。cora は、2708 の学術雑誌から構成され “Case-Based”, “Genetic_Algorithms”, “Neural_Networks”, “Probabilistic_Methods”, “Reinforcement_Learning”, “Rule_Learning”, “Theory” とラベル付けされた 7 クラスに分けられたデータである。cranmed は、2431 の文書から構成され、“cran” と “med” の 2 クラスへ分けられたデータである。fbis, hitech, la1, la2, la12, new3, reviews, sports, tr11, tr12, tr23, tr31, tr41, tr45 は、Text REtrieval Conference [30] から派生したデータである。fbis は、Foreign Broadcast Information Service (FBIS) のデータであり、TREC-5 [31] によって 17 クラスへ整形された。hitech, reviews, sports は、San Jose Mercury のニュース記事が元のデータである。hitech は、2301 の技術記事から構成され、“computer”, “electronics”, “medical”, “health”, “research”, “technology” の 6 クラスに分けられたデータである。reviews は、4069 の批評記事から構成され、“food”, “movie”, “music”, “radio”, “restaurant” の 5 クラスへ分類された。sports は、8580 のスポー

ツ記事から構成され, “baseball”, “basketball”, “football”, “hockey”, “boxing”, “bicycle”, “golf” の7クラスへ分類された. 1a1, 1a2, 1a12 は, Los Angeles Times から発行されたニュース記事を “Financial”, “Foreign”, “National”, “Metro”, “Sports”, “Entertainment” の6クラスへ分けられたデータである. k1a, k1b, wap は, WebACE project [32] によるデータある. mm は, 2521 の文書から構成されたデータであり, “movie” と “music” の2クラスへ分類されている. ohscal は, OHSUMED [33] から得られたデータであり, “DNA”, “Carcinoma”, “Tomography”, “In-Vitro”, “Antibodies”, “Risk-Factors”, “Prognosis”, “Receptors”, “Pregnancy”, “Molecular-Sequence-Data” とラベル付けされた10クラスへ分類された. re0 と re1 は, Reuters のニュース記事が元のデータである [34].

文書データは, 文書中の単語の頻出回数をベクトル化することで数値化される. ただし LINQS から得た citeseer, cora, webkb は, 単語が頻出したかどうかを0か1で表現されている. 表3.1に使用したデータを示す. 表中の実要素は, 頻度行列に存在する0でない要素数を意味し, スパース性は, $100 - \text{全要素数に対する実要素数の割合}$ である.

3.3 tf.idf

情報検索や文書分類の分野では, 単語頻度データを用いて一般的な単語の重みを下げ, 特定の文書にしか現れない単語の重みを上げる tf.idf がよく行われる [35]. tf (Term Frequency) とは単語の頻出回数に基づいて算出される. $tf_{k,\ell}$ は, $\#k$ 番目の文書中の $\#\ell$ 番目の単語の tf である. idf (Inverse Document Frequency) は, どの文書にも現れる一般語の重要度を下げ, 特定の文書にしか現れない単語の重要度を上げるフィルタの役割を持つ. idf_{ℓ} は, $\#\ell$ 番目の単語の idf である. $n_{k,\ell}$ を k 番目の文書の $\#\ell$ 番目の単語の頻出回数とし, N を全文書数, M を総単語数, df_{ℓ} をある単語 $\#\ell$ が出現する文書数とする. 本研究では, tf.idf を施さず頻度行列のまま扱ったデータを tf.idf なしと表記し, tf.idf なしと二通りの tf.idf 化によるデータを数値実験に採用する.

$$tf_{k,\ell}^A = \frac{n_{k,\ell}}{\sum_{r=1}^M n_{k,r}} \quad (3.1)$$

$$tf_{k,\ell}^B = \begin{cases} 1 + \log(n_{k,\ell}) & (tf_{k,\ell} > 0 \text{ のとき}), \\ 0 & (\text{それ以外のとき}). \end{cases} \quad (3.2)$$

$$idf_{\ell}^A = \log\left(\frac{N}{df_{\ell}}\right) \quad (3.3)$$

$$idf_{\ell}^B = \log\left(1 + \frac{N}{df_{\ell}}\right) \quad (3.4)$$

tf.idf は $tf_{k,\ell}$ と idf_{ℓ} の乗算から求まる:

$$tf.idf_{k,\ell}^A = tf_{k,\ell}^A idf_{\ell}^A, \quad (3.5)$$

$$tf.idf_{k,\ell}^B = tf_{k,\ell}^B idf_{\ell}^B. \quad (3.6)$$

式 (3.5) を tf.idfA, 式 (3.6) を tf.idfB と名付けた.



表 3.1: 数値実験に使用する文書データ

| データセット名 | 文書数 | クラス数 | 単語数 | 実要素 | スパース性 |
|----------|-------|------|-------|---------|----------|
| cacmcisi | 4663 | 2 | 14409 | 83181 | 99.8762% |
| citeseer | 3312 | 6 | 3703 | 105165 | 99.1425% |
| classic | 7094 | 4 | 41681 | 223839 | 99.9243% |
| classic4 | 7094 | 4 | 5896 | 247158 | 99.4091% |
| cora | 2708 | 7 | 1432 | 49216 | 98.7308% |
| cranmed | 2431 | 2 | 31720 | 140658 | 99.8176% |
| fbis | 2463 | 17 | 2000 | 393386 | 92.0141% |
| hitech | 2301 | 6 | 22498 | 346881 | 99.3299% |
| k1a | 2340 | 20 | 21839 | 349792 | 99.3155% |
| k1b | 2340 | 6 | 21839 | 349792 | 99.3155% |
| la1 | 3204 | 6 | 29714 | 484024 | 99.4916% |
| la2 | 3075 | 6 | 19692 | 455383 | 99.2480% |
| la12 | 6279 | 6 | 30125 | 939407 | 99.5034% |
| mm | 2521 | 2 | 29973 | 490062 | 99.3514% |
| new3 | 9558 | 44 | 70822 | 2295120 | 99.6609% |
| ohscal | 11162 | 10 | 11465 | 674365 | 99.4730% |
| re0 | 1504 | 13 | 2886 | 77808 | 98.2074% |
| re1 | 1657 | 25 | 3758 | 87328 | 98.5976% |
| reviews | 4069 | 5 | 36746 | 781635 | 99.4772% |
| sports | 8580 | 7 | 27673 | 1107980 | 99.5334% |
| tr11 | 414 | 9 | 6429 | 116613 | 95.6187% |
| tr12 | 313 | 8 | 5804 | 85640 | 95.2858% |
| tr23 | 204 | 6 | 5832 | 78609 | 93.3927% |
| tr31 | 927 | 7 | 10128 | 248903 | 97.3489% |
| tr41 | 878 | 10 | 7454 | 171509 | 97.3794% |
| tr45 | 690 | 10 | 8261 | 193605 | 96.6035% |
| wap | 1560 | 20 | 8460 | 220482 | 98.3294% |
| webkb | 877 | 5 | 1703 | 79365 | 94.6861% |

3.4 精度評価方法

3.4.1 Contingency Table

Contingency Table とはクラスタリングした結果を表で示したものである。 C をクラス数, $\#i$ 番目のクラス G_i^* に属しかつ $\#j$ 番目のクラス G_j にクラスタリングされたオブジェクト数を $n_{i,j}$ とする。 a を G^* に属すオブジェクト数の合計, b を G に属すオブジェクト数の合計, N を総オブジェクト数とすると Contingency Table は, 表 3.2 のように表される。

表 3.2: Contingency Table

| Cluster \ Class | G_1 | \cdots | G_C | Sums |
|-----------------|-----------|----------|-----------|----------|
| G_1^* | $n_{1,1}$ | \cdots | $n_{1,C}$ | a_1 |
| \vdots | \vdots | \ddots | \vdots | \vdots |
| G_C^* | $n_{C,1}$ | \cdots | $n_{C,C}$ | a_C |
| Sums | b_1 | \cdots | b_C | N |

3.4.2 Adjusted Rand Index (ARI)

Adjusted Rand index (ARI) [36] とは Contingency Table をから求め、クラスタリングの精度をスカラー値としたものである。ARI は -1 から 1 までの値を取り、高いほど精度が高いことを表す。以下に ARI の算出式を示す：

$$\text{ARI} = \frac{\sum_{i=1}^C \sum_{j=1}^C \binom{n_{i,j}}{2} - \left(\sum_{i=1}^C \binom{a_i}{2} \sum_{j=1}^C \binom{b_j}{2} \right) / \binom{N}{2}}{\frac{1}{2} \left(\sum_{i=1}^C \binom{a_i}{2} + \sum_{j=1}^C \binom{b_j}{2} \right) - \left(\sum_{i=1}^C \binom{a_i}{2} \sum_{j=1}^C \binom{b_j}{2} \right) / \binom{N}{2}} \quad (3.7)$$

3.5 数値実験 1

tf.idf なしと二通りの tf.idf によるデータ形式へ加工後、最適化問題の制約に基づき FCS では式 (3.8) に従い球面化し、FCCM と FCCMM では式 (3.9) に従い単体化することでデータの正規化を行う。 $x_k^{\text{Normalize1}}$ はデータ x を構成するオブジェクト $\#k$ の球面化を、 $x_k^{\text{Normalize2}}$ は単体化を表す。

$$x_k^{\text{Normalize1}} = \frac{x_k}{\|x_k\|_2}, \quad (3.8)$$

$$x_k^{\text{Normalize2}} = \frac{x_k}{\sum_{\ell=1}^M x_{k,\ell}}. \quad (3.9)$$

更新前と更新後の帰属度、クラスタ中心、クラスタサイズ調整変数それぞれの差における最大値ノルムの和が 10^{-6} を下回るとき収束したとみなし、更新回数が 10^4 を超えても収束しない場合は、そのパラメータでの処理を終える。収束前に処理を終えたパラメータは、クラスタリングが不完全に終わったとみなし数値実験の結果に反映しない。式 (3.10)、式 (3.11) に従い正解のファジィな帰属度とクラスタサイズ調整変数を初期値として与え、クラスタ数 C を各データの正解クラス数として設定し、表 3.3 に示すパラメータ範囲で数値実験 1 を行う。 $u_{i,k}^{\text{Init}}$ はクラスタ $\#i$ におけるオブジェクト $\#k$ の初期帰属度を、 π_i^{Init} はクラスタ $\#i$ における初期クラスタサイズ調整変数を表す。

$$u_{i,k}^{\text{Init}} = \begin{cases} 1 - (C - 1)/100 & (x_k \in G_i^*), \\ 0.01 & (x_k \notin G_i^*). \end{cases} \quad (3.10)$$

$$\pi_i^{\text{Init}} = \frac{1}{\sum_{k: x_k \in G_i^*} N} \quad (3.11)$$

以下に数値実験 1 の流れを示す：

実験の流れ 1 (数値実験 1).

表 3.3: 文書データ数値実験におけるファジィクラスタリングで採用したパラメータ範囲

| |
|--|
| $\lambda \in \{0.5, 1.0, 10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ |
| $m \in \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}, 1 + 10^{-2}, 1 + 10^{-1}\},$ |
| $t \in \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}, 0.5\}$ |

- STEP 1. 表 3.1 に示したデータの読み込み, 正解の帰属度の読み込みをする.
- STEP 2. tf.idf なしでない場合, 式 (3.5), 式 (3.6) に従い tf.idfA や tf.idfB に変換する.
- STEP 3. {B,KL,Q}FCS では式 (3.8) に従い球面化し, {B,KL,Q}FC{CM,CMM} では式 (3.9) に従い単体化する.
- STEP 4. クラスタ数を正解クラス数に設定する.
- STEP 5. 表 3.3 の範囲でパラメータを設定する.
- STEP 6. 式 (3.10) に従い初期帰属度を設定する.
- STEP 7. 式 (3.11) に従い初期クラスタサイズ調整変数を設定する.
- STEP 8. BFCS では式 (2.90), KLFCS では式 (2.104), QFCS では式 (2.118), KLFCCM では式 (2.140), BFCCM では式 (2.168), QFCCM では式 (2.154), KLFCCMM では式 (2.184), BFCCMM では式 (2.198), QFCCMM では式 (2.212) に従いクラスタ中心を更新する.
- STEP 9. {B,KL,Q}FCS では式 (2.78), {KL,Q,B}FCCM では式 (2.128), {KL,B,Q}FCCMM では式 (2.172) に従いオブジェクト-クラスタ間非類似度を更新する.
- STEP 10. BFCS では式 (2.88), KLFCS では式 (2.102), QFCS では式 (2.116), KLFCCM では式 (2.138), BFCCM では式 (2.166), QFCCM では式 (2.152), KLFCCMM では式 (2.182), BFCCMM では式 (2.196), QFCCMM では式 (2.210) に従い帰属度を更新する.
- STEP 11. BFCS では式 (2.92), KLFCS では式 (2.106), QFCS では式 (2.120), KLFCCM では式 (2.142), BFCCM では式 (2.170), QFCCM では式 (2.156), KLFCCMM では式 (2.186), BFCCMM では式 (2.200), QFCCMM では式 (2.214) に従いクラスタサイズ調整変数を更新する.
- STEP 12. 更新前と更新後の帰属度, クラスタ中心, クラスタサイズ調整変数の最大値ノルムを計算し, それらの和が 10^{-6} 以上であれば STEP 8. へ, また更新数が 10^4 を超えたらそのパラメータでのクラスタリングを不可として STEP 16. へ.
- STEP 13. 収束した帰属度をクリस्प化する.
- STEP 14. クリस्प化された帰属度と正解の帰属度を照らし合わせ Contingency Table を得る.
- STEP 15. Contingency Table から式 (3.7) に従い ARI を算出する.
- STEP 16. 次に設定すべきパラメータがあれば STEP 5. へ.
- STEP 17. 次に処理すべきデータ変換手法があれば STEP 2. へ.
- STEP 18. 次に処理すべきデータがあれば STEP 1. へ, そうでなければ終了する. ■

tf.idf なしを適用した各ファジィクラスタリング手法ごとの最高 ARI とそれらを達成したパラメータを表 3.4 から表 3.31 にまとめた. tf.idfA を適用した各ファジィクラスタリング手法ごとの最高 ARI とそれらを達成したパラメータを表 3.32 から表 3.59 にまとめた. tf.idfB を適用した各ファジィクラスタリング手法ごとの最高 ARI とそれらを達成したパラメータを表 3.60 から表 3.87 にまとめた. 九つのファジィクラスタリング手法のうち最高 ARI を達成した値を赤色で強調する. 27 の文書データにおける tf.idf なし, tf.idfA, tf.idfB を施した各ファジィクラスタリング手法ごとの最高 ARI を表 3.88,

表 3.89, 表 3.90 にまとめた.

表 3.4: cacmcisi dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|-----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.719538 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.719538 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.809621 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.809621 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.809621 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCs | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCS | -0.027071 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0 | $\lambda = \{0.5, 1.0\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, \dots, 1 + 10^{-1}\}$ |

表 3.5: citeseer dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.555670 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.555283 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.781548 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.781548 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.781548 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCs | 0.365114 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.365114 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.365114 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.6: classic dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.843743 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.842978 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.892567 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-2}\}$ |
| BFCCMM | 0.892567 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-2}\}$ |
| QFCCMM | 0.892567 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-2}\}$ |
| KLFCS | 0.411013 | $\lambda = \{10^2\}$ |
| BFCS | 0.410876 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.418826 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.7: classic4 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0.000746 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.487289 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.487289 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.783449 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.782738 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.782423 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.441396 | $\lambda = \{10^2\}$ |
| BFCS | 0.442482 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.447722 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.8: cora dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0.400710 | $\lambda = \{10\}$ |
| BFCCM | 0.558195 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.558224 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.745177 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}\}$ |
| BFCCMM | 0.745177 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}\}$ |
| QFCCMM | 0.745177 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}\}$ |
| KLFCS | 0.181326 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.182860 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.182860 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.9: cranmed dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0.978704 | $\lambda = \{10\}$ |
| BFCCM | 1.0 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 1.0 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.957636 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.957636 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.957636 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.10: fbis dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.548472 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.561336 | $\lambda = \{10\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.727048 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.727048 | $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.727048 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.465904 | $\lambda = \{10^3\}$ |
| BFCS | 0.466871 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.466871 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.11: hitech dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.500573 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.498074 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.928259 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.928259 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.928259 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.279298 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.279298 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.279298 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.12: k1a dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.951056 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.951056 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.976298 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.976298 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.976298 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.631614 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.632257 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.632257 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.13: k1b dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0.802794 | $\lambda = \{10\}$ |
| BFCCM | 0.894707 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.894707 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.930587 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.930587 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.930587 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.815651 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.815651 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.815651 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.14: la1 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.876117 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.875392 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.972516 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.972516 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.972516 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.296463 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.296463 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.296463 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.15: la2 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.670654 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.669151 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.955918 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.955918 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.955918 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.336259 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.336259 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.336259 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.16: la12 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.590707 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.586742 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.932916 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.932916 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.932916 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.290656 | $\lambda = \{10^3\}$ |
| BFCS | 0.287174 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.290867 | $\lambda = \{10^3\}, m = \{1 + 10^{-3}\}$ |

表 3.17: mm dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.913082 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.908538 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.996828 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.996828 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.996828 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.866667 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.866667 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.866667 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.18: new3 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.711023 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.702324 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.885880 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.885880 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.885880 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.457092 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.456898 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.457092 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.19: ohscal dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0.358778 | $\lambda = \{10\}$ |
| BFCCM | 0.396097 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.395148 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.711620 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.711786 | $m = \{1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.711786 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^5\}, t = \{10^{-4}\}$ |
| KLFCS | 0.229789 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.229587 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.229789 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.20: re0 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.260616 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.281009 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.624956 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.624781 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.624299 | $\lambda = \{10^3\}, m = \{1 + 10^{-4}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.292337 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.301650 | $m = \{1 + 10^{-1}\}$ |
| QFCS | 0.301444 | $\lambda = \{10^3\}, m = \{1 + 10^{-5}\}$ |

表 3.21: re1 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.801896 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.799243 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.837132 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.837132 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.843422 | $\lambda = \{10^2\}, m = \{1 + 10^{-3}\}, t = \{10^{-2}\}$ |
| KLFCs | 0.352648 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.355002 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.355002 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.22: reviews dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.887301 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.887301 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.986761 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.986761 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.986761 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCs | 0.700147 | $\lambda = \{10^2\}$ |
| BFCS | 0.700464 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.700451 | $\lambda = \{10^2\}, m = \{1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.23: sports dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.892871 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.891601 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.967107 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.967107 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.967107 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.431274 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.429228 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.490517 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.24: tr11 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0.671194 | $\lambda = \{10\}$ |
| BFCCM | 0.930590 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.930590 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.987652 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\}$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.987652 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.987652 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.820268 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.820268 | $m = \{1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.820268 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.25: tr12 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.942499 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.942499 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| BFCCMM | 0.980433 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| KLFCS | 0.695190 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.695190 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.695190 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.26: tr23 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0.382848 | $\lambda = \{10\}$ |
| BFCCM | 0.942754 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.942754 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.954420 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.954420 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.954420 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.296111 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}$ |
| BFCS | 0.305340 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.305340 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-2}\}$ |

表 3.27: tr31 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0.766094 | $\lambda = \{10\}$ |
| BFCCM | 0.913056 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.913056 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.971071 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.971071 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.971071 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.549255 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.554192 | $m = \{1 + 10^{-3}\}$ |
| QFCS | 0.554192 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-3}\}$ |

表 3.28: tr41 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.951001 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.951001 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.976203 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.976203 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.976203 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.602412 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.602412 | $m = \{1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.602412 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.29: tr45 dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.941595 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.941595 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.979758 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.979758 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.979758 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.503606 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.504575 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.504575 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.30: wap dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.941963 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.940369 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.980827 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.980827 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.980827 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.628756 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.626042 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.628756 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.31: webkb dataset の最高 ARI とパラメータ tf.idf なし

| 手法 | ARI | パラメータ |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.292994 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.299816 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.754128 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.754128 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.754128 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.267659 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.267659 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.267659 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.32: cacmcisi dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.778941 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.778941 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.858464 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.858464 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.858464 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.065465 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.065179 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.031167 | $\lambda = \{10^3\}, m = \{1 + 10^{-5}\}$ |

表 3.33: citeseer dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.613029 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.613467 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.815822 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.815822 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.815822 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.474599 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.474697 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.474697 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.34: classic dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.879056 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.879056 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.921936 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.921936 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.921936 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.462255 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.462271 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.462271 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.35: classic4 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0.000746 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.511101 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.510829 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.821297 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.821297 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.821297 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.473650 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.473650 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.473650 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.36: cora dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.582767 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.577253 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.762887 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.762887 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.762887 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.489489 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.489489 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.489489 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.37: cranmed dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 1.0 | $\lambda = \{10\}$ |
| BFCCM | 1.0 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 1.0 | $\lambda = \{10, 10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-2}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.998354 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.998354 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.998354 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.38: fbis dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.588115 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.588583 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.738970 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.738970 | $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.738970 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCs | 0.454868 | $\lambda = \{10^2\}$ |
| BFCS | 0.455623 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.457885 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.39: hitech dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.819096 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.815872 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.970478 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.970478 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.970478 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCs | 0.543066 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.543382 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.543382 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.40: k1a dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.973098 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.973098 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.982615 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.982615 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 0.982615 | $\lambda = \{10^3\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.842052 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.842052 | $m = \{1 + 10^{-5}, 1 + 10^{-5}\}$ |
| QFCS | 0.842052 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.41: k1b dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.937854 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.937854 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.948075 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.948075 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 0.948075 | $\lambda = \{10^3\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.846458 | $\lambda = \{10^2\}$ |
| BFCS | 0.831889 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.855643 | $\lambda = \{10^2\}, m = \{1 + 10^{-3}\}$ |

表 3.42: la1 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.955827 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.955061 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.985563 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.985563 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 0.985563 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.696012 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.696012 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.666012 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^5\}$ |

表 3.43: la2 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.931986 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.929407 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.979532 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.979532 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.979532 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCs | 0.721929 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.721929 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.721929 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.44: la12 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.882731 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.881583 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.964860 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.964860 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.964860 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCs | 0.663089 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.663153 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.663153 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.45: mm dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.945212 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.945212 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.998413 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.998413 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.998413 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.887479 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.887479 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.887479 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.46: new3 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.826857 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.825899 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.930602 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.930602 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.930602 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.519247 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.519247 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.519247 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.47: ohscal dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.456321 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.454839 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.756064 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.756064 | $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.756064 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.422681 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.422681 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.422681 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.48: re0 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.357040 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.357070 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.712491 | $\lambda = \{10^3\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.712491 | $m = \{1 + 10^{-4}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.712491 | $\lambda = \{10^3\}, m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.239827 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.243071 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.245803 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.49: `re1 dataset` の最高 ARI とパラメータ `tf.idfA`

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.775205 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.775205 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.849491 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.849491 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 0.849491 | $\lambda = \{10^3\}, m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.516297 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.516511 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.518583 | $\lambda = \{10^3\}, m = \{1 + 10^{-3}\}$ |

表 3.50: `reviews dataset` の最高 ARI とパラメータ `tf.idfA`

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.946783 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.946783 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.994732 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.994732 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.994732 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.715069 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.715499 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.715499 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.51: sports dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.934392 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.932660 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.977954 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.977954 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.977954 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.811925 | $\lambda = \{10^2\}$ |
| BFCS | 0.812071 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.807445 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |

表 3.52: tr11 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.981639 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.981639 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 1.0 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.873448 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.873448 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.873448 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.53: tr12 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.974436 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.974436 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\}$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| BFCCMM | 0.980433 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| KLFCS | 0.965572 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.965572 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |
| QFCS | 0.965572 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |

表 3.54: tr23 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0.005474 | $\lambda = \{10\}$ |
| BFCCM | 0.954420 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.954420 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.970179 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.970179 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.970179 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.714799 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}$ |
| BFCS | 0.714799 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |
| QFCS | 0.714799 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.55: tr31 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.950868 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.950868 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.981547 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.981547 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.981547 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.831028 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.832688 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.832688 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.56: tr41 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.972302 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.972302 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.984458 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.984458 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.984458 | $\lambda = \{10^3\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.901807 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}$ |
| BFCS | 0.901807 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |
| QFCS | 0.901807 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.57: tr45 dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.975641 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.975641 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.990437 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.990437 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.990437 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.906180 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.906180 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.906180 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.58: wap dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.977073 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.975461 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.984837 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.984837 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.984837 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.846868 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.846868 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.846868 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.59: webkb dataset の最高 ARI とパラメータ tf.idfA

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.452469 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.452469 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.810417 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.810417 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| QFCCMM | 0.810417 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.383982 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.386231 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.386231 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.60: cacmcisi dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.784261 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.784261 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.892475 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.892475 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.892475 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.027724 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.027724 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0 | $\lambda = \{0.5, 1.0, 10\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-1}\}$ |

表 3.61: citeseer dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.613457 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.611710 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.814505 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.814505 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.814505 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.463624 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.466704 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.466704 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.62: classic dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.884606 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.884606 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.930221 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.930221 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.930221 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.457151 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.458355 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.458625 | $\lambda = \{10^2\}, m = \{1 + 10^{-15}\}$ |

表 3.63: classic4 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0.000746 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.504125 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.504125 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.804391 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.804391 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.804391 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.468726 | $\lambda = \{10^3\}$ |
| BFCS | 0.468767 | $m = \{1 + 10^{-3}\}$ |
| QFCS | 0.468726 | $\lambda = \{10^3\}, m = \{1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.64: cora dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.577983 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.563768 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.761223 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.761223 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.761223 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.458978 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.458978 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.458978 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.65: cranmed dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 1.0 | $\lambda = \{10\}$ |
| BFCCM | 1.0 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 1.0 | $\lambda = \{10\}, m = \{1 + 10^{-5}, \dots, 1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-2}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10, 10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, \dots, 1 + 10^{-2}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.996709 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.996709 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.996709 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.66: fbis dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.608180 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.608180 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.732618 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.732618 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.732618 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.486375 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.486559 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.540674 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.67: hitech dataset の最高 ARI とパラメータ $tf.idfB$

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.824477 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.822401 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.975181 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.975181 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.975181 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.420789 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.420789 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.420789 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.68: k1a dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.973321 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.973321 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.983139 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.983139 | $m = \{1 + 10^{-15}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.983139 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.872057 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.872057 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.872057 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.69: k1b dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.947249 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.947249 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.954326 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.954326 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.954326 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.851167 | $\lambda = \{10^2\}$ |
| BFCS | 0.853199 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.854983 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.70: la1 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.956603 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.955727 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.990932 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.990932 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.990932 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.588466 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.588466 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.588466 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.71: 1a2 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.907764 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.906525 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.980613 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.980613 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.980613 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.608059 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.608059 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.623385 | $\lambda = \{10^2\}, m = \{1 + 10^{-5}\}$ |

表 3.72: 1a12 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.855851 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.852661 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.966958 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.966958 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.966958 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-15}\}$ |
| KLFCS | 0.567093 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.567093 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.567093 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.73: mm dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.929843 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.929843 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.908537 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.908537 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.908537 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.74: new3 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.830411 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.828033 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.944006 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.944006 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.943986 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.555790 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.555939 | $m = \{1 + 10^{-5}\}$ |
| QFCS | 0.555939 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.75: ohscal dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.434540 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.432073 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.748560 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.748560 | $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.748560 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.433619 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.433619 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.433619 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-15}\}$ |

表 3.76: re0 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.333506 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.332398 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.717821 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.717821 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.713674 | $\lambda = \{10^3\}, m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.256521 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.256521 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.256521 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.77: `re1 dataset` の最高 ARI とパラメータ `tf.idfB`

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.772153 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.767895 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.853335 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.853335 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.852029 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.572888 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.572888 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.572888 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.78: reviews dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.935865 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.934680 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.996465 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.996465 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.996465 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.759718 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.759718 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.759718 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.79: sports dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.930617 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.930361 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.977853 | $\lambda = \{10^3, 1.79769 \times 10^{308}\}, t = \{10^{-15}\}$ |
| BFCCMM | 0.977853 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-15}\}$ |
| QFCCMM | 0.977853 | $\lambda = \{10^3\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}\}$ |
| KLFCS | 0.587072 | $\lambda = \{10^2\}$ |
| BFCS | 0.590873 | $m = \{1 + 10^{-2}\}$ |
| QFCS | 0.692967 | $\lambda = \{10^2\}, m = \{1 + 10^{-2}\}$ |

表 3.80: tr11 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.993833 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.993833 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| KLFCS | 0.929941 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.929941 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.929941 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.81: tr12 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.965581 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.965581 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| BFCCMM | 0.980433 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| QFCCMM | 0.980433 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}, 10^{-1}\}$ |
| KLFCS | 0.947720 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.947720 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |
| QFCS | 0.947720 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\}$ |

表 3.82: tr23 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.969142 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.969142 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 1.0 | $\lambda = \{10^2, 10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 1.0 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 1.0 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}, 1 + 10^{-3}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| KLFCS | 0.939099 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.939099 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.939099 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.83: tr31 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.929030 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.929030 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.970143 | $\lambda = \{1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.970143 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.970143 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.736747 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.743767 | $m = \{1 + 10^{-4}\}$ |
| QFCS | 0.743767 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-4}\}$ |

表 3.84: tr41 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|--|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.974641 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.974641 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.986250 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.986250 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.986250 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.917612 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.917612 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.917612 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.85: tr45 dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.956969 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.956969 | $\lambda = \{10^2, 10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.997593 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| BFCCMM | 0.997593 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}, 10^{-2}\}$ |
| QFCCMM | 0.997593 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}, t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.912001 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.912001 | $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |
| QFCS | 0.912001 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}\}$ |

表 3.86: wap dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.974185 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.974185 | $\lambda = \{10^3\}, m = \{1 + 10^{-2}\}$ |
| KLFCCMM | 0.985715 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| BFCCMM | 0.985715 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-15}, 10^{-6}, 10^{-4}\}$ |
| QFCCMM | 0.985715 | $\lambda = \{10^3, 1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\},$ $t = \{10^{-6}, 10^{-4}\}$ |
| KLFCS | 0.871827 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.871827 | $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |
| QFCS | 0.871827 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}, 1 + 10^{-5}, 1 + 10^{-4}\}$ |

表 3.87: webkb dataset の最高 ARI とパラメータ tf.idfB

| Method | ARI | Parameter value |
|---------|----------|---|
| KLFCCM | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.444964 | $m = \{1 + 10^{-2}\}$ |
| QFCCM | 0.440257 | $\lambda = \{10^3\}, m = \{1 + 10^{-3}\}$ |
| KLFCCMM | 0.795540 | $\lambda = \{1.79769 \times 10^{308}\}, t = \{10^{-1}\}$ |
| BFCCMM | 0.795540 | $m = \{1 + 10^{-15}\}, t = \{10^{-1}\}$ |
| QFCCMM | 0.795540 | $\lambda = \{1.79769 \times 10^{308}\},$ $m = \{1 + 10^{-15}\}, t = \{10^{-1}\}$ |
| KLFCS | 0.387504 | $\lambda = \{1.79769 \times 10^{308}\}$ |
| BFCS | 0.397504 | $m = \{1 + 10^{-15}\}$ |
| QFCS | 0.392078 | $\lambda = \{1.79769 \times 10^{308}\}, m = \{1 + 10^{-5}\}$ |

表 3.88: 初期値に正解のファジィな帰属度を与えた最高 ARI tfidf なし

| 手法 データ | KLFCM | BFCCM | QFCCM | KLFCMM | BFCCMM | QFCCMM | KLFC | BFCS | QFCS |
|-----------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|
| cacmcisi | 0 | 0.719538 | 0.719538 | 0.809621 | 0.809621 | 0.809621 | 0 | -0.027071 | 0 |
| citepeer | 0 | 0.555670 | 0.555283 | 0.781548 | 0.781548 | 0.781548 | 0.365114 | 0.365114 | 0.365114 |
| classic | 0 | 0.843743 | 0.842978 | 0.892567 | 0.892567 | 0.892567 | 0.411013 | 0.410876 | 0.418826 |
| classic4 | 0.000746 | 0.487289 | 0.487289 | 0.783449 | 0.782738 | 0.782423 | 0.441396 | 0.442482 | 0.447722 |
| cora | 0.400710 | 0.558195 | 0.558224 | 0.745177 | 0.745177 | 0.745177 | 0.181326 | 0.182860 | 0.182860 |
| cranmed | 0.978704 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.957636 | 0.957636 | 0.957636 |
| fbis | 0 | 0.548472 | 0.561336 | 0.727048 | 0.727048 | 0.727048 | 0.465904 | 0.466871 | 0.466871 |
| hitech | 0 | 0.500573 | 0.498074 | 0.928259 | 0.928259 | 0.928259 | 0.279298 | 0.279298 | 0.279298 |
| kia | 0 | 0.951056 | 0.951056 | 0.976298 | 0.976298 | 0.976298 | 0.631614 | 0.632257 | 0.632257 |
| kib | 0.802794 | 0.894707 | 0.894707 | 0.930587 | 0.930587 | 0.930587 | 0.815651 | 0.815651 | 0.815651 |
| la1 | 0 | 0.876117 | 0.875392 | 0.972516 | 0.972516 | 0.972516 | 0.296463 | 0.296463 | 0.296463 |
| la2 | 0 | 0.670654 | 0.669151 | 0.955918 | 0.955918 | 0.955918 | 0.336259 | 0.336259 | 0.336259 |
| la12 | 0 | 0.590707 | 0.586742 | 0.932916 | 0.932916 | 0.932916 | 0.290656 | 0.287174 | 0.290867 |
| mm | 0 | 0.913082 | 0.908538 | 0.996828 | 0.996828 | 0.996828 | 0.866667 | 0.866667 | 0.866667 |
| new3 | 0 | 0.711023 | 0.702324 | 0.885880 | 0.885880 | 0.885880 | 0.457092 | 0.456898 | 0.457092 |
| ohscal | 0.358778 | 0.396097 | 0.395148 | 0.711620 | 0.711786 | 0.711786 | 0.229789 | 0.229587 | 0.229789 |
| re0 | 0 | 0.260616 | 0.281009 | 0.624956 | 0.624781 | 0.624299 | 0.292337 | 0.301650 | 0.301444 |
| re1 | 0 | 0.801896 | 0.799243 | 0.837132 | 0.837132 | 0.837132 | 0.352648 | 0.355002 | 0.355002 |
| reviews | 0 | 0.887301 | 0.887301 | 0.986761 | 0.986761 | 0.986761 | 0.700147 | 0.700464 | 0.700451 |
| sports | 0 | 0.892871 | 0.891601 | 0.967107 | 0.967107 | 0.967107 | 0.431274 | 0.429228 | 0.490517 |
| tr11 | 0.671194 | 0.930590 | 0.930590 | 0.987652 | 0.987652 | 0.987652 | 0.820268 | 0.820268 | 0.820268 |
| tr12 | 0 | 0.942499 | 0.942499 | 0.980433 | 0.980433 | 0.980433 | 0.695190 | 0.695190 | 0.695190 |
| tr23 | 0.382848 | 0.942754 | 0.942754 | 0.954420 | 0.954420 | 0.954420 | 0.296111 | 0.305340 | 0.305340 |
| tr31 | 0.766094 | 0.913056 | 0.913056 | 0.971071 | 0.971071 | 0.971071 | 0.549255 | 0.554192 | 0.554192 |
| tr41 | 0 | 0.951001 | 0.951001 | 0.976203 | 0.976203 | 0.976203 | 0.602412 | 0.602412 | 0.602412 |
| tr45 | 0 | 0.941595 | 0.941595 | 0.979758 | 0.979758 | 0.979758 | 0.503606 | 0.504575 | 0.504575 |
| wap | 0 | 0.941963 | 0.940369 | 0.980827 | 0.980827 | 0.980827 | 0.628756 | 0.626042 | 0.628756 |
| webkb | 0 | 0.292994 | 0.299816 | 0.754128 | 0.754128 | 0.754128 | 0.267659 | 0.267659 | 0.267659 |

表 3.89: 初期値に正解のファジィな帰属度を与えた最高 ARI tf.idfA

| データ | 手法 | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|----------|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | | | | | | | | | | |
| cacmcisi | | 0 | 0.778941 | 0.778941 | 0.858464 | 0.858464 | 0.858464 | 0.065465 | 0.065179 | 0.031167 |
| citepeer | | 0 | 0.613029 | 0.613467 | 0.815822 | 0.815822 | 0.815822 | 0.474599 | 0.474697 | 0.474697 |
| classic | | 0 | 0.879056 | 0.879056 | 0.921936 | 0.921936 | 0.921936 | 0.462255 | 0.462271 | 0.462271 |
| classic4 | | 0.000746 | 0.511101 | 0.510829 | 0.821297 | 0.821297 | 0.821297 | 0.473650 | 0.473650 | 0.473650 |
| cora | | 0 | 0.582767 | 0.577253 | 0.762887 | 0.762887 | 0.762887 | 0.489489 | 0.489489 | 0.489489 |
| cranmed | | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.998354 | 0.998354 | 0.998354 |
| fbis | | 0 | 0.588115 | 0.588583 | 0.738970 | 0.738970 | 0.738970 | 0.454868 | 0.455623 | 0.457885 |
| hitech | | 0 | 0.819096 | 0.815872 | 0.970478 | 0.970478 | 0.970478 | 0.543066 | 0.543382 | 0.543382 |
| kia | | 0 | 0.973098 | 0.973098 | 0.982615 | 0.982615 | 0.982615 | 0.842052 | 0.842052 | 0.842052 |
| kib | | 0 | 0.937854 | 0.937854 | 0.948075 | 0.948075 | 0.948075 | 0.846458 | 0.831889 | 0.855643 |
| la1 | | 0 | 0.955827 | 0.955061 | 0.985563 | 0.985563 | 0.985563 | 0.666012 | 0.696012 | 0.666012 |
| la2 | | 0 | 0.931986 | 0.929407 | 0.979532 | 0.979532 | 0.979532 | 0.721929 | 0.721929 | 0.721929 |
| la12 | | 0 | 0.882731 | 0.881583 | 0.964860 | 0.964860 | 0.964860 | 0.663089 | 0.663153 | 0.663153 |
| mm | | 0 | 0.945212 | 0.945212 | 0.998413 | 0.998413 | 0.998413 | 0.887479 | 0.887479 | 0.887479 |
| new3 | | 0 | 0.826857 | 0.825899 | 0.930602 | 0.930602 | 0.930602 | 0.519247 | 0.519247 | 0.519247 |
| ohscal | | 0 | 0.456321 | 0.454839 | 0.756064 | 0.756064 | 0.756064 | 0.422681 | 0.422681 | 0.422681 |
| re0 | | 0 | 0.357040 | 0.357070 | 0.712491 | 0.712491 | 0.712491 | 0.239827 | 0.243071 | 0.245803 |
| re1 | | 0 | 0.775205 | 0.775205 | 0.849491 | 0.849491 | 0.849491 | 0.516297 | 0.516511 | 0.518583 |
| reviews | | 0 | 0.946783 | 0.946783 | 0.994732 | 0.994732 | 0.994732 | 0.715069 | 0.715499 | 0.715499 |
| sports | | 0 | 0.934392 | 0.932660 | 0.977954 | 0.977954 | 0.977954 | 0.811925 | 0.812071 | 0.807445 |
| tr11 | | 0 | 0.981639 | 0.981639 | 1.0 | 1.0 | 1.0 | 0.873448 | 0.873448 | 0.873448 |
| tr12 | | 0 | 0.974436 | 0.974436 | 0.980433 | 0.980433 | 0.980433 | 0.965572 | 0.965572 | 0.965572 |
| tr23 | | 0.005474 | 0.954420 | 0.954420 | 0.970179 | 0.970179 | 0.970179 | 0.714799 | 0.714799 | 0.714799 |
| tr31 | | 0 | 0.950868 | 0.950868 | 0.981547 | 0.981547 | 0.981547 | 0.831028 | 0.832688 | 0.832688 |
| tr41 | | 0 | 0.972302 | 0.972302 | 0.984458 | 0.984458 | 0.984458 | 0.901807 | 0.901807 | 0.901807 |
| tr45 | | 0 | 0.975641 | 0.975641 | 0.990437 | 0.990437 | 0.990437 | 0.906180 | 0.906180 | 0.906180 |
| wap | | 0 | 0.977073 | 0.975461 | 0.984837 | 0.984837 | 0.984837 | 0.846868 | 0.846868 | 0.846868 |
| webkb | | 0 | 0.452469 | 0.452469 | 0.810417 | 0.810417 | 0.810417 | 0.383982 | 0.386231 | 0.386231 |

表 3.90: 初期値に正解のファジィな帰属度を与えた最高 ARI tf.idfB

| データ | 手法 | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | | | | | | | | | | |
| cacmcisi | 0 | 0.784261 | 0.784261 | 0.784261 | 0.892475 | 0.892475 | 0.892475 | 0.027724 | 0.027724 | 0 |
| citeseer | 0 | 0.613457 | 0.611710 | 0.611710 | 0.814505 | 0.814505 | 0.814505 | 0.463624 | 0.466704 | 0.466704 |
| classic | 0 | 0.884606 | 0.884606 | 0.884606 | 0.930221 | 0.930221 | 0.930221 | 0.457151 | 0.458355 | 0.458625 |
| classic4 | 0.000746 | 0.504125 | 0.504125 | 0.504125 | 0.804391 | 0.804391 | 0.804391 | 0.468726 | 0.468767 | 0.468726 |
| cora | 0 | 0.577983 | 0.563768 | 0.563768 | 0.761223 | 0.761223 | 0.761223 | 0.458978 | 0.458978 | 0.458978 |
| cranmed | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.996709 | 0.996709 | 0.996709 |
| fbis | 0 | 0.608180 | 0.608180 | 0.608180 | 0.732618 | 0.732618 | 0.732618 | 0.486375 | 0.486559 | 0.540674 |
| hitech | 0 | 0.824477 | 0.822401 | 0.822401 | 0.975181 | 0.975181 | 0.975181 | 0.420789 | 0.420789 | 0.420789 |
| kla | 0 | 0.973321 | 0.973321 | 0.973321 | 0.983139 | 0.983139 | 0.983139 | 0.872057 | 0.872057 | 0.872057 |
| kib | 0 | 0.947249 | 0.947249 | 0.947249 | 0.954326 | 0.954326 | 0.954326 | 0.851167 | 0.853199 | 0.854983 |
| la1 | 0 | 0.956603 | 0.955727 | 0.955727 | 0.990932 | 0.990932 | 0.990932 | 0.588466 | 0.588466 | 0.588466 |
| la2 | 0 | 0.907764 | 0.906525 | 0.906525 | 0.980613 | 0.980613 | 0.980613 | 0.608059 | 0.608059 | 0.623385 |
| la12 | 0 | 0.855851 | 0.852661 | 0.852661 | 0.966958 | 0.966958 | 0.966958 | 0.567093 | 0.567093 | 0.567093 |
| mm | 0 | 0.929843 | 0.929843 | 0.929843 | 1.0 | 1.0 | 1.0 | 0.908537 | 0.908537 | 0.908537 |
| new3 | 0 | 0.830411 | 0.828033 | 0.828033 | 0.944006 | 0.944006 | 0.943986 | 0.555790 | 0.555939 | 0.555939 |
| ohscal | 0 | 0.434540 | 0.432073 | 0.432073 | 0.748560 | 0.748560 | 0.748560 | 0.433619 | 0.433619 | 0.433619 |
| re0 | 0 | 0.333506 | 0.332398 | 0.332398 | 0.717821 | 0.717821 | 0.713674 | 0.256521 | 0.256521 | 0.256521 |
| re1 | 0 | 0.772153 | 0.767895 | 0.767895 | 0.853335 | 0.853335 | 0.852029 | 0.572888 | 0.572888 | 0.572888 |
| reviews | 0 | 0.935865 | 0.934680 | 0.934680 | 0.996465 | 0.996465 | 0.996465 | 0.759718 | 0.759718 | 0.759718 |
| sports | 0 | 0.930617 | 0.930361 | 0.930361 | 0.977853 | 0.977853 | 0.977853 | 0.587072 | 0.590873 | 0.692967 |
| tr11 | 0 | 0.993833 | 0.993833 | 0.993833 | 1.0 | 1.0 | 1.0 | 0.929941 | 0.929941 | 0.929941 |
| tr12 | 0 | 0.965581 | 0.965581 | 0.965581 | 0.980433 | 0.980433 | 0.980433 | 0.947720 | 0.947720 | 0.947720 |
| tr23 | 0 | 0.969142 | 0.969142 | 0.969142 | 1.0 | 1.0 | 1.0 | 0.939099 | 0.939099 | 0.939099 |
| tr31 | 0 | 0.929030 | 0.929030 | 0.929030 | 0.970143 | 0.970143 | 0.970143 | 0.736747 | 0.743767 | 0.743767 |
| tr41 | 0 | 0.974641 | 0.974641 | 0.974641 | 0.986250 | 0.986250 | 0.986250 | 0.917612 | 0.917612 | 0.917612 |
| tr45 | 0 | 0.956969 | 0.956969 | 0.956969 | 0.997593 | 0.997593 | 0.997593 | 0.912001 | 0.912001 | 0.912001 |
| wap | 0 | 0.974185 | 0.974185 | 0.974185 | 0.985715 | 0.985715 | 0.985715 | 0.871827 | 0.871827 | 0.871827 |
| webkb | 0 | 0.444964 | 0.440257 | 0.440257 | 0.795540 | 0.795540 | 0.795540 | 0.387504 | 0.392078 | 0.392078 |

3.5.1 数値実験 1 の結果

表 3.91 に各ファジィクラスタリング手法において最高 ARI を達成したデータ数を示す。tf.idf なし, tf.idfA, tf.idfB ごとで最高 ARI を達成したデータ数が一番多い手法のセルを赤色で強調する。数値実験 1 から FCS や FCCM と比べて FCCMM の三手法の精度が高く、過半数のデータにおいてクリスプに分けるようなパラメータを設定することで最高 ARI を達成し、その値は同値となることを確認した。tf.idf なしの場合, classic4 と re0 では KLFCCMM, ohscal では BFCCMM と QFCCMM, re1 では QFCCMM を適用するとどの手法よりも高い ARI を算出した。tf.idfA を施した場合, FCCMM の三つ手法は、全てのデータにおいて同値の最高 ARI を達成した。tf.idfB を施した場合, KLFCCMM と BFCCMM では、全てのデータにおいて同値の最高 ARI を達成したが, new3, re0, re1 を QFCCMM に適用すると他二手法より精度の低い結果となった。しかし数値実験 1 では, FCCMM 三手法ともに互角の精度で一概に優位性は比較できない。そこで、初期値依存性による精度評価をするため、初期値にランダムなクラスタ中心を設定する数値実験 2 を行う。

表 3.91: 最高 ARI を達成したデータ数

| データ 手法 | tf.idf なし | tf.idfA | tf.idfB |
|-----------|-----------|---------|---------|
| KLFCCM | 0 | 1 | 1 |
| BFCCM | 1 | 1 | 1 |
| QFCCM | 1 | 1 | 1 |
| KLFCCMM | 26 | 28 | 28 |
| BFCCMM | 25 | 28 | 28 |
| QFCCMM | 26 | 28 | 25 |
| KLFCS | 0 | 0 | 0 |
| BFCS | 0 | 0 | 0 |
| QFCS | 0 | 0 | 0 |

3.6 数値実験 2

初期値として k -means++ のアルゴリズムにおける平方ユークリッド距離の代用として各ファジィクラスタリング手法の非類似度を適用し、ランダムにオブジェクトを選びクラスタ中心へ代入する。初期クラスタサイズ調整変数に $1/\text{正解クラス数}$ を与え、パラメータは数値実験 1 と同様に表 3.3 の範囲を採用し、一組のパラメータにつき 10 通りの初期設定を行う。収束前に計算限界を迎えた場合、初期クラスタ中心を選択する乱数シード値をインクリメントすることで新たな初期値を与える。10 通りのクラスタリング結果を得る前に、10 回計算限界を迎えた場合、そのパラメータでの処理を終える。また、10 通りのうち 1 回でも更新回数が 10^4 を上回るとき、そのパラメータでの処理を終える。数値実験 1 同様、収束前に更新を終えたパラメータの結果を反映しない。データ規模が膨大な new3 は、処理時間の観点から数値実験 2 の実験データに含めない。以下に数値実験 2 の流れを示す：

実験の流れ 2 (数値実験 2).

- STEP 1. new3 を除いた表 3.1 に示すデータの読み込み，正解の帰属度の読み込みをする。
- STEP 2. tf.idf なしでない場合，式 (3.5)，式 (3.6) に従い tf.idfA や tf.idfB に変換する。
- STEP 3. {B,KL,Q}FCS では式 (3.8) に従い球面化し，{B,KL,Q}FC{CM,CMM} では式 (3.9) に従い単体化する。
- STEP 4. クラスタ数を正解クラス数に設定する。
- STEP 5. 表 3.3 の範囲でパラメータを設定する。
- STEP 6. *k*-means++ における平方ユークリッド距離の代わりにオブジェクト-クラスタ間非類似度を {B,KL,Q}FCS では式 (2.78)，{KL,Q,B}FCCM では式 (2.128)，{KL,B,Q}FCCMM では式 (2.172) に置き換えたアルゴリズム 3 に従い初期クラスタ中心を設定する。
- STEP 7. 初期クラスタサイズ調整変数を $1/\text{クラスタ数}$ に設定する。
- STEP 8. {B,KL,Q}FCS では式 (2.78)，{KL,Q,B}FCCM では式 (2.128)，{KL,B,Q}FCCMM では式 (2.172) に従いオブジェクト-クラスタ間非類似度を更新する。
- STEP 9. BFCS では式 (2.88)，KLFCFS では式 (2.102)，QFCS では式 (2.116)，KLFFCCM では式 (2.138)，BFCCM では式 (2.166)，QFCCM では式 (2.152)，KLFFCCMM では式 (2.182)，BFCCMM では式 (2.196)，QFCCMM では式 (2.210) に従い帰属度を更新する。
- STEP 10. BFCS では式 (2.90)，KLFCFS では式 (2.104)，QFCS では式 (2.118)，KLFFCCM では式 (2.140)，BFCCM では式 (2.168)，QFCCM では式 (2.154)，KLFFCCMM では式 (2.184)，BFCCMM では式 (2.198)，QFCCMM では式 (2.212) に従いクラスタ中心を更新する。
- STEP 11. BFCS では式 (2.92)，KLFCFS では式 (2.106)，QFCS では式 (2.120)，KLFFCCM では式 (2.142)，BFCCM では式 (2.170)，QFCCM では式 (2.156)，KLFFCCMM では式 (2.186)，BFCCMM では式 (2.200)，QFCCMM では式 (2.214) に従いクラスタサイズ調整変数を更新する。
- STEP 12. 更新前と更新後の帰属度，クラスタ中心，クラスタサイズ調整変数の最大値ノルムを計算し，それらの和が 10^{-6} 以上であれば STEP 8. へ，計算限界を迎えたら初期値設定乱数シードをインクリメントし STEP 6. へ，10 回計算限界を迎えたもしくは更新数が 10^4 を超えたらそのパラメータでのクラスタリングを不可として STEP 17. へ。
- STEP 13. 収束した帰属度をクリスプ化する。
- STEP 14. クリスプ化された帰属度と正解の帰属度を照らし合わせ Contingency Table を得る。
- STEP 15. Contingency Table から式 (3.7) に従い ARI を算出する。
- STEP 16. 10 通りの初期設定を行えたなら平均 ARI と標準偏差を算出する。そうでなければ STEP 6. へ。
- STEP 17. 次に設定すべきパラメータがあれば STEP 5. へ。
- STEP 18. 次に処理すべきデータ変換手法があれば STEP 2. へ。
- STEP 19. 次に処理すべきデータがあれば STEP 1. へ，そうでなければ終了する。 ■

表 3.3 のパラメータ範囲で結果を得られなかった手法は“—”と表記する。tf.idf なしを適用したクラスタリングにおける 10 通りの初期設定による平均 ARI，標準偏差とそれらを達成したパラメータを表 3.92 から表 3.118 にまとめた。tf.idfA を適用したクラスタリングにおける 10 通りの初期設定による平均 ARI，標準偏差とそれらを達成したパラメータを表 3.119 から表 3.145 にまとめた。tf.idfB を適用したクラスタリングにおける 10 通りの初期設定による平均 ARI，標準偏差とそれらを達成したパラメータを表 3.146 から表 3.172 にまとめた。データ対ファジィクラスタリング手法で最高平均 ARI を表 3.173，表 3.174，表 3.175 にまとめた。図 3.1，図 3.2，図 3.3 にそれぞれのデータ形式における

最高平均 ARI とその標準偏差を示す。

表 3.92: `cacmcisi dataset` の 10 通りの初期値による ARI とパラメータ `tf.idf` なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|-----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.133791 | 0.142445 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.151426 | 0.054539 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.252963 | 0.205589 | $(\lambda, t) = (10^3, 10^{-1})$ |
| BFCCMM | 0.253210 | 0.200353 | $(m, t) = (1 + 10^{-3}, 10^{-1})$ |
| QFCCMM | 0.254576 | 0.206241 | $(\lambda, m, t) = (10^3, 1 + 10^{-4}, 10^{-1})$ |
| KLFCs | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCS | -0.029153 | 0.001708 | $m = 1 + 10^{-5}$ |
| QFCS | 0 | 0 | $(\lambda, m) = (\{0.5, 1.0\}, \{1 + 10^{-5}, \dots, 1 + 10^{-1}\})$ |

表 3.93: `citeseer dataset` の 10 通りの初期値による ARI とパラメータ `tf.idf` なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.154531 | 0.034583 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.229337 | 0.049565 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.292533 | 0.047457 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.354346 | 0.047067 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.364302 | 0.050252 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.240857 | 0.041188 | $\lambda = 10^2$ |
| BFCS | 0.243215 | 0.036810 | $m = 1 + 10^{-4}$ |
| QFCS | 0.239188 | 0.041772 | $(\lambda, m) = (10^2, 1 + 10^{-15})$ |

表 3.94: classic dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.197311 | 0.091022 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.491005 | 0.060964 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.511592 | 0.016839 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.505424 | 0.043198 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.505398 | 0.042161 | $(\lambda, m, t) = (10^3, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.303158 | 0.048844 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | 0.310171 | 0.030906 | $m = 1 + 10^{-5}$ |
| QFCS | 0.294757 | 0.047942 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.95: classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|-----------|--|
| KLFCCM | 0.000746 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.412508 | 0.0800443 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.435832 | 0.0614389 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.490168 | 0.0137711 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.491468 | 0.0138263 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.491696 | 0.0138915 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.436888 | 0.0055927 | $\lambda = 10^2$ |
| BFCS | 0.433474 | 0.0174358 | $m = 1 + 10^{-2}$ |
| QFCS | 0.436886 | 0.0060191 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.96: cora dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.097231 | 0.033503 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.108148 | 0.037309 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.165854 | 0.041013 | $(\lambda, t) = (10, 10^{-2})$ |
| BFCCMM | 0.241122 | 0.055386 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.248180 | 0.051979 | $(\lambda, m, t) = (10^3, 1 + 10^{-2}, 10^{-1})$ |
| KLFCS | 0.127379 | 0.018520 | $\lambda = 10^2$ |
| BFCS | 0.137206 | 0.018773 | $m = 1 + 10^{-2}$ |
| QFCS | 0.126789 | 0.018351 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.97: cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0.871431 | 0.290635 | $\lambda = 10$ |
| BFCCM | 0.247833 | 0.190689 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.874819 | 0.291640 | $(\lambda, m) = (10, 1 + 10^{-15})$ |
| KLFCCMM | 0.991292 | 0.002436 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.983440 | 0.004699 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.993259 | 0.000884 | $(\lambda, m, t) = (10, 1 + 10^{-5}, 10^{-1})$ |
| KLFCS | 0.951199 | 0 | $\lambda = 10^2$ |
| BFCS | 0.951200 | 0 | $m = 1 + 10^{-2}$ |
| QFCS | 0.951200 | 0 | $(\lambda, m) = (10^2, 1 + 10^{-5})$ |

表 3.98: fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.430863 | 0.040651 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.424993 | 0.041320 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.488140 | 0.008470 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.461586 | 0.028190 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.490733 | 0.010797 | $(\lambda, m, t) = (10, 1 + 10^{-5}, 10^{-1})$ |
| KLFCS | 0.382091 | 0.024479 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.389150 | 0.021737 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.99: hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.175382 | 0.063880 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.194919 | 0.029058 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.270307 | 0.018316 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.274295 | 0.016691 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.274295 | 0.016691 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.265081 | 0.018546 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.264473 | 0.025150 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.100: k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.271709 | 0.100217 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.397914 | 0.061402 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.296765 | 0.042723 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.387499 | 0.049736 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.400672 | 0.046311 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCS | 0.275088 | 0.055997 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.263366 | 0.050494 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.101: k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.472567 | 0.184789 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.603584 | 0.147974 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.731733 | 0.091512 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.509426 | 0.108477 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.769328 | 0.093435 | $(\lambda, m, t) = (10, 1 + 10^{-3}, 10^{-1})$ |
| KLFCS | 0.509334 | 0.105984 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.528982 | 0.095244 | $(\lambda, m) = (10^2, 1 + 10^{-4})$ |

表 3.102: la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.214372 | 0.129540 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.403241 | 0.089426 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.501167 | 0.069939 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.521592 | 0.077635 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.534185 | 0.097557 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCS | 0.211776 | 0.028805 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.211880 | 0.028160 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.103: la2 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.309631 | 0.119197 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.337024 | 0.049826 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.470718 | 0.072788 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.482179 | 0.061800 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.488718 | 0.085481 | $(\lambda, m, t) = (10, 1 + 10^{-4}, 10^{-1})$ |
| KLFCS | 0.194797 | 0.020687 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.194815 | 0.020628 | $(\lambda, m) = (10^3, 1 + 10^{-4})$ |

表 3.104: la12 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.295090 | 0.156368 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.414950 | 0.091309 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.513915 | 0.072687 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.540250 | 0.081672 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.538959 | 0.082150 | $(\lambda, m, t) = (10^3, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.218632 | 0.042312 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.375149 | 0.092069 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.105: mm dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.308563 | 0.326837 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.312662 | 0.332486 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.812183 | 0.033988 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.820648 | 0.026322 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.820648 | 0.026322 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.776885 | 0.242812 | $\lambda = 10^2$ |
| BFCS | 0.007451 | 0.015776 | $m = 1 + 10^{-2}$ |
| QFCS | 0.776806 | 0.243051 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.106: ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0.295765 | 0.012326 | $\lambda = 10$ |
| BFCCM | 0.269983 | 0.011860 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.305045 | 0.006757 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.305922 | 0.010843 | $(\lambda, t) = (10, 10^{-2})$ |
| BFCCMM | 0.312164 | 0.011037 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.312144 | 0.012411 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.212066 | 0.009993 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.212288 | 0.009957 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.107: re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.082979 | 0.030113 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.136902 | 0.021253 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.170854 | 0.046396 | $(\lambda, t) = (10^2, 10^{-1})$ |
| BFCCMM | 0.172366 | 0.036398 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.172887 | 0.036420 | $(\lambda, m, t) = (10^3, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.169197 | 0.020895 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.180637 | 0.019906 | $(\lambda, m) = (10^2, 1 + 10^{-1})$ |

表 3.108: `rel dataset` の 10 通りの初期値による ARI とパラメータ `tf.idf` なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|-----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.096775 | 0.0156079 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.168108 | 0.0146363 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.226035 | 0.0140838 | $(\lambda, t) = (10^3, 0.5)$ |
| BFCCMM | 0.224395 | 0.0139157 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.226700 | 0.0160509 | $(\lambda, m, t) = (10^3, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.243798 | 0.0197782 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.247931 | 0.0198464 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.109: `reviews dataset` の 10 通りの初期値による ARI とパラメータ `tf.idf` なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.276202 | 0.158563 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.379397 | 0.120894 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.538596 | 0.050381 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.576558 | 0.043695 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.576558 | 0.043695 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.564279 | 0.111796 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.563625 | 0.112208 | $(\lambda, m) = (10^2, 1 + 10^{-5})$ |

表 3.110: sports dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.417823 | 0.083963 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.425279 | 0.079985 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.672250 | 0.087956 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.588483 | 0.100321 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.681617 | 0.059962 | $(\lambda, m, t) = (10, 1 + 10^{-4}, 10^{-1})$ |
| KLFCs | 0.309572 | 0.063869 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.309393 | 0.063973 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.111: tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.237391 | 0.103692 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.477699 | 0.103590 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.555134 | 0.088219 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.430620 | 0.149439 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.540338 | 0.136925 | $(\lambda, m, t) = (10, 1 + 10^{-4}, 10^{-1})$ |
| KLFCs | 0.439026 | 0.061130 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.432677 | 0.062738 | $(\lambda, m) = (10^3, 1 + 10^{-4})$ |

表 3.112: tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.176235 | 0.055910 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.306292 | 0.084528 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.365259 | 0.134997 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.368599 | 0.050131 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.368599 | 0.050131 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.382425 | 0.090190 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.375149 | 0.092069 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.113: tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.076868 | 0.115883 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.150237 | 0.076628 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.149527 | 0.066085 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.156809 | 0.050416 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.164993 | 0.082889 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.187104 | 0.034146 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.190638 | 0.042019 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.114: tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.325699 | 0.162885 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.489907 | 0.134557 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.615231 | 0.078339 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.577080 | 0.096805 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.628113 | 0.088658 | $(\lambda, m, t) = (10, 1 + 10^{-4}, 10^{-1})$ |
| KLFCs | 0.360381 | 0.052713 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.384971 | 0.073444 | $(\lambda, m) = (10^3, 1 + 10^{-1})$ |

表 3.115: tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.272910 | 0.077497 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.566572 | 0.069308 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.475561 | 0.058068 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.450730 | 0.093135 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.493992 | 0.039828 | $(\lambda, m, t) = (10, 1 + 10^{-4}, 10^{-1})$ |
| KLFCs | 0.371022 | 0.037765 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.369899 | 0.034944 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.116: tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.184500 | 0.089431 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.362001 | 0.051644 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.428516 | 0.089482 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.488709 | 0.044158 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.498007 | 0.052043 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.403686 | 0.100455 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.403126 | 0.100038 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.117: wap dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.265345 | 0.115918 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.326622 | 0.093091 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.266195 | 0.084345 | $(\lambda, t) = (10, 10^{-2})$ |
| BFCCMM | 0.331515 | 0.032726 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.353214 | 0.028493 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.261089 | 0.050942 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.262806 | 0.052256 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.118: webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idf なし

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.046977 | 0.088163 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.083637 | 0.078561 | $(\lambda, m) = (10, 1 + 10^{-15})$ |
| KLFCCMM | 0.224488 | 0.060421 | $(\lambda, t) = (10^2, 10^{-1})$ |
| BFCCMM | 0.293891 | 0.024005 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.297123 | 0.029493 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.204766 | 0.050574 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.203136 | 0.049414 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.119: cacmcisi dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.109125 | 0.142675 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.136767 | 0.096869 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.284956 | 0.167247 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.284956 | 0.167247 | $(m, t) = (1 + 10^{-15}, 0.5)$ |
| QFCCMM | 0.284956 | 0.167247 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-15}, 0.5)$ |
| KLFCs | 0.042454 | 0.011666 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | 0.034512 | 0.006331 | $m = 1 + 10^{-5}$ |
| QFCS | 0.034512 | 0.006331 | $(\lambda, m) = (1.79769 \times 10^{308}, 1 + 10^{-5})$ |

表 3.120: citeseer dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.090936 | 0.022432 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.210049 | 0.039664 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.298762 | 0.057549 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.381185 | 0.032282 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.381185 | 0.032282 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCS | 0.337738 | 0.046566 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.348793 | 0.045092 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.121: classic dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.160741 | 0.069047 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.394141 | 0.063778 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.513934 | 0.041412 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.515696 | 0.129654 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.515696 | 0.129654 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.374927 | 0.063872 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.376880 | 0.053666 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.122: classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0.000746 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.370652 | 0.090139 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.450194 | 0.038533 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.545673 | 0.052976 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.546794 | 0.051637 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.546794 | 0.051637 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.475238 | 0.013449 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.473808 | 0.007226 | $(\lambda, m) = (10^3, 1 + 10^{-4})$ |

表 3.123: cora dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.062849 | 0.013675 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.068871 | 0.021855 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.256137 | 0.048074 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.296871 | 0.030356 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.296871 | 0.030356 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCS | 0.274829 | 0.029142 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.279529 | 0.028608 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.124: cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0.861896 | 0.287454 | $\lambda = 10$ |
| BFCCM | 0.135181 | 0.098631 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.865918 | 0.288716 | $(\lambda, m) = (10, 1 + 10^{-4})$ |
| KLFCCMM | 0.983774 | 0.006983 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.960216 | 0.002907 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.996216 | 0.001052 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.991948 | 0.002482 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.992112 | 0.002411 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.125: fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0.085782 | 0.171568 | $\lambda = 10$ |
| BFCCM | 0.409896 | 0.043879 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.440917 | 0.045815 | $(\lambda, m) = (10, 1 + 10^{-15})$ |
| KLFCCMM | 0.475528 | 0.019528 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.433428 | 0.048637 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.484018 | 0.017647 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.391648 | 0.029260 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.395691 | 0.034307 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.126: hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.080154 | 0.042163 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.159716 | 0.034360 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.225428 | 0.018205 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.239141 | 0.017807 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.239141 | 0.017807 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.232820 | 0.022457 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.234890 | 0.025983 | $(\lambda, m) = (10^2, 1 + 10^{-15})$ |

表 3.127: k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.134708 | 0.033931 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.321256 | 0.051806 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.343359 | 0.056913 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.347930 | 0.040069 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.360423 | 0.077779 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.395083 | 0.054486 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.391710 | 0.071136 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.128: k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.171408 | 0.147967 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.353940 | 0.036440 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.441510 | 0.095274 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.408841 | 0.099533 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.743200 | 0.082768 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCS | 0.555370 | 0.093793 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.698865 | 0.051878 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.129: la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.094951 | 0.048768 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.344660 | 0.064683 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.446305 | 0.076325 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.481197 | 0.062403 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.481197 | 0.062403 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.483105 | 0.060481 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.480119 | 0.074935 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.130: la2 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.174640 | 0.052787 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.360006 | 0.086490 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.482917 | 0.086663 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.478598 | 0.028815 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.478598 | 0.028815 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.478404 | 0.059308 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.474630 | 0.059605 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.131: la12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.247759 | 0.071942 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.431583 | 0.066262 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.512399 | 0.069645 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.521798 | 0.074940 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.527401 | 0.075789 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.471867 | 0.058729 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.477453 | 0.062780 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.132: mm dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.080518 | 0.209036 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.083956 | 0.233156 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.553743 | 0.252690 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.612177 | 0.220762 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.612177 | 0.220762 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.192041 | 0.257752 | $\lambda = 10^3$ |
| BFCS | 0.225095 | 0.232140 | $m = 1 + 10^{-3}$ |
| QFCS | 0.222682 | 0.226075 | $(\lambda, m) = (1.79769 \times 10^{308}, 1 + 10^{-5})$ |

表 3.133: ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.209935 | 0.021364 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.278464 | 0.012973 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.331802 | 0.020392 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.336163 | 0.022972 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.336163 | 0.022972 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.347707 | 0.020842 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.347847 | 0.020875 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.134: re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.094556 | 0.020183 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.154778 | 0.024797 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.158646 | 0.017841 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.159402 | 0.017329 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.159402 | 0.017329 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.191924 | 0.038512 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.191389 | 0.038513 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.135: re1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.081188 | 0.007238 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.138863 | 0.019627 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.276294 | 0.025565 | $(\lambda, t) = (10^3, 0.5)$ |
| BFCCMM | 0.251336 | 0.029216 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.277660 | 0.026963 | $(\lambda, m, t) = (10^3, 1 + 10^{-15}, 0.5)$ |
| KLFCs | 0.276253 | 0.023244 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.280549 | 0.020389 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.136: reviews dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.153187 | 0.137916 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.401909 | 0.109104 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.505193 | 0.057373 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.520718 | 0.057912 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.520718 | 0.057912 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.504051 | 0.143371 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.503057 | 0.144119 | $(\lambda, m) = (10^2, 1 + 10^{-15})$ |

表 3.137: sports dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.433167 | 0.064182 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.506027 | 0.075015 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.628163 | 0.104662 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.590318 | 0.085109 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.628723 | 0.100516 | $(\lambda, m, t) = (10, 1 + 10^{-3}, 10^{-1})$ |
| KLFCs | 0.457123 | 0.061962 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.458599 | 0.060718 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.138: tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.118043 | 0.067231 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.396187 | 0.073628 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.446295 | 0.055548 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.490933 | 0.047703 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.490933 | 0.047703 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.482650 | 0.096112 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.516922 | 0.107709 | $(\lambda, m) = (10^3, 1 + 10^{-1})$ |

表 3.139: tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.091214 | 0.041928 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.273812 | 0.090227 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.385571 | 0.066407 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.465958 | 0.057549 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.465958 | 0.057549 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.401174 | 0.075296 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.435601 | 0.079483 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.140: tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.042816 | 0.105301 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.187563 | 0.085512 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.232405 | 0.072949 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.265932 | 0.047353 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.265932 | 0.047353 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.218663 | 0.058184 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.234309 | 0.065914 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.141: tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.249294 | 0.114118 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.401356 | 0.101408 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.535034 | 0.085494 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.583081 | 0.063729 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.616494 | 0.087527 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.466838 | 0.072225 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.461950 | 0.080825 | $(\lambda, m) = (10^3, 1 + 10^{-4})$ |

表 3.142: tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|-----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.277744 | 0.054134 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.455327 | 0.059202 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.474838 | 0.058751 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.489535 | 0.049393 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.495037 | 0.063867 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.518547 | 0.052520 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.559420 | 0.0528151 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.143: tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.147922 | 0.080790 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.442745 | 0.097349 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.526169 | 0.093997 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.614977 | 0.064971 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.614977 | 0.064971 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.572268 | 0.065158 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.572986 | 0.077081 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.144: wap dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.170395 | 0.076814 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.325678 | 0.058722 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.285389 | 0.054284 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.318639 | 0.047735 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.318639 | 0.047735 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCS | 0.326335 | 0.042201 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.326264 | 0.044037 | $(\lambda, m) = (10^2, 1 + 10^{-4})$ |

表 3.145: webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idfA

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.068781 | 0.102159 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.079534 | 0.101500 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |
| KLFCCMM | 0.232046 | 0.099956 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.237386 | 0.106916 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.237386 | 0.106916 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.258446 | 0.034667 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.255425 | 0.038884 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.146: cacmcisi dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|-----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.146084 | 0.136570 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.156668 | 0.071142 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.326915 | 0.204115 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.326915 | 0.204115 | $(m, t) = (1 + 10^{-15}, 0.5)$ |
| QFCCMM | 0.326915 | 0.204115 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-15}, 0.5)$ |
| KLFCS | 0.020327 | 0.006226 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.0123986 | 0.003790 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.147: citeseer dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.090786 | 0.025397 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.240385 | 0.043250 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.303662 | 0.053417 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.384980 | 0.034734 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.384980 | 0.034734 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCS | 0.357742 | 0.047811 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.357949 | 0.041691 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.148: classic dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = 0.5$ |
| BFCCM | 0.191957 | 0.081837 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.402000 | 0.090523 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.506000 | 0.048138 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.524295 | 0.118017 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.524295 | 0.118017 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.374725 | 0.042484 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.370945 | 0.036030 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.149: classic4 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0.000746 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.388606 | 0.062684 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.421254 | 0.042442 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.558087 | 0.085505 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.571168 | 0.078395 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.571168 | 0.078395 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.483460 | 0.018791 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.483625 | 0.011782 | $(\lambda, m) = (10^2, 1 + 10^{-15})$ |

表 3.150: cora dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.050843 | 0.036342 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.064279 | 0.030900 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.257442 | 0.033776 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.296053 | 0.018181 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.296053 | 0.018181 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.262604 | 0.019459 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.261630 | 0.020096 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.151: cranmed dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.138791 | 0.118087 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.801970 | 0.272559 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.975489 | 0.012810 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.956504 | 0.001259 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.993424 | 0.001641 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.990636 | 0.002080 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.990800 | 0.001967 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.152: fbis dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.372325 | 0.032015 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.391067 | 0.042849 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.477401 | 0.020437 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.441018 | 0.025571 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.472601 | 0.024586 | $(\lambda, m, t) = (10, 1 + 10^{-15}, 10^{-1})$ |
| KLFCs | 0.389542 | 0.025405 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.393266 | 0.027514 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.153: hitech dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.064805 | 0.048423 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.154299 | 0.020889 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.206864 | 0.022628 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.220055 | 0.022153 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.220055 | 0.022153 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.206742 | 0.035431 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.207062 | 0.036231 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.154: k1a dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.066485 | 0.023312 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.332651 | 0.051052 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.376125 | 0.078922 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.401520 | 0.074412 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.401520 | 0.074412 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCS | 0.402427 | 0.038965 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.406370 | 0.045081 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.155: k1b dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.067867 | 0.155801 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.326351 | 0.047718 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.479155 | 0.110229 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.652203 | 0.073294 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.745656 | 0.116143 | $(\lambda, m, t) = (10, 1 + 10^{-2}, 10^{-1})$ |
| KLFCS | 0.602017 | 0.088651 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.720695 | 0.045921 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.156: la1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.100608 | 0.041090 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.346041 | 0.088169 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.411963 | 0.076460 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.455582 | 0.038264 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.455582 | 0.038264 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.473526 | 0.060187 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.474960 | 0.059162 | $(\lambda, m) = (10^3, 1 + 10^{-4})$ |

表 3.157: 1a2 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.208849 | 0.057779 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.296429 | 0.047110 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.401194 | 0.063439 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.437037 | 0.059789 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.437037 | 0.059789 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.364763 | 0.074776 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.364951 | 0.074687 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.158: 1a12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.185928 | 0.052289 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.386146 | 0.075021 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.448854 | 0.051828 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.499549 | 0.076947 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.503052 | 0.078066 | $(\lambda, m, t) = (10^3, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.411265 | 0.063704 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.410668 | 0.064063 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.159: mm dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.038481 | 0.078583 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.017393 | 0.047574 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.698228 | 0.192429 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.792001 | 0.147385 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.792001 | 0.147385 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.331688 | 0.373452 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.332158 | 0.375256 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.160: ohscal dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.221036 | 0.018428 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.275916 | 0.009250 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.379881 | 0.012299 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.384179 | 0.012913 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.384179 | 0.012913 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCS | 0.381167 | 0.017430 | $\lambda = 1.79769 \times 10^{308}$ |
| BFCS | — | — | — |
| QFCS | 0.378429 | 0.018609 | $(\lambda, m) = (10^3, 1 + 10^{-15})$ |

表 3.161: re0 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.087242 | 0.040987 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.140500 | 0.014605 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.155346 | 0.027625 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.164255 | 0.028867 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.164255 | 0.028867 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.179165 | 0.021706 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.179302 | 0.021361 | $(\lambda, m) = (10^3, 1 + 10^{-5})$ |

表 3.162: re1 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.061528 | 0.016681 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.131502 | 0.010158 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.265534 | 0.024452 | $(\lambda, t) = (10^3, 0.5)$ |
| BFCCMM | 0.225312 | 0.032720 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.267582 | 0.020817 | $(\lambda, m, t) = (10^3, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.233989 | 0.024873 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.239262 | 0.030099 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.163: reviews dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.233776 | 0.133371 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.374538 | 0.118173 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.493033 | 0.066293 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.534516 | 0.040708 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.534516 | 0.040708 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.491035 | 0.115832 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.494287 | 0.114671 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.164: sports dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.325368 | 0.086438 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.429930 | 0.094684 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.595054 | 0.088441 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.613764 | 0.065201 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.613764 | 0.065201 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.454039 | 0.073525 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.465900 | 0.079251 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.165: tr11 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.147713 | 0.050927 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.271684 | 0.058202 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.411278 | 0.055195 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.465033 | 0.050288 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.465033 | 0.050288 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.433422 | 0.070026 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.504960 | 0.090395 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.166: tr12 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.080273 | 0.028080 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.201499 | 0.048816 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.386998 | 0.071048 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.523105 | 0.086898 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.523105 | 0.086898 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.334139 | 0.067482 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.343193 | 0.071808 | $(\lambda, m) = (10^2, 1 + 10^{-3})$ |

表 3.167: tr23 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.005831 | 0.056321 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.113751 | 0.076977 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.225981 | 0.087225 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.230222 | 0.085270 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.230222 | 0.085270 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.224356 | 0.058207 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.223668 | 0.057893 | $(\lambda, m) = (10^2, 1 + 10^{-5})$ |

表 3.168: tr31 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.201881 | 0.091338 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.322651 | 0.062360 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.498754 | 0.137001 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.480441 | 0.121409 | $(m, t) = (1 + 10^{-2}, 10^{-1})$ |
| QFCCMM | 0.507071 | 0.131933 | $(\lambda, m, t) = (10^2, 1 + 10^{-2}, 10^{-1})$ |
| KLFCs | 0.399557 | 0.101332 | $\lambda = 10^3$ |
| BFCS | — | — | — |
| QFCS | 0.403708 | 0.101632 | $(\lambda, m) = (10^3, 1 + 10^{-3})$ |

表 3.169: tr41 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $0 \lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.172046 | 0.077597 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.449944 | 0.066188 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.470666 | 0.074076 | $(\lambda, t) = (10, 10^{-1})$ |
| BFCCMM | 0.471996 | 0.070701 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.471996 | 0.070701 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.450939 | 0.073247 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.472094 | 0.088799 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |

表 3.170: tr45 dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.145830 | 0.073807 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.387115 | 0.070077 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.437964 | 0.068271 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.467573 | 0.087518 | $(m, t) = (1 + 10^{-5}, 0.5)$ |
| QFCCMM | 0.467573 | 0.087518 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-5}, 0.5)$ |
| KLFCs | 0.468671 | 0.080844 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.470413 | 0.081648 | $(\lambda, m) = (10^2, 1 + 10^{-5})$ |

表 3.171: wap dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|---|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.204559 | 0.097821 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.332294 | 0.048574 | $(\lambda, m) = (10, 1 + 10^{-2})$ |
| KLFCCMM | 0.294766 | 0.064499 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.330461 | 0.049304 | $(m, t) = (1 + 10^{-4}, 0.5)$ |
| QFCCMM | 0.330461 | 0.049304 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-4}, 0.5)$ |
| KLFCs | 0.379167 | 0.096147 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.379668 | 0.094029 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.172: webkb dataset の 10 通りの初期値による ARI とパラメータ tf.idfB

| 手法 | 平均 ARI | 標準偏差 | パラメータ |
|---------|----------|----------|--|
| KLFCCM | 0 | 0 | $\lambda = \{0.5, 1.0\}$ |
| BFCCM | 0.050811 | 0.095281 | $m = 1 + 10^{-2}$ |
| QFCCM | 0.074100 | 0.099911 | $(\lambda, m) = (10^2, 1 + 10^{-2})$ |
| KLFCCMM | 0.226381 | 0.073353 | $(\lambda, t) = (1.79769 \times 10^{308}, 0.5)$ |
| BFCCMM | 0.226381 | 0.073353 | $(m, t) = (1 + 10^{-15}, 0.5)$ |
| QFCCMM | 0.226381 | 0.073353 | $(\lambda, m, t) = (1.79769 \times 10^{308}, 1 + 10^{-15}, 0.5)$ |
| KLFCs | 0.285620 | 0.022926 | $\lambda = 10^2$ |
| BFCS | — | — | — |
| QFCS | 0.285538 | 0.024658 | $(\lambda, m) = (10^3, 1 + 10^{-2})$ |

表 3.173: 10 通りの初期値による最高平均 ARI tfidf なし

| 手法 データ | | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|-----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|
| cacmcisi | 0 | 0.133791 | 0.151426 | 0.252963 | 0.253210 | 0.253210 | 0.254576 | 0 | -0.029153 | 0 |
| citepeer | 0 | 0.154531 | 0.229337 | 0.292533 | 0.354346 | 0.354346 | 0.364302 | 0.240857 | 0.243215 | 0.239188 |
| classic | 0 | 0.197311 | 0.491005 | 0.511592 | 0.505424 | 0.505424 | 0.505398 | 0.303158 | 0.310171 | 0.294757 |
| classic4 | 0.000746 | 0.412508 | 0.435832 | 0.490168 | 0.491468 | 0.491468 | 0.491696 | 0.436888 | 0.433474 | 0.436886 |
| cora | 0 | 0.097231 | 0.108148 | 0.165854 | 0.241122 | 0.241122 | 0.248180 | 0.127379 | 0.137206 | 0.126789 |
| cranmed | 0.871431 | 0.247833 | 0.872565 | 0.991292 | 0.983440 | 0.983440 | 0.993259 | 0.951199 | 0.951200 | 0.951200 |
| fbis | 0 | 0.430863 | 0.424993 | 0.488140 | 0.461586 | 0.461586 | 0.490733 | 0.382091 | — | 0.389150 |
| hitech | 0 | 0.175382 | 0.194919 | 0.270307 | 0.274295 | 0.274295 | 0.274295 | 0.265081 | — | 0.264473 |
| k1a | 0 | 0.271709 | 0.397914 | 0.296765 | 0.387499 | 0.387499 | 0.400672 | 0.275088 | — | 0.263366 |
| k1b | 0 | 0.472567 | 0.603584 | 0.731733 | 0.509426 | 0.509426 | 0.769328 | 0.509334 | — | 0.528982 |
| la1 | 0 | 0.214372 | 0.403241 | 0.501167 | 0.521592 | 0.521592 | 0.534185 | 0.211776 | — | 0.211880 |
| la2 | 0 | 0.309631 | 0.337024 | 0.470718 | 0.482179 | 0.482179 | 0.488718 | 0.194797 | — | 0.194815 |
| la12 | 0 | 0.295090 | 0.414950 | 0.513915 | 0.540250 | 0.538959 | 0.538959 | 0.218632 | — | 0.375149 |
| mm | 0 | 0.308563 | 0.312662 | 0.812183 | 0.820648 | 0.820648 | 0.820648 | 0.776885 | 0.007451 | 0.776806 |
| ohscal | 0.295765 | 0.269983 | 0.305045 | 0.305922 | 0.312164 | 0.312144 | 0.312144 | 0.212066 | — | 0.212288 |
| re0 | 0 | 0.082979 | 0.136902 | 0.170854 | 0.172366 | 0.172366 | 0.172887 | 0.169197 | — | 0.180637 |
| re1 | 0 | 0.096775 | 0.168108 | 0.226035 | 0.224395 | 0.224395 | 0.226700 | 0.243798 | — | 0.247931 |
| reviews | 0 | 0.276202 | 0.379397 | 0.538596 | 0.576558 | 0.576558 | 0.576558 | 0.564279 | — | 0.563625 |
| sports | 0 | 0.417823 | 0.425279 | 0.672250 | 0.588483 | 0.588483 | 0.681617 | 0.309572 | — | 0.309393 |
| tr11 | 0 | 0.237391 | 0.477699 | 0.555134 | 0.430620 | 0.430620 | 0.540338 | 0.439026 | — | 0.432677 |
| tr12 | 0 | 0.176235 | 0.306292 | 0.365259 | 0.368599 | 0.368599 | 0.368599 | 0.382425 | — | 0.375149 |
| tr23 | 0 | 0.076868 | 0.150237 | 0.149527 | 0.156809 | 0.156809 | 0.164993 | 0.187104 | — | 0.190638 |
| tr31 | 0 | 0.325699 | 0.489907 | 0.615231 | 0.577080 | 0.577080 | 0.628113 | 0.360381 | — | 0.384971 |
| tr41 | 0 | 0.272910 | 0.566572 | 0.475561 | 0.450730 | 0.450730 | 0.493992 | 0.371022 | — | 0.369899 |
| tr45 | 0 | 0.184500 | 0.362001 | 0.428516 | 0.488709 | 0.488709 | 0.498007 | 0.403686 | — | 0.403126 |
| wap | 0 | 0.265345 | 0.326622 | 0.266195 | 0.331515 | 0.331515 | 0.353214 | 0.261089 | — | 0.262806 |
| webkb | 0 | 0.046977 | 0.083637 | 0.224488 | 0.293891 | 0.293891 | 0.297123 | 0.204766 | — | 0.203136 |

表 3.174: 10 通りの初期値による最高平均 ARI tf.idfA

| 手法 データ | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|-----------|----------|----------|----------|-----------------|-----------------|-----------------|-----------------|----------|-----------------|
| cacmcisi | 0 | 0.109125 | 0.136767 | 0.284956 | 0.284956 | 0.284956 | 0.042454 | 0.034512 | 0.034512 |
| citeseer | 0 | 0.090936 | 0.210049 | 0.298762 | 0.381185 | 0.381185 | 0.337738 | — | 0.348793 |
| classic | 0 | 0.160741 | 0.394141 | 0.513934 | 0.515696 | 0.515696 | 0.374927 | — | 0.376880 |
| classic4 | 0.000746 | 0.370652 | 0.450194 | 0.545673 | 0.546794 | 0.546794 | 0.475238 | — | 0.473808 |
| cora | 0 | 0.062849 | 0.068871 | 0.256137 | 0.296871 | 0.296871 | 0.274829 | — | 0.279529 |
| cranmed | 0.861896 | 0.135181 | 0.865918 | 0.983774 | 0.960216 | 0.996216 | 0.991620 | — | 0.992112 |
| fbis | 0.085782 | 0.409896 | 0.440917 | 0.475528 | 0.433428 | 0.484018 | 0.391648 | — | 0.395691 |
| hitech | 0 | 0.080154 | 0.159716 | 0.225428 | 0.239141 | 0.239141 | 0.232820 | — | 0.234890 |
| k1a | 0 | 0.134708 | 0.321256 | 0.343359 | 0.347930 | 0.360423 | 0.395083 | — | 0.391710 |
| k1b | 0 | 0.171408 | 0.353940 | 0.441510 | 0.408841 | 0.743200 | 0.555370 | — | 0.698865 |
| la1 | 0 | 0.094951 | 0.344660 | 0.446305 | 0.481197 | 0.463277 | 0.483105 | — | 0.480119 |
| la2 | 0 | 0.174640 | 0.360006 | 0.482917 | 0.478598 | 0.467290 | 0.478404 | — | 0.474630 |
| la12 | 0 | 0.247759 | 0.431583 | 0.512399 | 0.521798 | 0.527401 | 0.471867 | — | 0.477453 |
| mm | 0 | 0.080518 | 0.083956 | 0.553743 | 0.612177 | 0.612177 | 0.192041 | 0.225095 | 0.222682 |
| ohscal | 0 | 0.209935 | 0.278464 | 0.331802 | 0.336163 | 0.294857 | 0.347707 | — | 0.347847 |
| re0 | 0 | 0.094556 | 0.154778 | 0.158646 | 0.159402 | 0.158908 | 0.191924 | — | 0.191389 |
| re1 | 0 | 0.081188 | 0.138863 | 0.276294 | 0.251336 | 0.277660 | 0.276253 | — | 0.280549 |
| reviews | 0 | 0.153187 | 0.401909 | 0.505193 | 0.520718 | 0.520718 | 0.504051 | — | 0.503057 |
| sports | 0 | 0.433167 | 0.506027 | 0.628163 | 0.590318 | 0.628723 | 0.457123 | — | 0.458599 |
| tr11 | 0 | 0.118043 | 0.396187 | 0.446295 | 0.490933 | 0.490933 | 0.482650 | — | 0.516922 |
| tr12 | 0 | 0.091214 | 0.273812 | 0.385571 | 0.465958 | 0.465958 | 0.401174 | — | 0.435601 |
| tr23 | 0 | 0.042816 | 0.187563 | 0.232405 | 0.265932 | 0.265932 | 0.218663 | — | 0.234309 |
| tr31 | 0 | 0.249294 | 0.401356 | 0.535034 | 0.583081 | 0.616494 | 0.461767 | — | 0.461950 |
| tr41 | 0 | 0.277744 | 0.455327 | 0.474838 | 0.489535 | 0.495037 | 0.518547 | — | 0.559420 |
| tr45 | 0 | 0.147922 | 0.442745 | 0.526169 | 0.614977 | 0.614977 | 0.572268 | — | 0.572986 |
| wap | 0 | 0.170395 | 0.325678 | 0.285389 | 0.318639 | 0.318639 | 0.326335 | — | 0.326264 |
| webkb | 0 | 0.068781 | 0.079534 | 0.232046 | 0.237386 | 0.237386 | 0.258446 | — | 0.255425 |

表 3.175: 10 通りの初期値による最高平均 ARI tf.idfB

| 手法 データ | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|-----------|----------|----------|----------|-----------------|-----------------|-----------------|-----------------|------|-----------------|
| cacmcisi | 0 | 0.146084 | 0.156668 | 0.326915 | 0.326915 | 0.326915 | 0.020327 | — | 0.012399 |
| citeseer | 0 | 0.090786 | 0.240385 | 0.303662 | 0.384980 | 0.384980 | 0.357742 | — | 0.357949 |
| classic | 0 | 0.191957 | 0.402000 | 0.506000 | 0.524295 | 0.524295 | 0.374725 | — | 0.370945 |
| classic4 | 0.000746 | 0.388606 | 0.421254 | 0.558087 | 0.571168 | 0.571168 | 0.483460 | — | 0.483625 |
| cora | 0 | 0.050843 | 0.064279 | 0.257442 | 0.296053 | 0.296053 | 0.262604 | — | 0.261630 |
| cranmed | 0 | 0.138791 | 0.801970 | 0.975489 | 0.956504 | 0.993424 | 0.990636 | — | 0.990800 |
| fbis | 0 | 0.372325 | 0.391067 | 0.477401 | 0.441018 | 0.472601 | 0.389542 | — | 0.393266 |
| hitech | 0 | 0.064805 | 0.154299 | 0.206864 | 0.220055 | 0.220055 | 0.206742 | — | 0.207062 |
| k1a | 0 | 0.066485 | 0.332651 | 0.376125 | 0.401520 | 0.401520 | 0.402427 | — | 0.406370 |
| k1b | 0 | 0.067867 | 0.326351 | 0.479155 | 0.652203 | 0.745656 | 0.602017 | — | 0.720695 |
| la1 | 0 | 0.100608 | 0.346041 | 0.411963 | 0.455582 | 0.455582 | 0.473526 | — | 0.474960 |
| la2 | 0 | 0.208849 | 0.296429 | 0.401194 | 0.437037 | 0.437037 | 0.364763 | — | 0.364951 |
| la12 | 0 | 0.185928 | 0.386146 | 0.448854 | 0.499549 | 0.503052 | 0.411265 | — | 0.410668 |
| mm | 0 | 0.038481 | 0.017393 | 0.698228 | 0.792001 | 0.792001 | 0.331688 | — | 0.332158 |
| ohscal | 0 | 0.221036 | 0.275916 | 0.379881 | 0.384179 | 0.384179 | 0.381167 | — | 0.378429 |
| re0 | 0 | 0.087242 | 0.140500 | 0.153346 | 0.164255 | 0.164255 | 0.179165 | — | 0.179302 |
| re1 | 0 | 0.061528 | 0.131502 | 0.265534 | 0.225312 | 0.267582 | 0.233989 | — | 0.239262 |
| reviews | 0 | 0.233776 | 0.374538 | 0.493033 | 0.534516 | 0.534516 | 0.491035 | — | 0.494287 |
| sports | 0 | 0.325368 | 0.429930 | 0.595054 | 0.613764 | 0.613764 | 0.454039 | — | 0.465900 |
| tr11 | 0 | 0.147713 | 0.271684 | 0.411278 | 0.465033 | 0.465033 | 0.433422 | — | 0.504960 |
| tr12 | 0 | 0.080273 | 0.201499 | 0.386998 | 0.523105 | 0.523105 | 0.334139 | — | 0.343193 |
| tr23 | 0 | 0.005831 | 0.113751 | 0.225981 | 0.230222 | 0.230222 | 0.224356 | — | 0.223668 |
| tr31 | 0 | 0.201881 | 0.322651 | 0.498754 | 0.480441 | 0.507071 | 0.399557 | — | 0.403708 |
| tr41 | 0 | 0.172046 | 0.449944 | 0.470666 | 0.471996 | 0.471996 | 0.450939 | — | 0.472094 |
| tr45 | 0 | 0.145830 | 0.387115 | 0.437964 | 0.467573 | 0.467573 | 0.468671 | — | 0.470413 |
| wap | 0 | 0.204559 | 0.332294 | 0.294766 | 0.330461 | 0.330461 | 0.379167 | — | 0.379668 |
| webkb | 0 | 0.050811 | 0.074100 | 0.226381 | 0.226381 | 0.226381 | 0.285620 | — | 0.285538 |

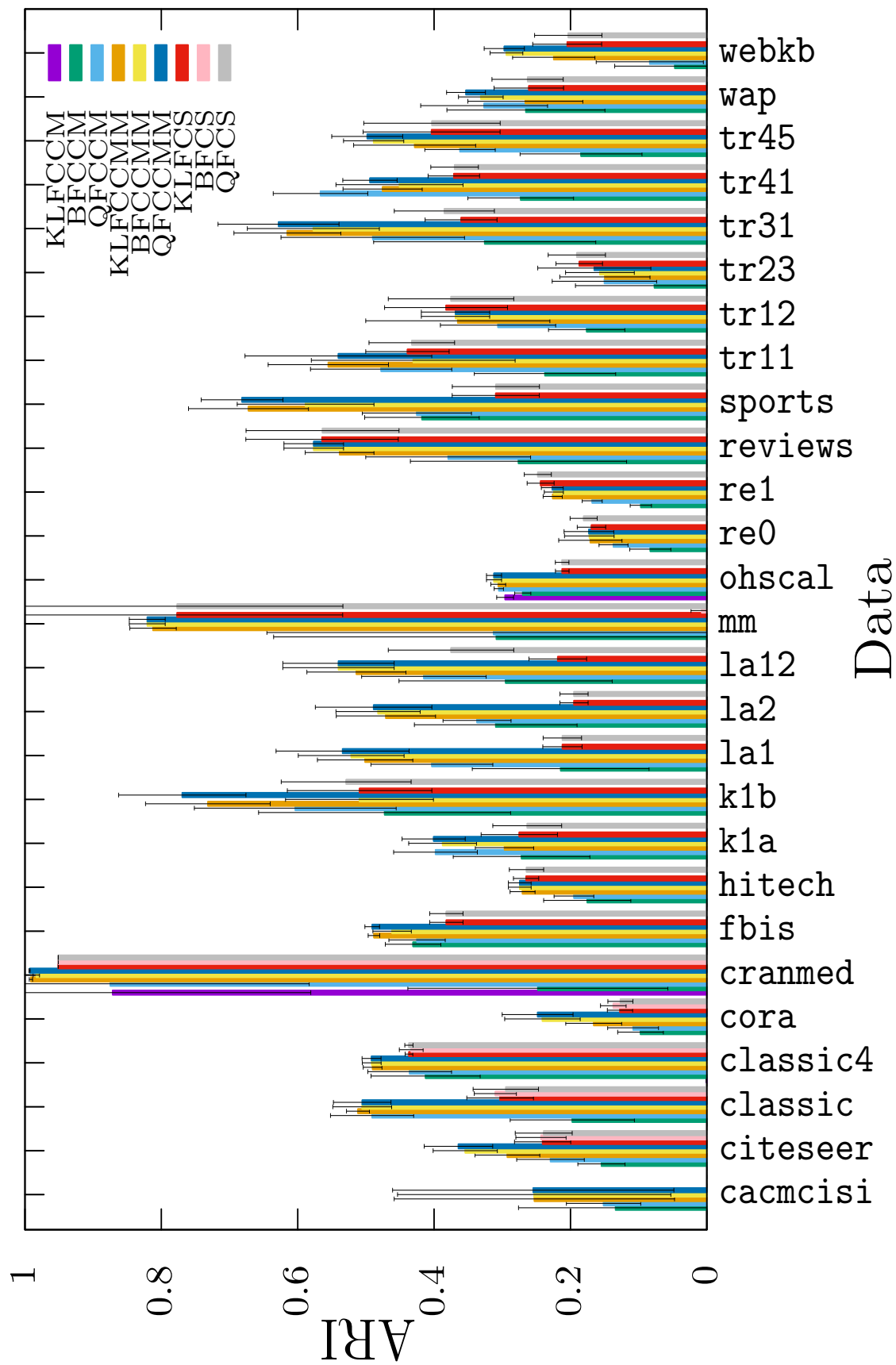


図 3.1: 10 通りの初期値による最高平均 ARI と標準偏差 tf.idf なし

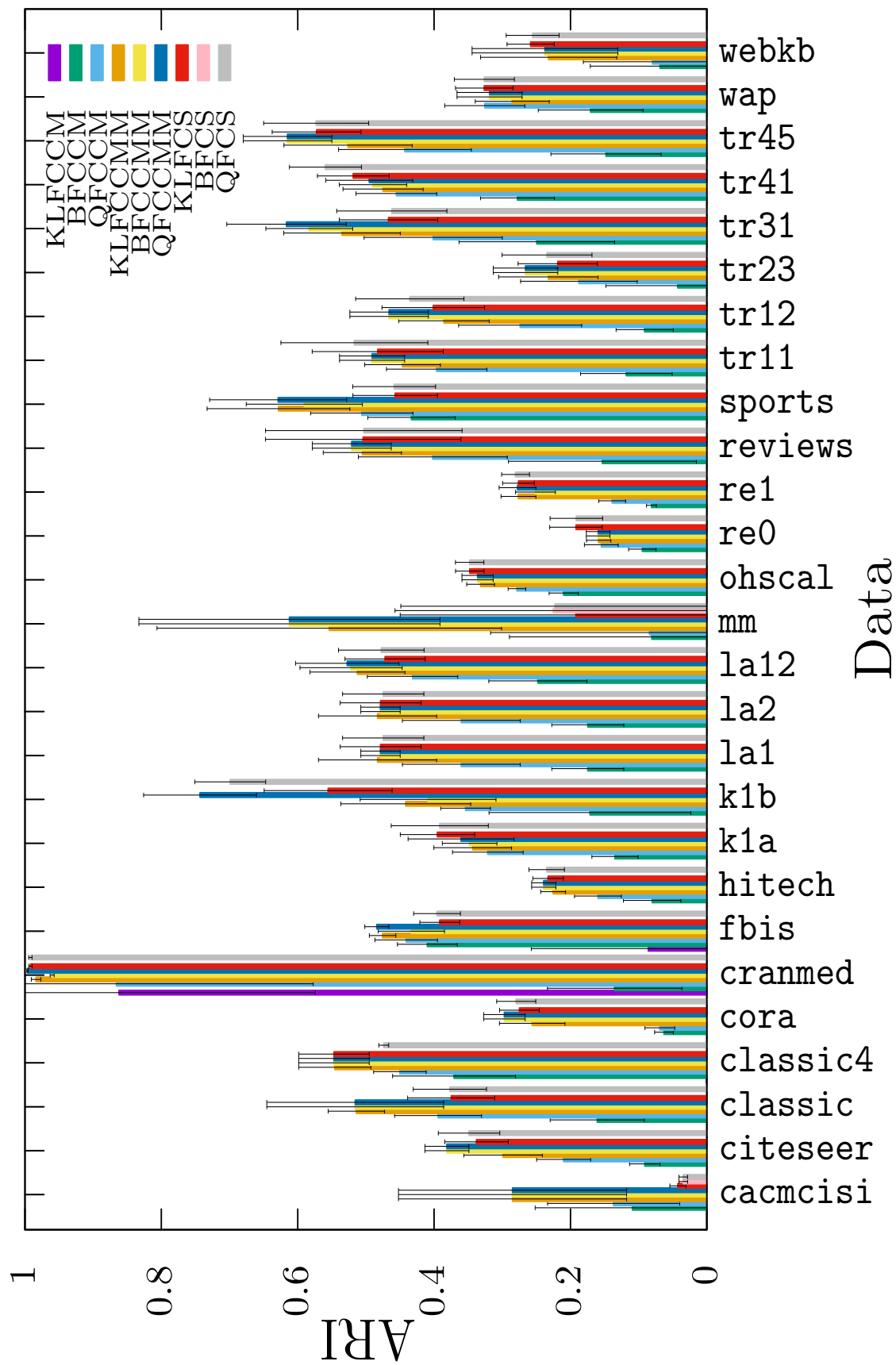


図 3.2: 10 通りの初期値による最高平均 ARI と標準偏差 tf.idfa

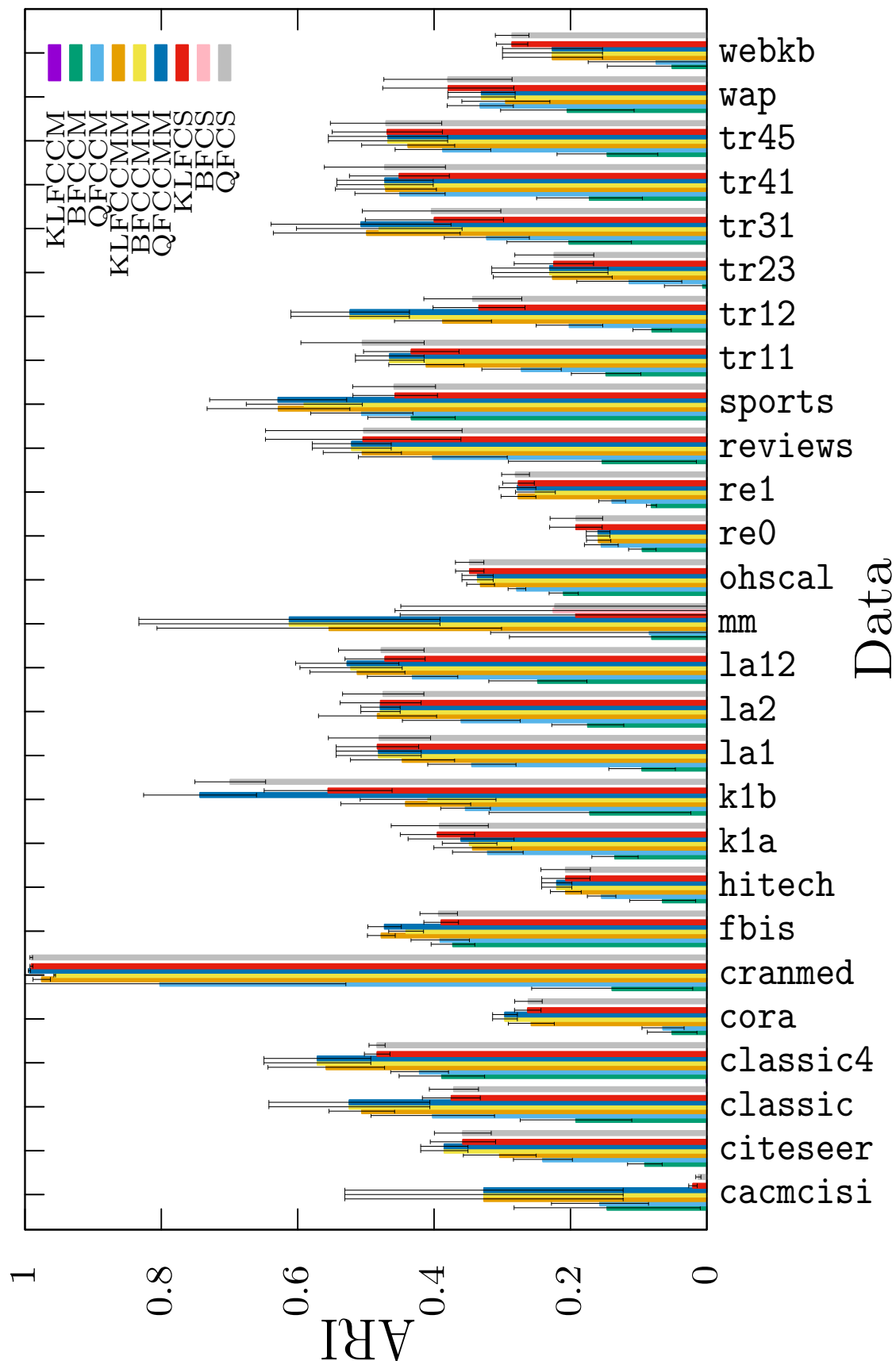


図 3.3: 10 通りの初期値による最高平均 ARI と標準偏差 tf.idfB

3.6.1 数値実験2の結果

表 3.176 に、10 通りの初期設定による平均 ARI の最高値を達成した文書データ数を示す。数値実験 2 では、三つのデータ形式において、QFCCMM を適用すると半数以上のデータで最高平均 ARI を達成し、QFCS や QFCCM と比べて高精度であることを確認した。

表 3.176: 10 通りの初期設定による平均 ARI の最高値を達成したデータ数

| データ 手法 | tf.idf なし | tf.idfA | tf.idfB |
|-----------|-----------|---------|---------|
| KLFCCM | 0 | 0 | 0 |
| BFCCM | 0 | 0 | 0 |
| QFCCM | 1 | 0 | 0 |
| KLFCCMM | 1 | 2 | 2 |
| BFCCMM | 5 | 11 | 13 |
| QFCCMM | 19 | 17 | 18 |
| KLFCS | 1 | 5 | 1 |
| BFCS | 0 | 0 | 0 |
| QFCS | 3 | 4 | 7 |

次に初期値の設定数を 10 通りから 100 通りに変更した数値実験 3 を行う。平均 ARI だけでなく、目的関数の最小値を達成したクラスタリング結果による ARI も算出し、精度比較を行う。

3.7 数値実験 3

数値実験 2 で得た 10 通りの平均で最高 ARI を達成したパラメータを適用して 100 通りのクラスタ中心を設定しクラスタリングを行う。以下に数値実験 3 の流れを示す：

実験の流れ 3 (数値実験 3).

- STEP 1. new を除いた表 3.1 に示すデータの読み込み、正解の帰属度の読み込みをする。
- STEP 2. tf.idf なしでない場合、式 (3.5)、式 (3.6) に従い tf.idfA や tf.idfB に変換する。
- STEP 3. {B,KL,Q}FCS では式 (3.8) に従い球面化し、{B,KL,Q}FC{CM,CMM} では式 (3.9) に従い単体化する。
- STEP 4. クラスタ数を正解クラス数に設定する。
- STEP 5. 表 3.92 から表 3.172 に示した最高平均 ARI を達成したパラメータを設定する。
- STEP 6. k -means++ における平方ユークリッド距離の代わりにオブジェクト-クラスタ間非類似度を {B,KL,Q}FCS では式 (2.78)、{KL,Q,B}FCCM では式 (2.128)、{KL,B,Q}FCCMM では式 (2.172) に置き換えたアルゴリズム 3 に従い初期クラスタ中心を設定する。
- STEP 7. 初期クラスタサイズ調整変数を $1/C$ に設定する。
- STEP 8. {B,KL,Q}FCS では式 (2.78)、{KL,Q,B}FCCM では式 (2.128)、{KL,B,Q}FCCMM では式 (2.172) に従いオブジェクト-クラスタ間非類似度を更新する。
- STEP 9. BFCS では式 (2.88)、KLFCS では式 (2.102)、QFCS では式 (2.116)、KLFCCM では

- 式 (2.138), BFCCM では式 (2.166), QFCCM では式 (2.152), KLFCCMM では式 (2.182), BFCCMM では式 (2.196), QFCCMM では式 (2.210) に従い帰属度を更新する.
- STEP 10. BFCS では式 (2.90), KLFCS では式 (2.104), QFCS では式 (2.118), KLFCCM では式 (2.140), BFCCM では式 (2.168), QFCCM では式 (2.154), KLFCCMM では式 (2.184), BFCCMM では式 (2.198), QFCCMM では式 (2.212) に従いクラスタ中心を更新する.
- STEP 11. BFCS では式 (2.92), KLFCS では式 (2.106), QFCS では式 (2.120), KLFCCM では式 (2.142), BFCCM では式 (2.170), QFCCM では式 (2.156), KLFCCMM では式 (2.186), BFCCMM では式 (2.200), QFCCMM では式 (2.214) に従いクラスタサイズ調整変数を更新する.
- STEP 12. 更新前と更新後の帰属度, クラスタ中心, クラスタサイズ調整変数の最大値ノルムを計算し, それらの和が 10^{-6} 以上であれば STEP 8. へ, 計算限界を迎えたら初期値設定乱数シードをインクリメントし STEP 6. へ, 100 回計算限界を迎えたもしくは更新数が 10^4 を超えたらそのパラメータでのクラスタリングを不可として STEP 17. へ.
- STEP 13. 収束した帰属度をクリस्प化する.
- STEP 14. クリस्प化された帰属度と正解の帰属度を照らし合わせ Contingency Table を得る.
- STEP 15. Contingency Table から式 (3.7) に従い ARI を算出する. また, 収束した変数から目的関数値を算出し保持する.
- STEP 16. 100 通りの初期設定を行えたなら平均 ARI と標準偏差を算出し目的関数値最小の ARI を保持する. そうでなければ STEP 6. へ.
- STEP 17. 次に処理すべきデータ変換手法があれば STEP 2. へ.
- STEP 18. 次に処理すべきデータがあれば STEP 1. へ, そうでなければ終了する. ■

データ対ファジィクラスタリング手法で最高平均 ARI を表 3.177, 表 3.178, 表 3.179 にまとめた. また, 100 通りのクラスタ中心による初期設定のうち, 目的関数値最小の ARI を表 3.180, 表 3.181, 表 3.182 にまとめた. 図 3.4, 図 3.5, 図 3.6 にそれぞれのデータ形式における最高平均 ARI とその標準偏差を示す.

表 3.177: 100 通りの初期値による最高平均 ARI tf.idf なし

| 手法 データ | KLFCM | BFCCM | QFCCM | KLFCMM | BFCCMM | QFCCMM | KLFC | BFCS | QFCS |
|-----------|----------|----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------|-----------------|
| cacmcisi | 0 | 0.166645 | 0.142322 | 0.283865 | 0.289642 | 0.285182 | 0 | -0.018982 | 0 |
| citepeer | 0 | 0.168043 | 0.264315 | 0.304926 | 0.350448 | 0.360149 | 0.233270 | 0.024322 | 0.231342 |
| classic | 0 | 0.259816 | 0.450451 | 0.509701 | 0.498806 | 0.499302 | 0.295040 | 0.054937 | 0.285103 |
| classic4 | 0.000746 | 0.367455 | 0.449094 | 0.479857 | 0.484389 | 0.481971 | 0.426011 | 0.064093 | 0.425747 |
| cora | 0 | 0.089773 | 0.107108 | 0.173154 | 0.247325 | 0.248075 | 0.135150 | — | 0.134880 |
| cranmed | — | 0.232967 | 0.116486 | 0.990865 | 0.983016 | 0.992865 | 0.951199 | 0.494624 | 0.951200 |
| fbis | 0 | 0.433379 | 0.434088 | 0.493359 | 0.462550 | 0.492960 | 0.385391 | 0 | 0.389754 |
| hitech | 0 | 0.169380 | 0.198223 | 0.277287 | 0.280048 | 0.280048 | 0.262884 | — | 0.261897 |
| kia | 0 | 0.284882 | 0.403279 | 0.310827 | 0.396553 | 0.413778 | 0.268018 | — | 0.262514 |
| kib | 0 | 0.400893 | 0.590156 | 0.767848 | 0.513222 | 0.801983 | 0.460784 | — | 0.475212 |
| la1 | 0 | 0.231360 | 0.398493 | 0.497610 | 0.511310 | 0.515710 | 0.202503 | — | 0.201754 |
| la2 | 0 | 0.252491 | 0.345959 | — | 0.507942 | — | 0.203750 | — | 0.203698 |
| la12 | 0 | 0.309273 | 0.398648 | 0.491976 | 0.544447 | 0.543901 | 0.210063 | — | 0.172110 |
| mm | 0 | 0.178664 | 0.190486 | 0.824364 | 0.817318 | 0.817318 | 0.747365 | 0.266047 | 0.738827 |
| ohscal | 0.294896 | 0.277171 | 0.306387 | 0.305964 | 0.317419 | 0.316486 | 0.206059 | — | 0.206173 |
| re0 | 0 | 0.113858 | 0.154791 | 0.161060 | 0.173185 | 0.173208 | 0.173046 | — | 0.187234 |
| re1 | 0 | 0.107869 | 0.163192 | 0.224592 | 0.221173 | 0.224544 | 0.249363 | — | 0.252373 |
| reviews | 0 | 0.306915 | 0.417431 | 0.533443 | 0.558713 | 0.558713 | 0.561838 | — | 0.560367 |
| sports | 0 | 0.447012 | 0.457660 | 0.673685 | 0.594683 | 0.674525 | 0.315026 | — | 0.315364 |
| tr11 | 0 | 0.230217 | 0.471350 | 0.556106 | 0.405982 | 0.552637 | 0.416776 | — | 0.425922 |
| tr12 | 0 | 0.147067 | 0.308860 | 0.394484 | 0.347881 | 0.347881 | 0.333967 | — | 0.326053 |
| tr23 | 0 | 0.039284 | 0.143254 | 0.167575 | 0.161515 | 0.171591 | 0.187172 | — | 0.190185 |
| tr31 | 0 | 0.339244 | 0.512857 | 0.589968 | 0.500192 | 0.583999 | 0.396499 | — | 0.403703 |
| tr41 | 0 | 0.288322 | 0.517846 | 0.509781 | 0.463186 | 0.510653 | 0.379404 | — | 0.370300 |
| tr45 | 0 | 0.195030 | 0.382463 | 0.434422 | 0.438565 | 0.449758 | 0.370892 | — | 0.373409 |
| wap | 0 | 0.292066 | 0.332595 | 0.258239 | 0.352628 | 0.370035 | 0.245744 | — | 0.248110 |
| webkb | 0 | 0.032621 | 0.026399 | 0.211233 | 0.294572 | 0.296122 | 0.199115 | — | 0.200271 |

表 3.178: 100 通りの初期値による最高平均 ARI tfidfA

| データ \ 手法 | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|----------|-----------|----------|----------|----------|----------|----------|----------|----------|----------|
| caacmcsi | -0.000644 | 0.142391 | 0.177645 | 0.340284 | 0.340284 | 0.340284 | 0.04255 | 0.041524 | 0.041524 |
| citepeer | 0 | 0.092975 | 0.214870 | 0.294899 | 0.374117 | 0.374117 | 0.348508 | — | 0.349288 |
| classic | 0 | 0.182779 | 0.214870 | 0.294899 | 0.374117 | 0.374117 | 0.348508 | — | 0.394668 |
| classic4 | 0.000746 | 0.386157 | 0.443772 | 0.529791 | 0.533026 | 0.533026 | 0.470941 | — | 0.472590 |
| cora | 0 | 0.057761 | 0.071222 | 0.253775 | 0.291895 | 0.291895 | 0.281116 | — | 0.286066 |
| cranmed | — | 0.105482 | 0.086592 | 0.984084 | 0.950526 | 0.995460 | 0.991505 | — | 0.992014 |
| fbis | — | 0.416036 | 0.044092 | 0.480070 | 0.436727 | — | 0.391011 | — | 0.399467 |
| hitech | 0 | 0.071469 | 0.169422 | 0.232412 | 0.243796 | 0.243796 | 0.234732 | — | 0.231242 |
| kla | 0 | 0.101177 | 0.326715 | 0.353459 | 0.345347 | — | 0.384439 | — | 0.379897 |
| k1b | 0 | 0.121166 | 0.353376 | 0.411140 | 0.389655 | 0.701987 | 0.546387 | — | 0.683672 |
| la1 | 0 | 0.108588 | 0.353467 | 0.450172 | 0.487470 | 0.487470 | 0.454550 | — | 0.467192 |
| la2 | 0 | 0.165456 | 0.351520 | 0.507871 | 0.490160 | 0.490160 | 0.484780 | — | 0.473589 |
| la12 | 0 | 0.234865 | 0.437481 | 0.537210 | 0.512477 | 0.525503 | 0.504049 | — | 0.508355 |
| mm | 0 | 0.081552 | 0.154588 | 0.544615 | 0.621069 | 0.621069 | 0.411422 | 0.273304 | 0.277868 |
| ohscal | 0 | 0.206735 | 0.271147 | 0.328427 | 0.333661 | 0.333661 | 0.345320 | — | 0.345847 |
| re0 | 0 | 0.084719 | 0.142143 | 0.153099 | 0.153035 | 0.153035 | 0.176216 | — | 0.176010 |
| re1 | 0 | 0.079558 | 0.138960 | 0.285035 | 0.266287 | 0.283916 | 0.278263 | — | 0.284456 |
| reviews | 0 | 0.225141 | 0.396967 | 0.493361 | 0.503267 | 0.503267 | 0.476082 | — | 0.468640 |
| sports | 0 | 0.402299 | 0.465475 | 0.575152 | 0.594237 | 0.583914 | 0.460515 | — | 0.455559 |
| tr11 | 0 | 0.143884 | 0.338947 | 0.428528 | 0.474030 | 0.474030 | 0.511037 | — | 0.476599 |
| tr12 | 0 | 0.105652 | 0.276779 | 0.379637 | 0.463524 | 0.463524 | 0.440733 | — | 0.466708 |
| tr23 | 0 | 0.054083 | 0.151210 | 0.250321 | 0.278812 | 0.278812 | 0.238992 | — | 0.243207 |
| tr31 | 0 | 0.272747 | 0.429999 | 0.509055 | 0.545476 | 0.582000 | 0.468788 | — | 0.470628 |
| tr41 | 0 | 0.235188 | 0.450486 | 0.476295 | 0.495255 | 0.495255 | 0.517451 | — | 0.554450 |
| tr45 | 0 | 0.159490 | 0.442490 | 0.527498 | 0.578244 | 0.578244 | 0.607299 | — | 0.613865 |
| wap | 0 | 0.193573 | 0.333470 | 0.290692 | 0.316917 | 0.316917 | 0.362780 | — | 0.362512 |
| webkb | 0 | 0.028800 | 0.031427 | 0.237077 | 0.241132 | 0.241132 | 0.219101 | — | 0.226854 |

表 3.179: 100 通りの初期値による最高平均 ARI tf.idfb

| 手法 データ | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|-----------|----------|----------|----------|-----------------|-----------------|-----------------|-----------------|------|-----------------|
| cacmcisi | 0 | 0.132167 | 0.161552 | 0.351469 | 0.351469 | 0.351469 | 0.020136 | — | 0.011422 |
| citeseer | 0 | 0.099243 | 0.223319 | 0.298870 | 0.379902 | 0.379902 | 0.350791 | — | 0.336887 |
| classic | 0 | 0.178765 | 0.406345 | 0.503574 | 0.492546 | 0.492546 | 0.384526 | — | 0.378331 |
| classic4 | 0.000746 | 0.346646 | 0.426605 | 0.554391 | 0.560867 | 0.560867 | 0.480201 | — | 0.468791 |
| cora | 0 | 0.059303 | 0.075624 | 0.257895 | 0.289776 | 0.289776 | 0.261500 | — | 0.261155 |
| cranmed | 0 | 0.115662 | 0.700507 | 0.979978 | 0.956795 | 0.993686 | 0.990324 | — | 0.990423 |
| fbis | 0 | 0.386763 | 0.405511 | — | 0.441183 | 0.462840 | 0.386600 | — | 0.385936 |
| hitech | 0 | 0.061711 | 0.161851 | 0.214560 | 0.224607 | 0.224607 | 0.205972 | — | 0.205688 |
| k1a | 0 | 0.096396 | 0.344495 | 0.380180 | 0.373240 | 0.373298 | 0.399000 | — | 0.395358 |
| k1b | 0 | 0.102393 | 0.356261 | 0.430772 | — | 0.765973 | 0.620562 | — | 0.731402 |
| la1 | 0 | 0.103965 | 0.328833 | 0.428652 | 0.460026 | 0.460026 | 0.425807 | — | 0.426184 |
| la2 | 0 | 0.154029 | 0.308761 | 0.418677 | 0.456001 | 0.456001 | 0.380987 | — | 0.380221 |
| la12 | 0 | 0.203921 | 0.375690 | 0.411004 | 0.469283 | 0.468603 | 0.418539 | — | 0.416637 |
| mm | 0 | 0.061890 | 0.078535 | 0.659559 | 0.753207 | 0.753207 | 0.369443 | — | 0.370422 |
| ohscal | 0 | 0.223640 | 0.280622 | 0.379911 | 0.384581 | 0.384581 | 0.372908 | — | 0.377067 |
| re0 | 0 | 0.081434 | 0.132053 | 0.166196 | 0.172488 | 0.172488 | 0.172751 | — | 0.132053 |
| re1 | 0 | 0.071593 | 0.129418 | 0.267428 | 0.240095 | 0.268081 | 0.231431 | — | 0.227810 |
| reviews | 0 | 0.185081 | 0.367015 | 0.491336 | 0.506837 | 0.506837 | 0.502380 | — | 0.505438 |
| sports | 0 | 0.357761 | 0.418854 | 0.606266 | 0.629785 | 0.629785 | 0.453244 | — | 0.457071 |
| tr11 | 0 | 0.137588 | 0.274617 | 0.399555 | 0.441173 | 0.441173 | 0.457261 | — | 0.495420 |
| tr12 | 0 | 0.088311 | 0.209599 | 0.376059 | 0.477272 | 0.477272 | 0.325285 | — | 0.327342 |
| tr23 | 0 | 0.017294 | 0.122252 | 0.221339 | 0.226915 | 0.226915 | 0.218947 | — | 0.219504 |
| tr31 | 0 | 0.197083 | 0.314786 | 0.427518 | 0.379292 | 0.392291 | 0.343128 | — | 0.342418 |
| tr41 | 0 | 0.191025 | 0.433354 | 0.472566 | 0.467723 | 0.467723 | 0.458254 | — | 0.477050 |
| tr45 | 0 | 0.146642 | 0.368210 | 0.427570 | 0.459773 | 0.459773 | 0.472110 | — | 0.470279 |
| wap | 0 | 0.224491 | 0.358794 | 0.310296 | 0.339327 | 0.339327 | 0.369683 | — | 0.365864 |
| webkb | 0 | 0.023899 | 0.038335 | 0.229172 | 0.229172 | 0.229172 | 0.263781 | — | 0.260066 |

表 3.180: 100 通りの初期値による目的関数値最小 ARI tfidf なし

| Method | | | | | | | | | | |
|----------|----------|-----------|-----------|----------|----------|----------|----------|-----------|----------|--|
| Data | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS | |
| cacmcisi | 0 | 0.049724 | 0.084280 | 0.155794 | 0.064399 | 0.064762 | 0 | -0.031884 | 0 | |
| citeseer | 0 | 0.222650 | 0.337127 | 0.331782 | 0.401365 | 0.401642 | 0.311094 | 0.303470 | 0.197176 | |
| classic | 0 | 0.527548 | 0.506853 | 0.509598 | 0.493329 | 0.492909 | 0.333834 | 0.280467 | 0.263483 | |
| classic4 | 0.000746 | 0.479721 | 0.475278 | 0.479105 | 0.477812 | 0.478577 | 0.441466 | 0.442287 | 0.441517 | |
| cora | 0 | 0.172700 | 0.121635 | 0.210422 | 0.255959 | 0.255932 | 0.152640 | — | 0.153502 | |
| cranmed | — | 0.733991 | 0 | 0.993423 | 0.993423 | 0.995065 | 0.951199 | 0.951200 | 0.951200 | |
| fbis | 0 | 0.481751 | 0.414263 | 0.507884 | 0.465291 | 0.500887 | 0.374255 | 0 | 0.396684 | |
| hitech | 0 | 0.140195 | 0.224388 | 0.290646 | 0.261170 | 0.261170 | 0.237324 | — | 0.239672 | |
| k1a | 0 | 0.428245 | 0.408580 | 0.388029 | 0.358126 | 0.365196 | 0.314449 | — | 0.311117 | |
| k1b | 0 | 0.279550 | 0.736734 | 0.828980 | 0.513695 | 0.859760 | 0.488568 | — | 0.511607 | |
| la1 | 0 | 0.268379 | 0.513226 | 0.603402 | 0.558001 | 0.553402 | 0.246980 | — | 0.278557 | |
| la2 | 0 | 0.274936 | 0.351170 | — | 0.588409 | — | 0.223056 | — | 0.223125 | |
| la12 | 0 | 0.465040 | 0.496666 | 0.386858 | 0.567552 | 0.566735 | 0.206244 | — | 0.180924 | |
| mm | 0 | -0.001617 | -0.001965 | 0.850483 | 0.844637 | 0.844637 | 0.857823 | 0.856353 | 0.857823 | |
| ohscal | 0.298374 | 0.302531 | 0.315337 | 0.320411 | 0.322802 | 0.326590 | 0.217969 | — | 0.222877 | |
| re0 | 0 | 0.157287 | 0.113178 | 0.178989 | 0.172170 | 0.173334 | 0.159689 | — | 0.157578 | |
| re1 | 0 | 0.163575 | 0.146668 | 0.235219 | 0.211695 | 0.239767 | 0.266342 | — | 0.258118 | |
| reviews | 0 | 0.587653 | 0.332441 | 0.463440 | 0.628220 | 0.628220 | 0.700147 | — | 0.700451 | |
| sports | 0 | 0.501852 | 0.589233 | 0.769838 | 0.637599 | 0.770794 | 0.216338 | — | 0.324979 | |
| tr11 | 0 | 0.348830 | 0.540709 | 0.628390 | 0.487029 | 0.625421 | 0.543650 | — | 0.438437 | |
| tr12 | 0 | 0.246505 | 0.285605 | 0.427181 | 0.448870 | 0.448870 | 0.404799 | — | 0.373693 | |
| tr23 | 0 | 0.189152 | 0.197313 | 0.211762 | 0.172042 | 0.194008 | 0.201560 | — | 0.163189 | |
| tr31 | 0 | 0.315997 | 0.319596 | 0.710484 | 0.534565 | 0.712131 | 0.390227 | — | 0.402031 | |
| tr41 | 0 | 0.366605 | 0.558510 | 0.510102 | 0.449456 | 0.509413 | 0.474615 | — | 0.402290 | |
| tr45 | 0 | 0.385074 | 0.567928 | 0.377287 | 0.576684 | 0.621482 | 0.359808 | — | 0.355743 | |
| wap | 0 | 0.330663 | 0.467156 | 0.160717 | 0.348750 | 0.386484 | 0.248561 | — | 0.172778 | |
| webkb | 0 | 0.120495 | 0.062951 | 0.284013 | 0.340781 | 0.330122 | 0.249908 | — | 0.200422 | |

表 3.181: 100 通りの初期値による目的関数値最小 ARI tf.idfA

| Method | | | | | | | | | | |
|----------|-----------|-----------|-----------|----------|----------|----------|----------|----------|----------|----------|
| Data | | KLFCCM | BFCCM | QFCCM | KLFCCMM | BFCCMM | QFCCMM | KLFCs | BFCS | QFCS |
| cacmci | -0.000233 | 0.152283 | 0.111388 | 0.208632 | 0.123142 | 0.123142 | 0.123142 | 0.041160 | 0.032006 | 0.032006 |
| citeseer | 0 | 0.020047 | 0.276594 | 0.349269 | 0.391945 | 0.391945 | 0.391945 | 0.392300 | — | 0.391625 |
| classic | 0 | 0.368436 | 0.276594 | 0.349269 | 0.391945 | 0.391945 | 0.391945 | 0.392300 | — | 0.394936 |
| classic4 | 0.000746 | 0.483575 | 0.478034 | 0.541076 | 0.488563 | 0.488563 | 0.488563 | 0.470852 | — | 0.472101 |
| cora | 0 | 0.101333 | 0.116085 | 0.324755 | 0.304158 | 0.304158 | 0.304158 | 0.312847 | — | 0.306796 |
| cranmed | — | 0.537324 | 0.968952 | 0.995065 | 0.965710 | 0.965710 | 0.996709 | 0.990143 | — | 0.993423 |
| fbis | — | 0.428938 | 0.508517 | 0.482436 | 0.425628 | — | — | 0.393062 | — | 0.401069 |
| hitech | 0 | 0.125535 | 0.162269 | 0.233772 | 0.251259 | 0.251259 | 0.251259 | 0.199227 | — | 0.250646 |
| kia | 0 | 0.120585 | 0.343233 | 0.390292 | 0.342800 | — | — | 0.347262 | — | 0.335674 |
| k1b | 0 | 0.250181 | 0.361798 | 0.440085 | 0.533257 | 0.533257 | 0.783683 | 0.642550 | — | 0.718070 |
| la1 | 0 | 0.156275 | 0.415115 | 0.552764 | 0.494746 | 0.494746 | 0.494746 | 0.513617 | — | 0.519880 |
| la2 | 0 | 0.273027 | 0.385114 | 0.577571 | 0.481067 | 0.481067 | 0.481067 | 0.516075 | — | 0.489628 |
| la12 | 0 | 0.255187 | 0.520307 | 0.573796 | 0.564258 | 0.565256 | 0.565256 | 0.516229 | — | 0.520688 |
| mm | 0 | -0.002914 | -0.006252 | 0.400952 | 0.851947 | 0.851947 | 0.851947 | 0.001683 | 0.001683 | 0.001450 |
| ohscal | 0 | 0.220903 | 0.271662 | 0.326212 | 0.361831 | 0.361831 | 0.361831 | 0.361563 | — | 0.346776 |
| re0 | 0 | 0.088975 | 0.150354 | 0.149105 | 0.140017 | 0.140017 | 0.140017 | 0.186310 | — | 0.213631 |
| re1 | 0 | 0.084114 | 0.141287 | 0.269922 | 0.302546 | 0.302546 | 0.308053 | 0.283311 | — | 0.272173 |
| reviews | 0 | 0.378996 | 0.544090 | 0.495624 | 0.603144 | 0.603144 | 0.603144 | 0.155650 | — | 0.473216 |
| sports | 0 | 0.361194 | 0.468154 | 0.622344 | 0.584772 | 0.627099 | 0.627099 | 0.448624 | — | 0.447037 |
| tr11 | 0 | 0.083936 | 0.300452 | 0.365763 | 0.521941 | 0.521941 | 0.521941 | 0.559358 | — | 0.171197 |
| tr12 | 0 | 0.109914 | 0.212463 | 0.303890 | 0.484725 | 0.484725 | 0.484725 | 0.490339 | — | 0.479380 |
| tr23 | 0 | 0.069892 | 0.253590 | 0.312887 | 0.287989 | 0.287989 | 0.287989 | 0.264637 | — | 0.276403 |
| tr31 | 0 | 0.440639 | 0.410089 | 0.547713 | 0.665605 | 0.577044 | 0.577044 | 0.524223 | — | 0.481057 |
| tr41 | 0 | 0.345606 | 0.507110 | 0.556288 | 0.560065 | 0.560065 | 0.560065 | 0.586519 | — | 0.634543 |
| tr45 | 0 | 0.255199 | 0.562714 | 0.513604 | 0.515408 | 0.515408 | 0.515408 | 0.730891 | — | 0.756658 |
| wap | 0 | 0.368858 | 0.368763 | 0.258570 | 0.359373 | 0.359373 | 0.359373 | 0.355921 | — | 0.351415 |
| webkb | 0 | 0.090367 | 0.090367 | 0.245660 | 0.260739 | 0.260739 | 0.260739 | 0.267135 | — | 0.318516 |

表 3.182: 100 通りの初期値による目的関数値最小 ARI tf.idfB

| Data | Method | KLFCM | BFCCM | QFCCM | KLFCMM | BFCCMM | QFCCMM | KLFCS | BFCS | QFCS |
|----------|----------|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------|-----------------|
| | | | | | | | | | | |
| cacmcisi | 0 | 0.143597 | 0.109073 | 0.202722 | 0.116823 | 0.116823 | 0.116823 | 0.025542 | — | 0.002516 |
| citepeer | 0 | 0.131275 | 0.281833 | 0.349663 | 0.396558 | 0.396558 | 0.396558 | 0.397743 | — | 0.400028 |
| classic | 0 | 0.489573 | 0.510646 | 0.514869 | 0.307868 | 0.307868 | 0.307868 | 0.342717 | — | 0.368794 |
| classic4 | 0.000746 | 0.476177 | 0.478190 | 0.546887 | 0.458694 | 0.458694 | 0.458694 | 0.505163 | — | 0.467471 |
| cora | 0 | 0.074008 | 0.084781 | 0.306438 | 0.291357 | 0.291357 | 0.291357 | 0.283334 | — | 0.282334 |
| cranmed | 0 | 0.481114 | 0.973824 | 0.991782 | 0.959245 | 0.959245 | 0.995065 | 0.988505 | — | 0.990143 |
| fbis | 0 | 0.468845 | 0.449822 | — | 0.464592 | 0.464592 | 0.447934 | 0.406712 | — | 0.413455 |
| hitech | 0 | 0.088104 | 0.170732 | 0.198595 | 0.245515 | 0.245515 | 0.245515 | 0.224903 | — | 0.211855 |
| kla | 0 | 0.232311 | 0.396884 | 0.439962 | 0.520521 | 0.520521 | 0.520521 | 0.436662 | — | 0.433620 |
| kib | 0 | 0.107617 | 0.368085 | 0.646402 | — | — | 0.821272 | 0.750009 | — | 0.783243 |
| la1 | 0 | 0.238303 | 0.499730 | 0.429080 | 0.488444 | 0.488444 | 0.488444 | 0.545304 | — | 0.542669 |
| la2 | 0 | 0.283307 | 0.313352 | 0.329108 | 0.532603 | 0.532603 | 0.532603 | 0.349577 | — | 0.349577 |
| la12 | 0 | 0.259935 | 0.463755 | 0.479718 | 0.550968 | 0.550968 | 0.550142 | 0.534024 | — | 0.455925 |
| mm | 0 | -0.002774 | -0.005777 | 0.863710 | 0.917636 | 0.917636 | 0.917636 | -0.005671 | — | -0.005506 |
| ohscal | 0 | 0.200999 | 0.292896 | 0.403296 | 0.406332 | 0.406332 | 0.406332 | 0.384917 | — | 0.353184 |
| re0 | 0 | 0.054867 | 0.168810 | 0.150194 | 0.163332 | 0.163332 | 0.163332 | 0.190039 | — | 0.168810 |
| re1 | 0 | 0.062634 | 0.151082 | 0.216634 | 0.227832 | 0.227832 | 0.210669 | 0.235207 | — | 0.216454 |
| reviews | 0 | 0.242396 | 0.545149 | 0.495748 | 0.521456 | 0.521456 | 0.521456 | 0.577867 | — | 0.574099 |
| sports | 0 | 0.423954 | 0.527910 | 0.482241 | 0.621275 | 0.621275 | 0.621275 | 0.531927 | — | 0.529283 |
| tr11 | 0 | 0.144088 | 0.273546 | 0.413735 | 0.433728 | 0.433728 | 0.433728 | 0.566463 | — | 0.629849 |
| tr12 | 0 | 0.125985 | 0.239534 | 0.344636 | 0.518575 | 0.518575 | 0.518575 | 0.419559 | — | 0.505244 |
| tr23 | 0 | 0.185312 | 0.195889 | 0.278101 | 0.238004 | 0.238004 | 0.238004 | 0.273424 | — | 0.273424 |
| tr31 | 0 | 0.198044 | 0.374028 | 0.258224 | 0.455487 | 0.455487 | 0.257018 | 0.254816 | — | 0.269398 |
| tr41 | 0 | 0.388643 | 0.525827 | 0.512396 | 0.536816 | 0.536816 | 0.536816 | 0.481267 | — | 0.509897 |
| tr45 | 0 | 0.207043 | 0.408448 | 0.423223 | 0.567975 | 0.567975 | 0.567975 | 0.562005 | — | 0.558268 |
| wap | 0 | 0.374950 | 0.461312 | 0.23245 | 0.425698 | 0.425698 | 0.425698 | 0.398608 | — | 0.338850 |
| webkb | 0 | 0.173655 | 0.166699 | 0.303875 | 0.232000 | 0.232000 | 0.232000 | 0.303194 | — | 0.315604 |

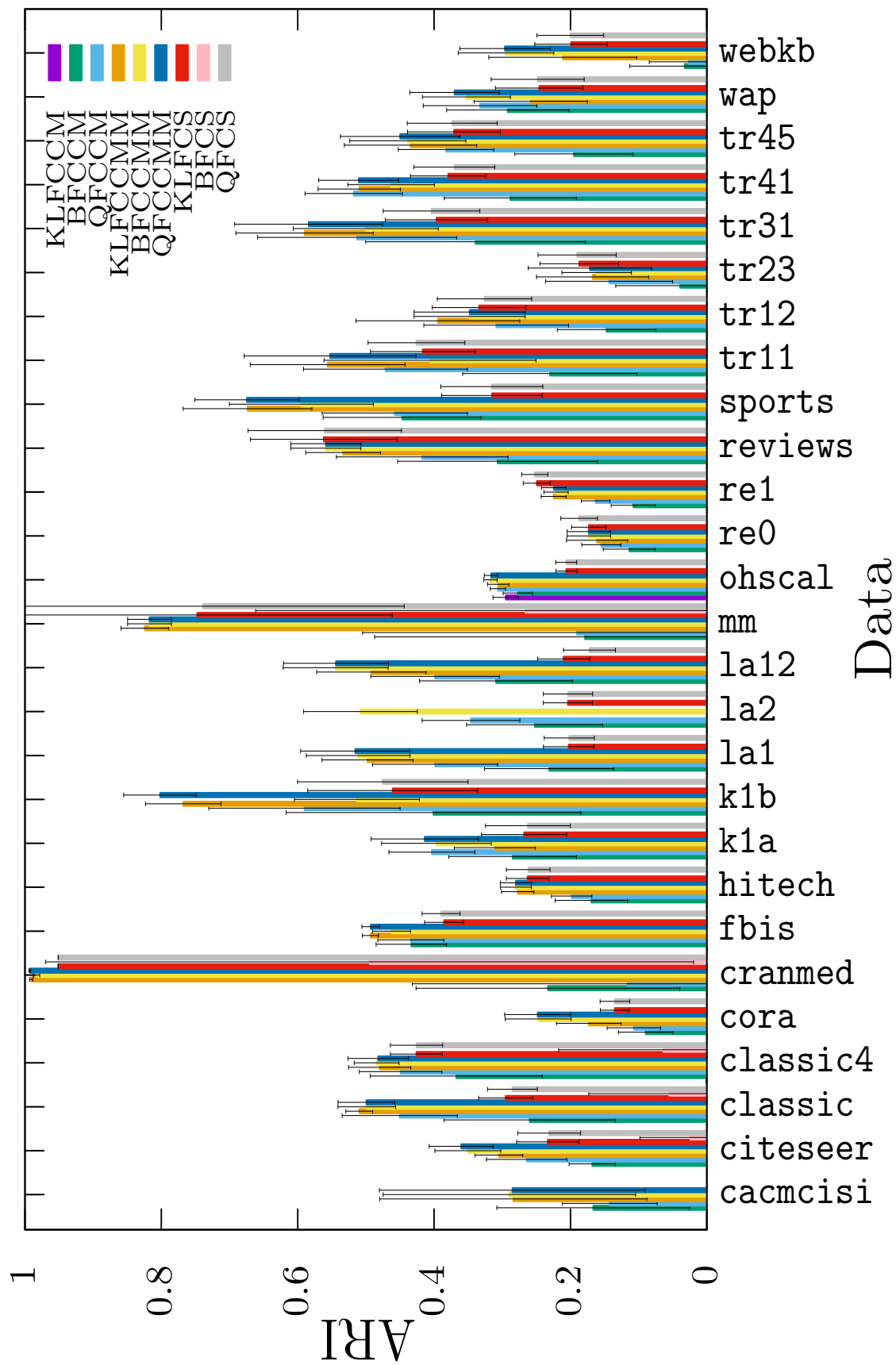


図 3.4: 100 通りの初期値による最高平均 ARI と標準偏差 tf.idf なし

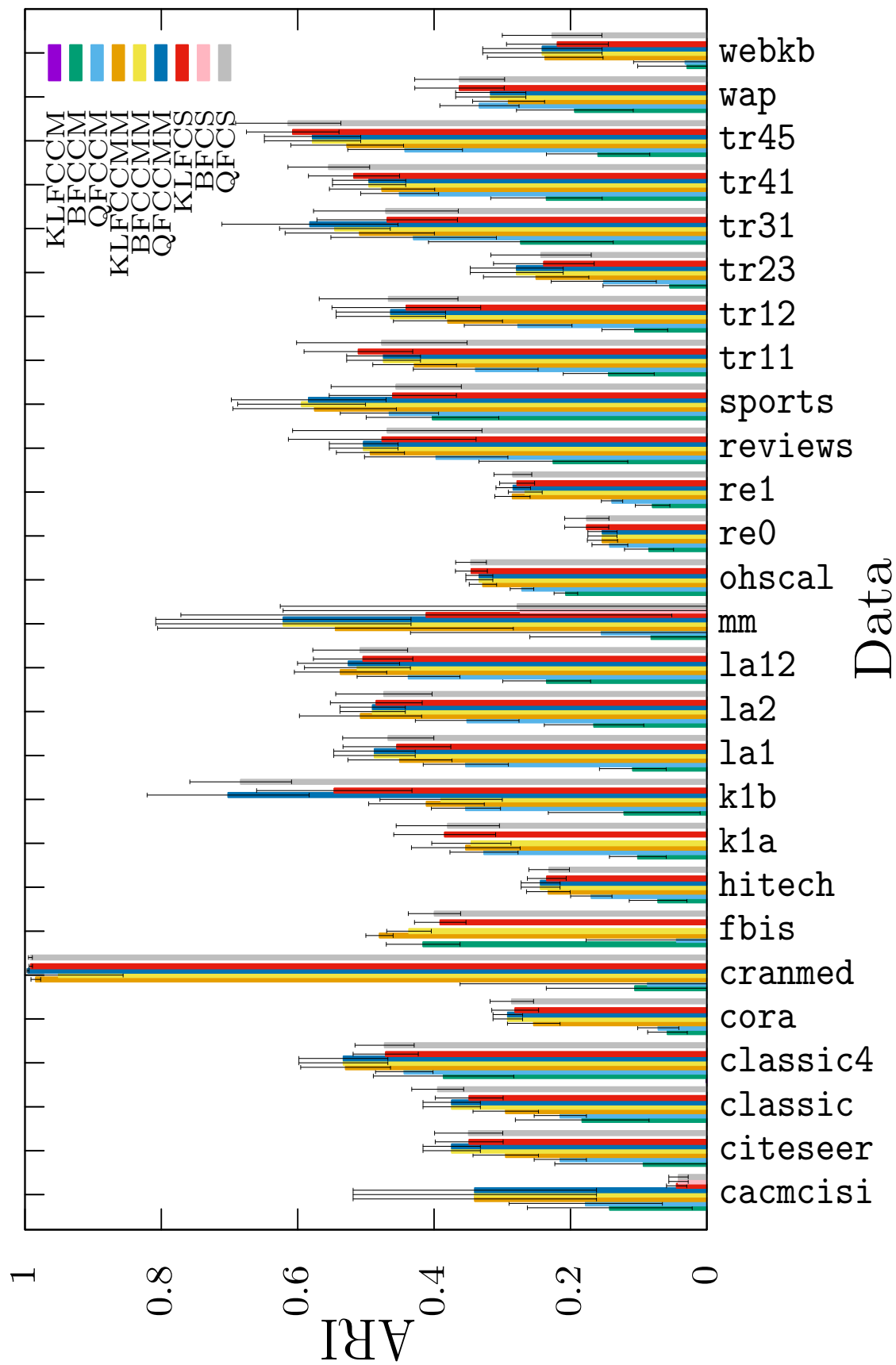


図 3.5: 100 通りの初期値による最高平均 ARI と標準偏差 tf.idfA

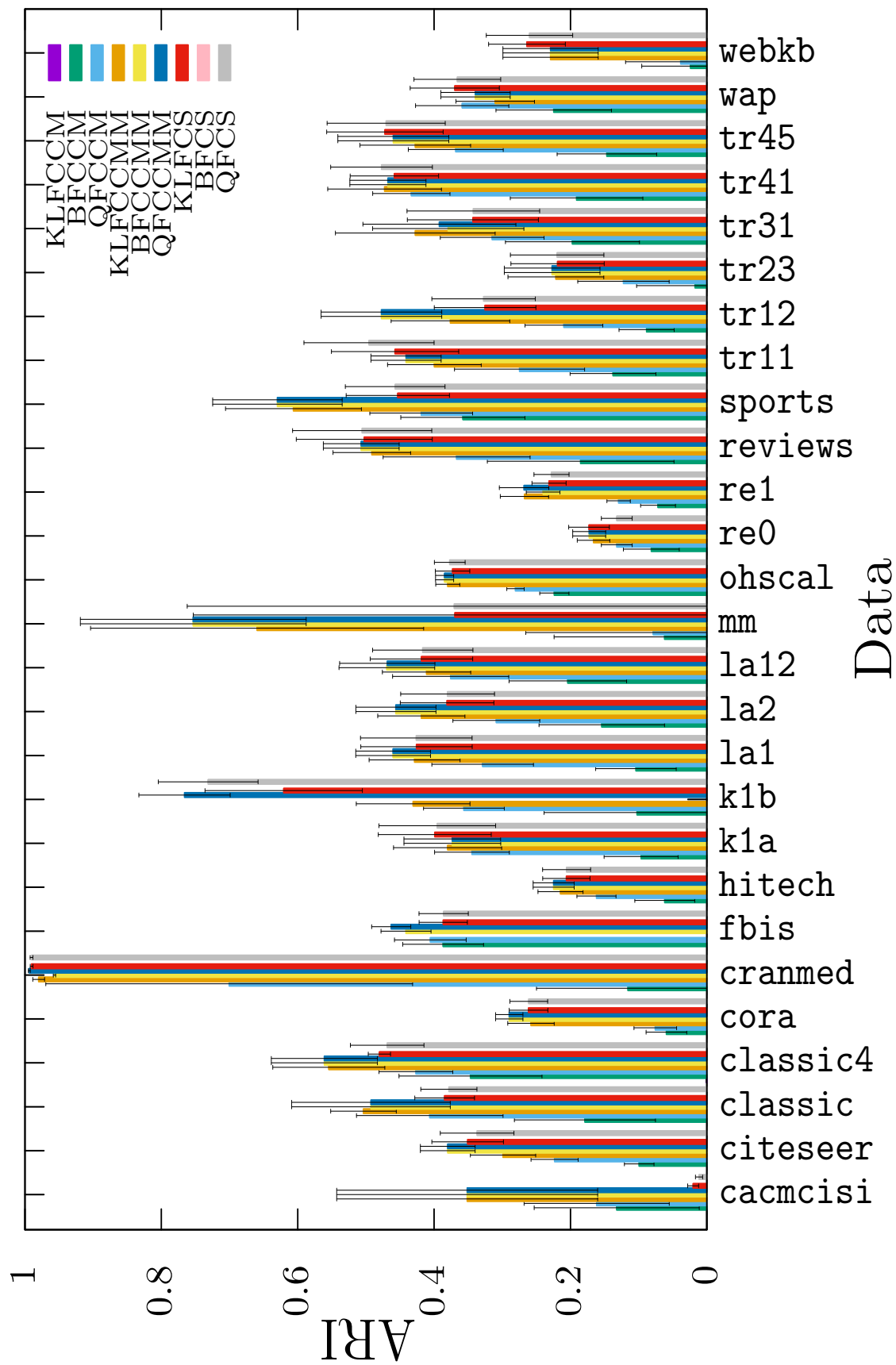


図 3.6: 100 通りの初期値による最高平均 ARI と標準偏差 tf.idfB

3.7.1 数値実験3の結果

表 3.183 に 100 通りの初期設定による平均 ARI の最高値を達成したデータ数を，表 3.184 に目的関数値最小の ARI の最高値を達成したデータ数を示す．初期設定を増やした数値実験 3 においても，tf.idf なし，tf.idfA，tf.idfB において，QFCCMM を適用すると多くのデータで最高平均 ARI を達成し，QFCS や QFCCM と比べて高精度であることを確認した．

表 3.183: 100 通りの初期設定による平均 ARI の最高値を達成したデータ数

| データ 手法 | tf.idf なし | tf.idfA | tf.idfB |
|-----------|-----------|---------|---------|
| KLFCCM | 0 | 0 | 0 |
| BFCCM | 0 | 0 | 0 |
| QFCCM | 1 | 0 | 0 |
| KLFCCMM | 6 | 5 | 3 |
| BFCCMM | 7 | 11 | 15 |
| QFCCMM | 12 | 13 | 18 |
| KLFCS | 1 | 4 | 4 |
| BFCS | 0 | 0 | 0 |
| QFCS | 2 | 5 | 2 |

表 3.184: 目的関数値最小の ARI の最高値を達成したデータ数

| データ 手法 | tf.idf なし | tf.idfA | tf.idfB |
|-----------|-----------|---------|---------|
| KLFCCM | 0 | 0 | 0 |
| BFCCM | 2 | 1 | 0 |
| QFCCM | 2 | 1 | 1 |
| KLFCCMM | 8 | 8 | 4 |
| BFCCMM | 5 | 5 | 12 |
| QFCCMM | 8 | 8 | 13 |
| KLFCS | 2 | 3 | 4 |
| BFCS | 0 | 0 | 0 |
| QFCS | 2 | 5 | 3 |

3.8 考察

数値実験 1 では，表 3.88 から提案法である QFCCMM は，fidf なしの re1 を適用した場合のみ優位性を確認できたが，他のデータを適用したとき QFCCMM は他のファジィクラスタリング手法と比べて精度が劣ることがあった．例えば表 3.90 から tf.idfB を施した new3, re0, re1 を適用した場

合 $\{KL,B\}$ FCCMM に劣ることが確認された。QFCCMM は、パラメータを調整することで理論的に $\{KL,B\}$ FCCMM どちらの性質も包括しているが、この三データにおいて $\{KL,B\}$ FCCMM で最高 ARI を達成するようなパラメータに調整し $\{KL,B\}$ FCCMM の特性に近づけた QFCCMM は更新の途中で計算限界を迎えた。これは、new3, re0, re1 のクラス数が他のデータと比べて多いことや tf.idfB による変換の影響がこの要因だと考えられる。また、この三つのデータでは、 $\lambda = 1.79769 \times 10^{308}$, $m = 1 + 10^{-15}$ と極端にクリスプに分割するパラメータだったことも計算限界の原因だと考えられるため、QFCCMM においてより適当なパラメータ設定が必要である。

数値実験 2 では、表 3.176 から QFCCMM は、tf.idf なし、tf.idfA, tf.idfB それぞれにおいて最高平均 ARI を達成したデータの数が多く、他のファジィクラスタリング手法と比べて高精度であることが確認された。ここで、QFCCMM の精度が BFCCMM より劣っていた tf.idf なしの 1a12 に着目する。表 3.104 における QFCCMM の λ を大きくするほど BFCCMM に近づくが、 $\lambda = 1.79769 \times 10^{308}$ では計算限界であった。よって tf.idf なしの 1a12 に QFCCMM を適用するとき、 λ を 10^3 より大きい 1.79769×10^{308} より小さく処理の行える範囲に設定することで BFCCMM のクラスタリング精度に近づくと期待できる。同様に表 3.94 の tf.idf なしの classic において、KLFCMM と QFCCMM の関係から QFCCMM のパラメータを $\lambda = 10$, $t = 10^{-1}$ と設定し m を 1 に近づけることで KLFCMM の最高平均 ARI に近づくと予想できる。

数値実験 3 では、表 3.183 から、100 通りの初期設定においても QFCCMM が他のファジィクラスタリング手法と比べて高精度であることが確認された。表 3.184 から目的関数値最小の最高 ARI を達成したデータ数は、tf.idf なしと tf.idfA では QFCCMM と KLFCMM を適用した場合で同一数となり、tf.idfB では QFCCMM が最高数を達成した。この結果から平均 ARI だけでなく、目的関数値による初期値選択を適用したクラスタリングにおいても QFCCMM の有効性が確認された。ここで、数値実験 2 において QFCCMM が最高精度であったが、数値実験 3 では他手法より平均 ARI が劣ったデータに着目する。tf.idf なしの結果である図 3.1 と図 3.4 における classic4 を比較すると、他のデータと比べ標準偏差の差が大きいことが視覚的に分かる。これより、数値実験 2 で選択された最適なファジィ化パラメータが、100 通りの初期値設定である数値実験 3 では適当でなかったと考えられる。次に、tf.idf なしの 1a2 の平均 ARI に着目する。QFCCMM を適用した場合、数値実験 2 では 0.488718 で最高平均 ARI を達成したが、数値実験 3 では 15 回目の初期値設定で更新数が 10^4 を超えてクラスタリング不可となった。同様に、QFCCMM を適用した tf.idfA の fbis と k1a においても 100 通りの初期値設定によるクラスタリング結果を得る前に更新数が 10^4 を超えた。したがってこれらのデータにおいて、ファジィ化パラメータの再設定や、実験条件を見直す必要がある。次に、QFCCMM を適用した tf.idfA の classic に着目する。数値実験 2 の図 3.2 と数値実験 3 の図 3.5 を比べると、QFCCMM における classic の平均 ARI と標準偏差が低下していることから、最初の 10 通りの初期値による依存度が大きいと考えられる。tf.idfB において、KLFCMM を適用した fbis や BFCCMM を適用した k1b では、100 通りの初期値による結果を得る前に更新数が 10^4 を超えクラスタリング不可になった一方、QFCCMM を適用した場合、最高精度を達成した。数値実験 2 の結果である表 3.176 と数値実験 3 の結果である表 3.183 を比較すると、tf.idf なしや tf.idfA を適用した場合は QFCCMM の最高平均 ARI を達成したデータ数が少なくなっているが、tf.idfB を適用した場合はその数は同一であった。tf.idfB によるデータ変換によって初期値依存性が緩和されたと考えられる。

3.9 おわりに

本章では、九つのファジィクラスタリング手法 $\{KL, B, Q\}FC\{S, CM, CMM\}$ における文書データを使用したファジィクラスタリングにおける数値実験について述べた。まず CLUTO や LINQS から得た 27 の文書データについて記述した。次に tf.idf による二通りのデータ変換方法を示した。次にクラスタリング精度の指標となる Contingency Table とそれに基づいて算出される ARI について記述した。次に正解の帰属度とクラスタサイズ調整変数を初期値とする数値実験 1 について記述し、その結果を示した。次にランダムなクラスタ中心を 10 通り設定する数値実験 2 について記述し、その結果を示した。次に数値実験 2 で最高平均 ARI を達成したパラメータを採用し、初期値設定数を 100 に変更した数値実験 3 について記述し、その結果を示した。最後に文書データを使用したファジィクラスタリングにおける数値実験 1, 2, 3 の考察を行った。数値実験 1 の結果から、混合多項分布モデルのファジィクラスタリング手法が文書データに対して高精度な分割を行えることを示した。数値実験 2, 3 の結果から、提案法である QFCCMM の有効性を示した。

今後の展望として、まず、クラスタリング不可とみなしたデータに対してパラメータ範囲や収束条件などの実験設定を詳細に見直して、提案アルゴリズムをより堅牢にすることによって、クラスタリングに関する知識のないユーザにも使えるようになるだろう。また、提案法に、可能性理論的アプローチやカーネル法を導入することによってさらなる高精度化や高性能化が期待できる。

第 4 章

協調フィルタリング

4.1 はじめに

本章では、従来の協調フィルタリング手法である FireFly アルゴリズム [37] と GroupLens アルゴリズム [11] を示した後に、ファジィクラスタリングと GroupLens を組み合わせたアルゴリズムを提案する。まず第 4.2 節でユーザ間類似度として用いられるピアソン相関係数とそれに基づいて未評価値を予測する従来の協調フィルタリングについて二手法を記述する。次に第 4.3 節でファジィクラスタリングを協調フィルタリングに応用したアルゴリズムを提案する。最後に 4.4 節で精度評価方法について記述する。評価指標として MAE (Mean Absolute Error), F-measure, ROC 曲線 (Receiver Operating Characteristic) から算出される AUC (Area Under the Curve) の三つを採用した。

4.2 従来の協調フィルタリング

従来法としてユーザ間の類似度から予測する二手法を採用した。ユーザの既評価値と類似度から求める方法を FireFly とし、FireFly にユーザ既評価平均値を考慮に入れた方法を GroupLens とする。ユーザ $\#k$ に対するアイテム $\#\ell$ の評価値行列 $x_{k,\ell}$ ($k = \{1, \dots, N\}, \ell = \{1, \dots, M\}$) を用いて未評価値の予測を行う。数値実験では既評価をランダムに欠損させ、欠損させた評価値を予測し正解の評価値と比較することで精度評価を行う。ユーザ $\#k$ の既評価平均値を $\bar{x}_{k,\cdot}$ とし、ユーザ $\#k$ と $\#k'$ の類似度を $\text{sim}(k, k')$ とする。またユーザ $\#k$ に対するアイテム $\#\ell$ が既評価である場合 1, 未評価である場合 0 である既評価行列を $h \in \mathbb{R}^{N \times M}$ とする。

$$\text{既評価行列} : h_{k,\ell} = \begin{cases} 1 & (x_{k,\ell} \text{が評価済}), \\ 0 & (x_{k,\ell} \text{が未評価}). \end{cases} \quad (4.1)$$

ユーザ $\#k$ と $\#k'$ の類似度はピアソン相関係数を用いる。ピアソン相関係数は $-1 \leq \text{sim}(k, k') \leq 1$ であり 1 に近いほどユーザ $\#k$ と $\#k'$ の類似性が高い。以下にピアソン相関係数の計算式を示す：

$$\text{sim}(k, k') = \frac{\sum_{\ell: h_{k,\ell} h_{k',\ell} = 1} (x_{k,\ell} - \bar{x}_{k,\cdot})(x_{k',\ell} - \bar{x}_{k',\cdot})}{\sqrt{\sum_{\ell: h_{k,\ell} h_{k',\ell} = 1} (x_{k,\ell} - \bar{x}_{k,\cdot})^2} \sqrt{\sum_{\ell: h_{k,\ell} h_{k',\ell} = 1} (x_{k',\ell} - \bar{x}_{k',\cdot})^2}}. \quad (4.2)$$

ここで、 $\sum_{\ell: h_{k,\ell} h_{k',\ell} = 1}$ はユーザ $\#k$ と $\#k'$ どちらもが評価済みのアイテム $\#\ell$ のみ計算に考慮されることを意味する。効率的な処理を行うため式変形を行う [38]。まず、式 (4.2) の分母の因子を以下のよ

うに変形する：

$$\begin{aligned} & \sqrt{\sum_{\ell:h_{k,\ell}h_{k',\ell}=1} (x_{k,\ell} - \bar{x}_{k,\cdot})^2} \\ &= \sqrt{\sum_{\ell:h_{k,\ell}h_{k',\ell}=1} x_{k,\ell}^2 - 2\bar{x}_{k,\cdot} \sum_{\ell:h_{k,\ell}h_{k',\ell}=1} x_{k,\ell} + \bar{x}_{k,\cdot}^2 \sum_{\ell:h_{k,\ell}h_{k',\ell}=1} 1}. \end{aligned} \quad (4.3)$$

ここで、 M' をユーザ $\#k$ と $\#k'$ のお互いの既評価数し、 $\sum_{\ell:h_{k,\ell}h_{k',\ell}=1}$ を \sum_{ℓ} と省略する． $\bar{x}_{k,\cdot} = \frac{1}{M'} \sum_{\ell} x_{k,\ell}$ より式 (4.3) は以下のように変形できる：

$$\begin{aligned} & \sqrt{\sum_{\ell} x_{k,\ell}^2 - 2\bar{x}_{k,\cdot} \sum_{\ell} x_{k,\ell} + \bar{x}_{k,\cdot}^2 \sum_{\ell} 1} \\ &= \sqrt{\sum_{\ell} x_{k,\ell}^2 - 2\bar{x}_{k,\cdot} \times M' \bar{x}_{k,\cdot} + \bar{x}_{k,\cdot}^2 \times M'} \\ &= \sqrt{\sum_{\ell} x_{k,\ell}^2 - M' \bar{x}_{k,\cdot}^2}. \end{aligned} \quad (4.4)$$

次に式 (4.2) の分子を以下のように変形する：

$$\begin{aligned} & \sum_{\ell} (x_{k,\ell} - \bar{x}_{k,\cdot})(x_{k',\ell} - \bar{x}_{k',\cdot}) \\ &= \sum_{\ell} x_{k,\ell}x_{k',\ell} - \sum_{\ell} x_{k,\ell}\bar{x}_{k',\cdot} - \sum_{\ell} x_{k',\ell}\bar{x}_{k,\cdot} + \sum_{\ell} \bar{x}_{k,\cdot}\bar{x}_{k',\cdot} \\ &= \sum_{\ell} x_{k,\ell}x_{k',\ell} - M'\bar{x}_{k,\cdot}\bar{x}_{k',\cdot} - M'\bar{x}_{k',\cdot}\bar{x}_{k,\cdot} + M'\bar{x}_{k,\cdot}\bar{x}_{k',\cdot} \\ &= \sum_{\ell} x_{k,\ell}x_{k',\ell} - M'\bar{x}_{k,\cdot}\bar{x}_{k',\cdot}. \end{aligned} \quad (4.5)$$

最終的に式 (4.2) は、式 (4.4)、式 (4.5) から以下のように変形される：

$$\text{sim}(k, k') = \frac{\sum_{\ell} x_{k,\ell}x_{k',\ell} - \frac{1}{M'} \sum_{\ell} x_{k,\ell} \sum_{\ell} x_{k',\ell}}{\sqrt{\sum_{\ell} x_{k,\ell}^2 - \frac{1}{M'} (\sum_{\ell} x_{k,\ell})^2} \sqrt{\sum_{\ell} x_{k',\ell}^2 - \frac{1}{M'} (\sum_{\ell} x_{k',\ell})^2}}. \quad (4.6)$$

$\sum_{k':\text{sim}(k,k')>0}$ はピアソン相関係数が 0 より大きいユーザ情報のみ計算に扱うことを意味する．ユーザ $\#k$ のアイテム $\#l$ に対する FireFly の予測評価値 $\hat{x}_{k,\ell}$ を以下に示す：

$$\hat{x}_{k,\ell} = \frac{\sum_{k':\text{sim}(k,k')>0} h_{k',\ell}x_{k',\ell}\text{sim}(k, k')}{\sum_{k':\text{sim}(k,k')>0} h_{k',\ell}\text{sim}(k, k')}. \quad (4.7)$$

以上より、FireFly は次のアルゴリズムにまとめられる：

アルゴリズム 8 (FireFly).

STEP 1. 式 (4.6) に従いピアソン相関係数を算出する．

STEP 2. 式 (4.7) に従い予測評価値を算出する． ■

ユーザ $\#k$ のアイテム $\#l$ に対する GroupLens の予測評価値 $\hat{x}_{k,\ell}$ を以下に示す：

$$\hat{x}_{k,\ell} = \bar{x}_{k,\cdot} + \frac{\sum_{k':\text{sim}(k,k')>0} h_{k',\ell}\text{sim}(k, k')(x_{k',\ell} - \bar{x}_{k',\cdot})}{\sum_{k':\text{sim}(k,k')>0} h_{k',\ell}\text{sim}(k, k')}. \quad (4.8)$$

以上より，GroupLens は次のアルゴリズムにまとめられる：

アルゴリズム 9 (GroupLens).

STEP 1. 式 (4.6) に従いピアソン相関係数を算出する.

STEP 2. 式 (4.8) に従い予測評価値を算出する. ■

4.3 提案法

$f(x_{k'}) \equiv f(x_k)$ は，ユーザ $\#k$ と $\#k'$ が同じクラスタに属することを表す．クラスタリングと GroupLens を組み合わせた提案法は以下のように示される：

$$\hat{x}_{k,\ell} = \bar{x}_{k,\cdot} + \frac{\sum_{\substack{k': \text{sim}(k,k') > 0, \\ f(x_{k'}) \equiv f(x_k)}} \text{sim}(k, k')(x_{k',\ell} - \bar{x}_{k',\cdot})}{\sum_{\substack{k': \text{sim}(k,k') > 0, \\ f(x_{k'}) \equiv f(x_k)}} \text{sim}(k, k')}. \quad (4.9)$$

以上より，提案法は次のアルゴリズムにまとめられる：

アルゴリズム 10 (提案法).

STEP 1. 式 (4.6) に従いピアソン相関係数を算出する.

STEP 2. アルゴリズム 7 に従いユーザをクラスタリングする.

STEP 3. 式 (4.9) に従い予測評価値を算出する. ■

ここで，STEP 2. におけるアルゴリズム 7 では QFCCMM が応用されたが， $\{B, KL, Q\}FC$ $\{S, CM, CMM\}$ でも同様に適用できる．

4.4 精度評価方法

4.4.1 Mean Absolute Error (MAE)

MAE は予測評価値 $\hat{x}_{k,\ell}$ と真の評価値 $x_{k,\ell}$ の絶対誤差の平均であり W を欠損数とすることで，式 (4.10) によって求めることができる．MAE は 0 に近いほど精度が高い．

$$\text{MAE} = \frac{\sum_{k=1}^N \sum_{\ell=1}^M |\hat{x}_{k,\ell} - x_{k,\ell}|}{W} \quad (4.10)$$

4.4.2 F-measure

設定した閾値に基づき，推薦すべきアイテムを推薦した数を TP (True Positive)，推薦すべきでないアイテムを推薦した数を FP (False Positive)，推薦すべきアイテムを推薦しなかった数を FN (False Negative)，推薦すべきでないアイテムを推薦しなかった数を TN True Negative とする．表 4.1 にこれらをまとめたスレットスコアを示す．Precision (適合率) とは推薦したアイテムのうちの推薦すべきアイテム数であり，式 (4.11) によって求めることができる．Recall (再現率) とは推薦すべきアイテムのうちの推薦したアイテム数であり，式 (4.12) によって求めることができる．これら二つは割合であるため

調和平均によって F-measure を求める．推薦システムでは，Precision と Recall どちらの精度も重要視されるため，単純に加算平均ではなく調和平均が採用される．例えば，Precision = 1, Recall = 0 であるとき，加算平均の場合は F-measure = 0.5 であるが，調和平均であれば F-measure = 0 となる．いくら Precision が良かったとしても Recall が 0 であれば，その推薦システムの評価は低くなる．したがって F-measure には調和平均が適用される．F-measure の計算式を式 (4.13) に示す．F-measure $\in [0, 1]$ であり，1 に近いほど精度が高い．

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.12)$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.13)$$

表 4.1: スレットスコア

| | 推薦すべき | 推薦すべきでない |
|---------|-------|----------|
| 推薦した | TP | FP |
| 推薦しなかった | FN | TN |

4.4.3 Receiver Operating Characteristic 曲線 (ROC 曲線)

縦軸に Recall を，横軸に推薦すべきでないアイテムのうちの推薦されたアイテムの割合を取り，閾値を変えてプロットをすると曲線を描くことができる．横軸を Fallout と呼び，以下の式で求まる：

$$\text{Fallout} = \frac{FP}{FP + TN} \quad (4.14)$$

この曲線の内側の面積が 1 に近いほど推薦精度が高く，この面積を AUC (Are under the curve) という [39, 40]．この AUC の値で精度比較を行う．図 4.1 に ROC 曲線の例を示す．

4.5 おわりに

本章では，協調フィルタリングについて従来法を述べた後に，ファジィクラスタリングを援用して新たな協調フィルタリング手法を提案した．まず第 4.2 節でユーザ間類似度として用いられるピアソン相関係数とそれに基づいて未評価値を予測する従来の協調フィルタリングである GroupLens と FireFly について記述した．次に第 4.3 節でファジィクラスタリングを協調フィルタリングに応用した提案法を示した．最後に第 4.4 節で協調フィルタリングの精度評価方法である MAE, F-measure, ROC 曲線について記述した．

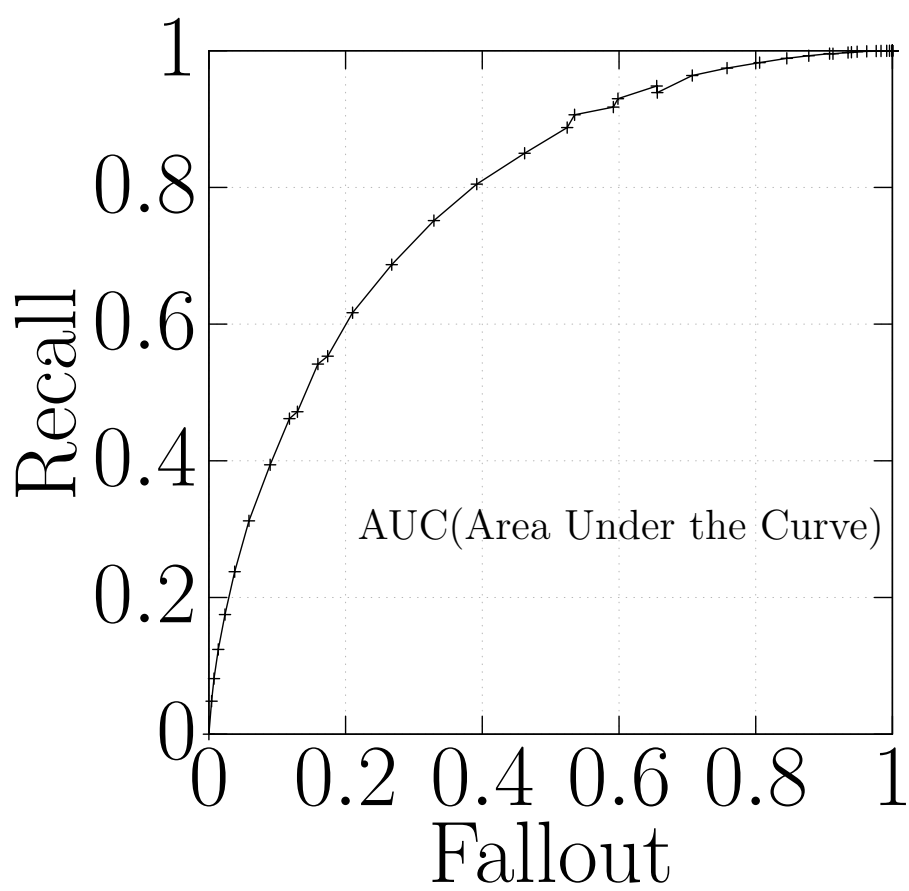


図 4.1: ROC 曲線の例

第 5 章

協調フィルタリングにおける数値実験

5.1 はじめに

本章では、評価データを使用した協調フィルタリングにおける数値実験について記述する。第 2 章で述べた九つのファジィクラスタリング手法 $\{B, KL, Q\}FC\{S, CM, CMM\}$ を提案法に適用する。まず人工データを使用した数値実験の設定、その結果と考察について述べる。100 人のユーザが 100 個のアイテムを評価したと見立てた人工データを生成し、ランダムに既評価値を欠損させ、従来法及び提案法によって欠損箇所を推定し、正解解の要素と比較をすることで精度評価を行う。人工データ数値実験では、ファジィクラスタリングの初期値に帰属度とクラスサイズ調整変数を与えることで初期値依存性を解消し、ファジィクラスタリングを応用した協調フィルタリングの有効性を確認する。次に実データを使用した数値実験について述べる。ユーザがアイテムを評価した六つの評価データを紹介し、データ整形方法を記す。次に実データ数値実験の設定と結果について示す。人工データ数値実験と同様に既評価である要素をランダムに欠損させ予測することで精度評価を行う。実データ数値実験では、ファジィクラスタリングの初期値にランダムなクラス中心とクラスサイズ調整を与えることで、初期値依存性による精度差を検証する。最後に協調フィルタリングにおける数値実験のまとめを記述する。

5.2 人工データ数値実験

5.2.1 設定

人工データとして行をユーザ、列をアイテムと見立て 100 行 100 列の評価値行列を生成した。表 5.1 に生成した人工データを示す。人工データは、一つのクラスに対して同じ既評価値を持った 20 ユーザから構成され、5 クラスになるようにした。協調フィルタリングにおける数値実験では、既評価値をランダムに欠損させ、予測した評価値と正解の評価値から精度比較を行う。人工データ数値実験では、精度の差が顕著に現れる MAE で比較を行う。欠損数を 500 から 7,500 まで 500 刻みで増やし、10 通りの欠損パターンの MAE を算出する。欠損箇所を決める乱数シード値は、0 から設定された数を欠損させるまでインクリメントすることで変化させる。データの欠損後に FCS の三手法は式 (3.8) に従い球面化し、FCCM と FCCMM の六手法では式 (3.9) に従い単体化することでデータの正規化を行いクラスタリングする。式 (5.1)、式 (5.2) に従いファジィな帰属度とクラスサイズ調整変数を初期値として与える。表 5.2 は、 $C = 5$ としたときの初期帰属度を表す。表 5.3 に示すパラメータ範囲で数値実験を行う。更新前と更新後の帰属度、クラス中心、クラスサイズ調整変数それぞれの差における最大値ノルムの和が 10^{-10} を下回るとき収束したとみなす。更新回数を 10^3 を超えても収束しない場合や、更新

式計算の途中で0除算や無限大が発生した場合は、そのパラメータでの処理を終えてクラスタリング不可であったとみなす。設定された欠損数ごとの10通りの平均MAEを算出することで手法間の精度比較を行う。

$$u_{i,k}^{\text{Init}} = \begin{cases} 1 - (C - 1)/100 & (k \bmod C = i), \\ 0.01 & (\text{それ以外のとき}). \end{cases} \quad (5.1)$$

$$\pi_i^{\text{Init}} = \frac{1}{C} \quad (5.2)$$

表 5.1: 人工データ

| Item \ User | 1 | ... | 20 | 21 | ... | 40 | 41 | ... | 60 | 61 | ... | 80 | 81 | ... | 100 |
|-------------|---|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|-----|
| 1 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 20 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 |
| 21 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 40 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 |
| 41 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 60 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 | 3 | ... | 3 |
| 61 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 80 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 | 2 | ... | 2 |
| 81 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 100 | 2 | ... | 2 | 3 | ... | 3 | 4 | ... | 4 | 5 | ... | 5 | 1 | ... | 1 |

表 5.2: 式 (5.1) において $C = 5$ としたときの初期帰属度 $u_{i,k}^{\text{Init}}$

| $u_{i,k}^{\text{Init}}$ $i \backslash k$ | 0 | 1 | 2 | 3 | 4 | 5 | \dots | 99 |
|---|------|------|------|------|------|------|---------|------|
| 0 | 0.96 | 0.01 | 0.01 | 0.01 | 0.01 | 0.96 | \dots | 0.01 |
| 1 | 0.01 | 0.96 | 0.01 | 0.01 | 0.01 | 0.01 | \dots | 0.01 |
| 2 | 0.01 | 0.01 | 0.96 | 0.01 | 0.01 | 0.01 | \dots | 0.01 |
| 3 | 0.01 | 0.01 | 0.01 | 0.96 | 0.01 | 0.01 | \dots | 0.01 |
| 4 | 0.01 | 0.01 | 0.01 | 0.01 | 0.96 | 0.01 | \dots | 0.96 |

以下に協調フィルタリングにおける人工データ数値実験の流れを示す：

表 5.3: 人工データを使用した数値実験で採用したパラメータ範囲

| 手法 \ パラメータ | λ | m | t |
|------------|--|-----------------------|--------------------|
| KLFCS | {1.0, 10, 35, 100} | | |
| BFCS | | {1.01, 1.3, 1.4, 1.5} | |
| QFCS | {1.0, 10, 35, 100} | {1.01, 1.3, 1.4, 1.5} | |
| KLFCCM | {1.0, 10, 100} | | |
| BFCCM | | {1.008, 1.009, 1.01} | |
| QFCCM | {1.0, 10, 100} | {1.008, 1.009, 1.01} | |
| KLFCCMM | {1.0, 10, 14, 100} | | {0.001, 0.01, 0.1} |
| BFCCMM | | {1.008, 1.009, 1.01} | {0.001, 0.01, 0.1} |
| QFCCMM | {1.0, 10, 100} | {1.008, 1.009, 1.01} | {0.001, 0.01, 0.1} |
| クラスタ数 | $C \in \{2, 3, \dots, 10, 15, \dots, 50\}$ | | |

実験の流れ 4 (協調フィルタリングにおける人工データ数値実験).

- STEP 1. 表 5.1 に示すデータの読み込み.
- STEP 2. 表 5.3 の範囲でパラメータを設定する.
- STEP 3. 欠損数を設定する.
- STEP 4. 既評価値を欠損する.
- STEP 5. $\{B, KL, Q\}FCS$ では式 (3.8) に従い球面化し, $\{B, KL, Q\}FC\{CM, CMM\}$ では式 (3.9) に従い単体化する.
- STEP 6. 式 (5.1), 式 (5.2) に従い初期値を設定する.
- STEP 7. BFCS では式 (2.90), KLFCS では式 (2.104), QFCS では式 (2.118), KLFCCM では式 (2.140), BFCCM では式 (2.168), QFCCM では式 (2.154), KLFCCMM では式 (2.184), BFCCMM では式 (2.198), QFCCMM では式 (2.212) に従いクラスタ中心を更新する.
- STEP 8. $\{B, KL, Q\}FCS$ では式 (2.78), $\{KL, Q, B\}FCCM$ では式 (2.128), $\{KL, B, Q\}FCCMM$ では式 (2.172) に従いオブジェクト-クラスタ間非類似度を更新する.
- STEP 9. BFCS では式 (2.88), KLFCS では式 (2.102), QFCS では式 (2.116), KLFCCM では式 (2.138), BFCCM では式 (2.166), QFCCM では式 (2.152), KLFCCMM では式 (2.182), BFCCMM では式 (2.196), QFCCMM では式 (2.210) に従い帰属度を更新する.
- STEP 10. BFCS では式 (2.92), KLFCS では式 (2.106), QFCS では式 (2.120), KLFCCM では式 (2.142), BFCCM では式 (2.170), QFCCM では式 (2.156), KLFCCMM では式 (2.186), BFCCMM では式 (2.200), QFCCMM では式 (2.214) に従いクラスタサイズ調整変数を更新する.
- STEP 11. 更新前と更新後の帰属度, クラスタ中心, クラスタサイズ調整変数の最大値ノルムを計算し,

それらの和が 10^{-10} 以上であれば STEP 7. へ、更新数が 10^3 を超えた場合、もしくは計算限界を迎えた場合はそのパラメータでのクラスタリングを不可として STEP 17. へ。

STEP 12. 収束した帰属度をクリスプ化する。

STEP 13. アルゴリズム 10 に従って欠損された評価値を予測する。

STEP 14. 予測値と正解の評価値を照らし合わせ式 4.10 に従い MAE を算出する。

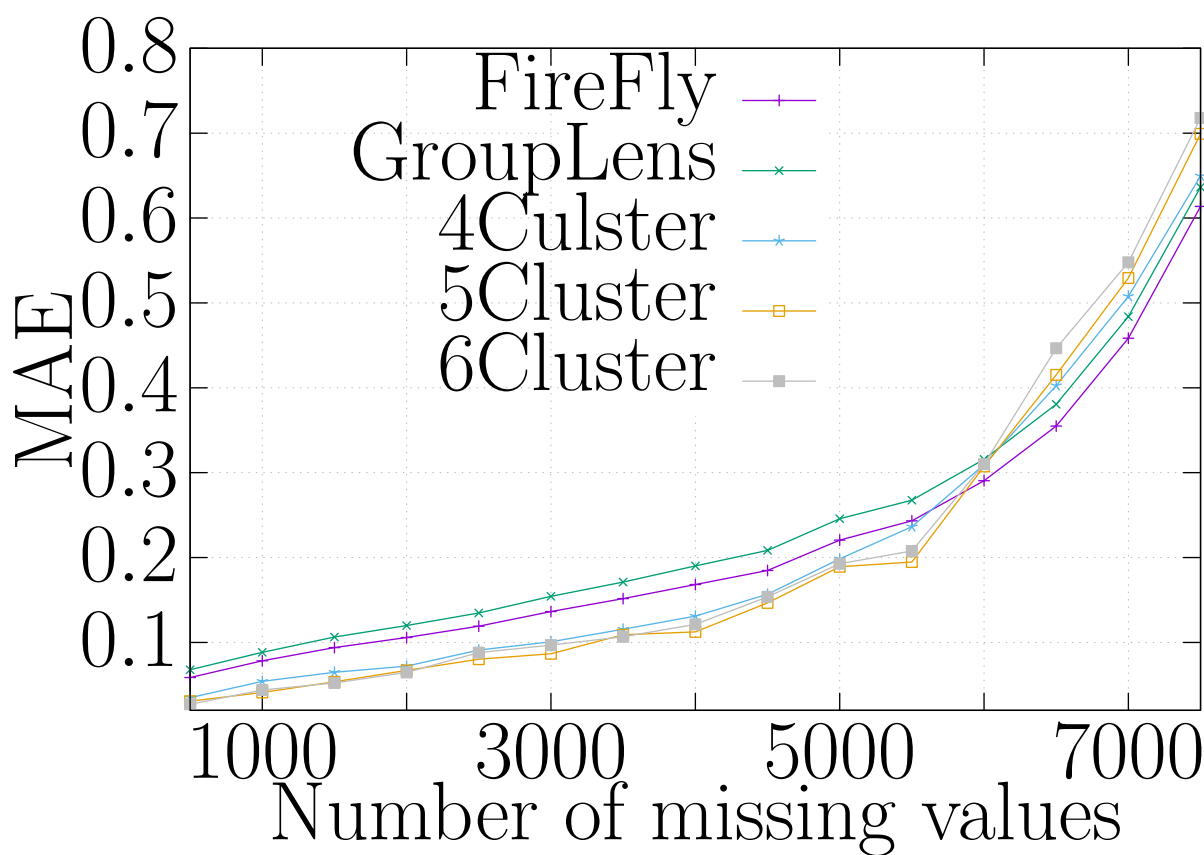
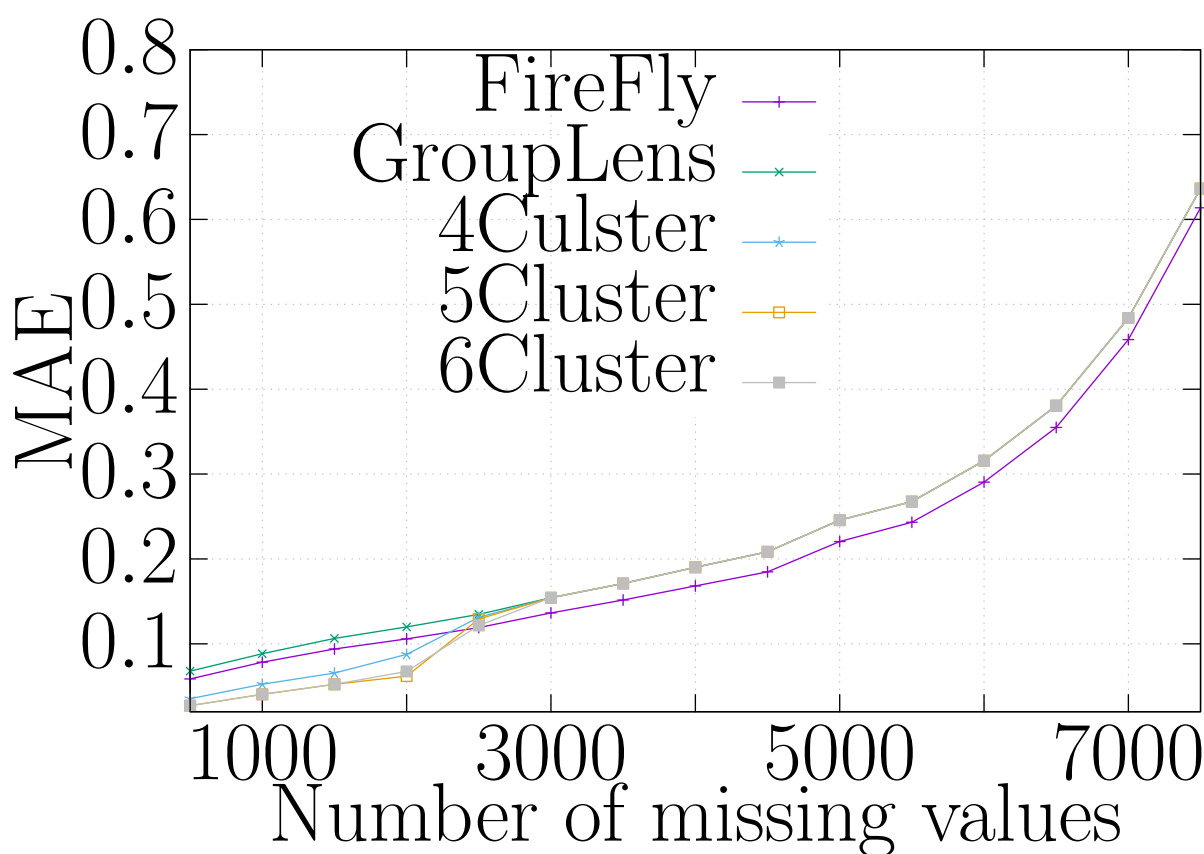
STEP 15. 10 通りの欠損パターンで結果を得たら、その平均を算出する。そうでなければ欠損パターン発生乱数シード値をインクリメントし STEP 4. へ。

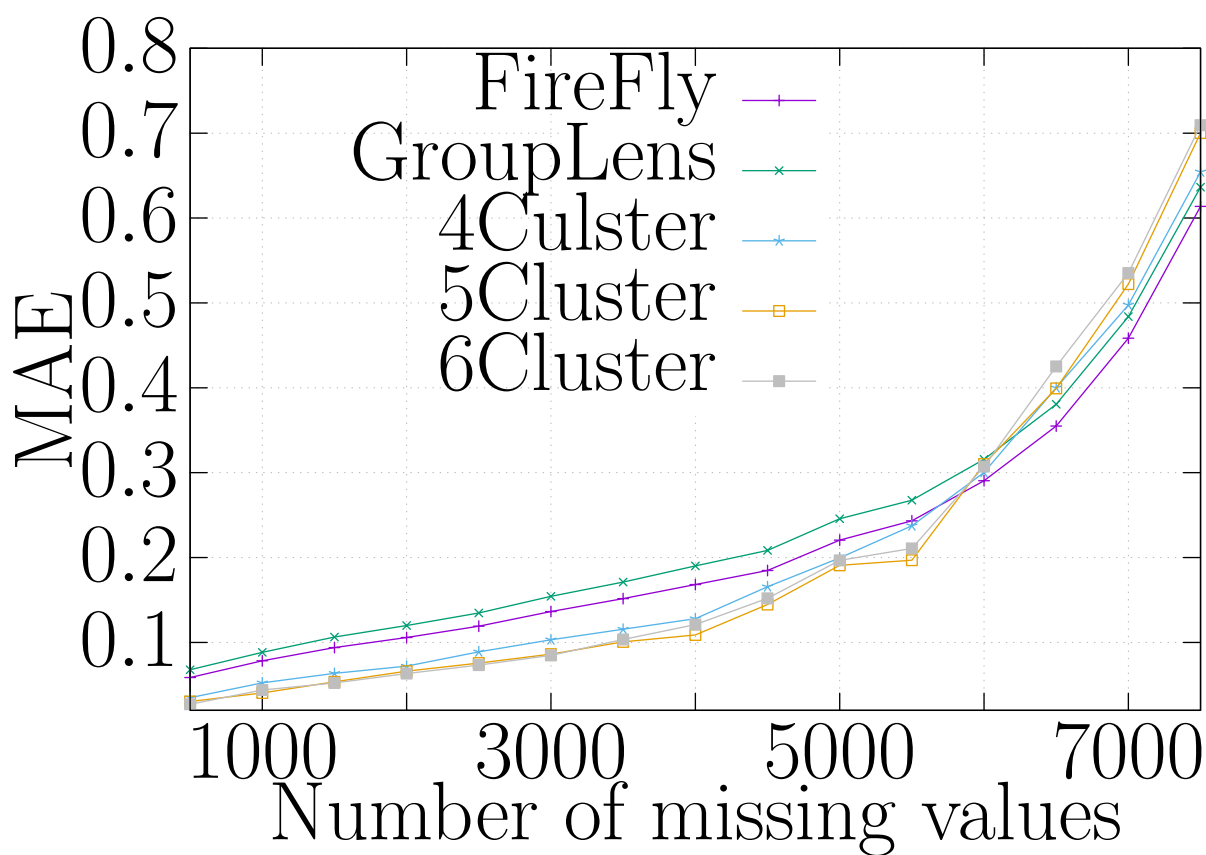
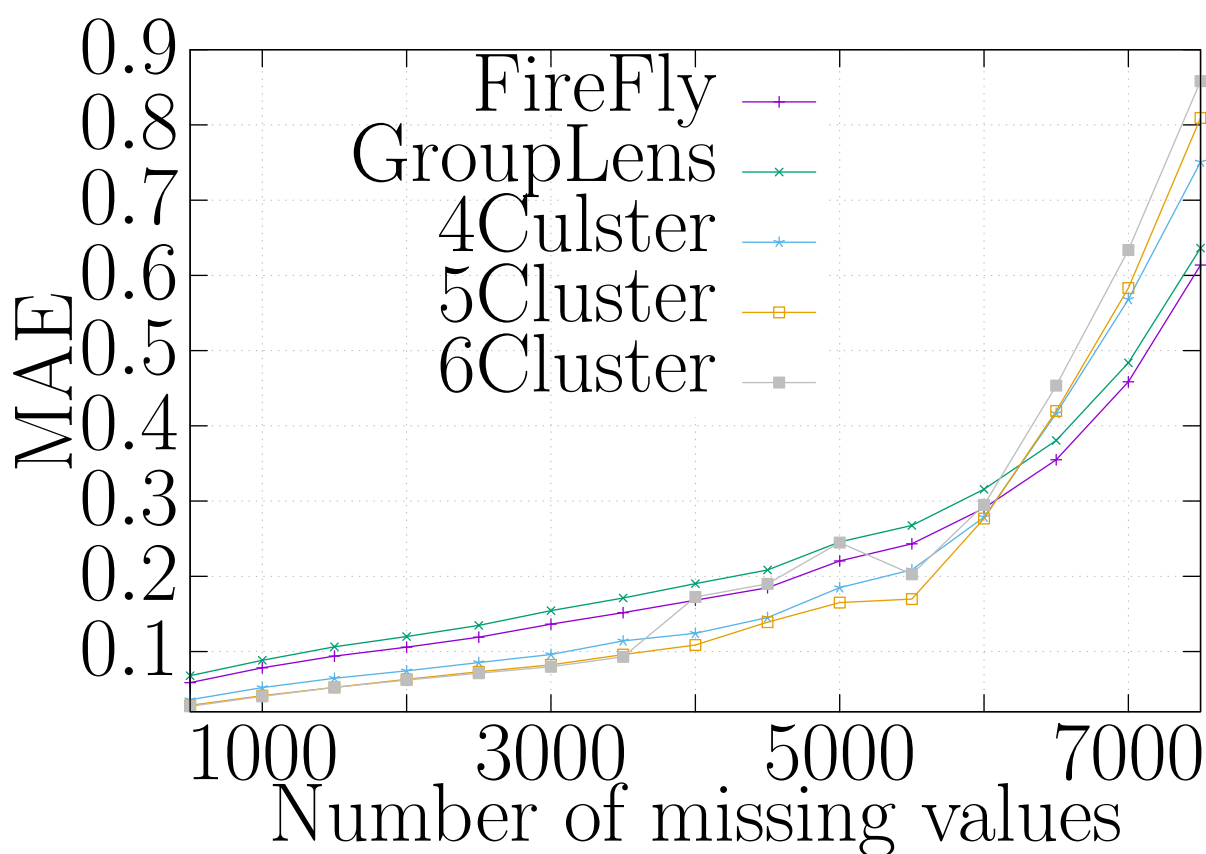
STEP 16. 欠損数が 7,500 でなければ欠損数を 500 増やし STEP 3. へ。

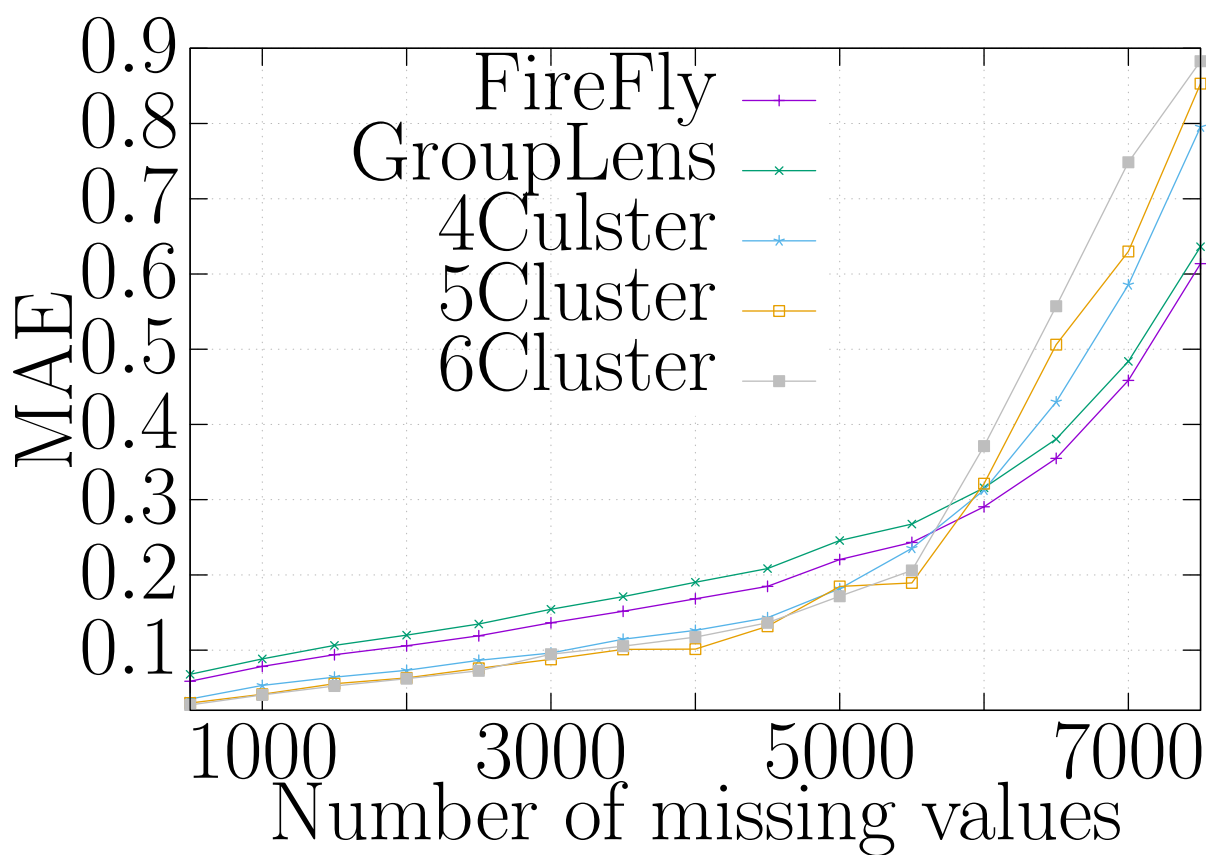
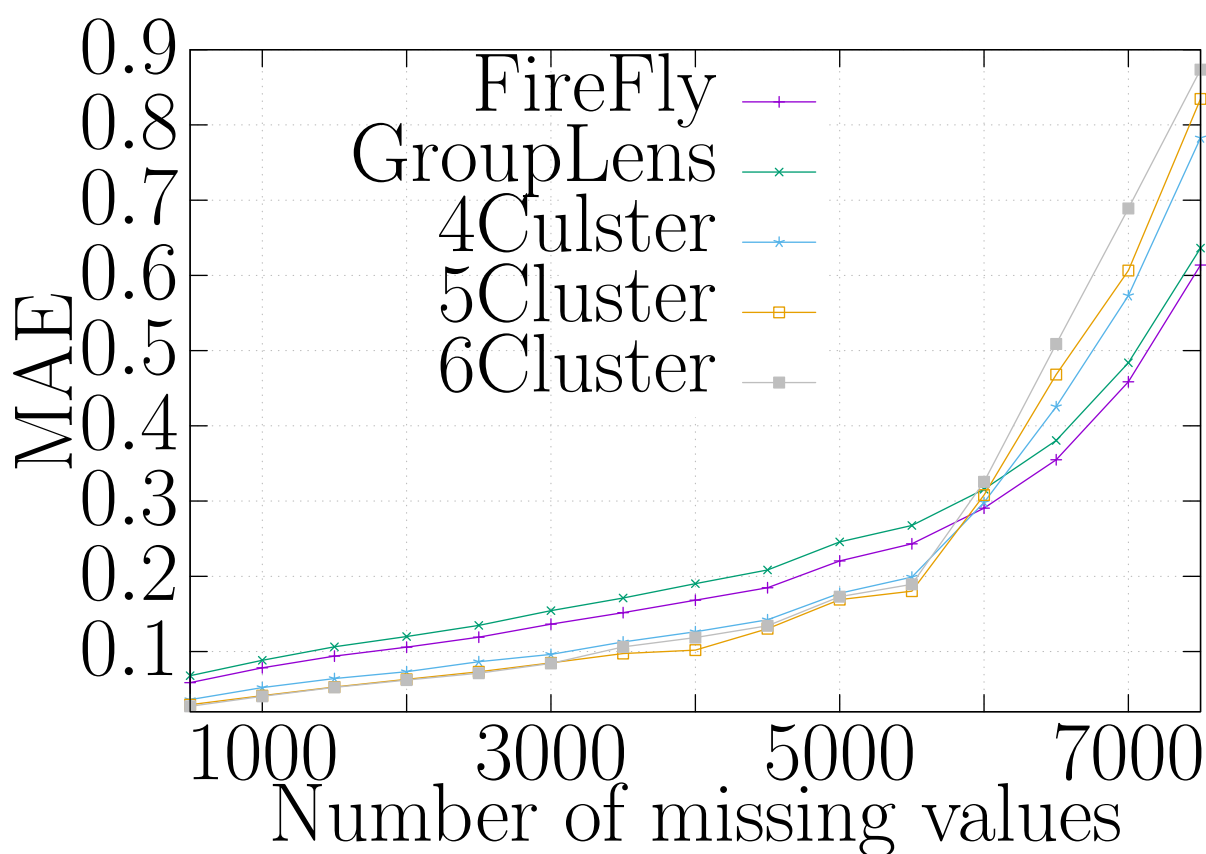
STEP 17. 次に設定すべきパラメータがあれば STEP 2. へ。そうでなければ終了。 ■

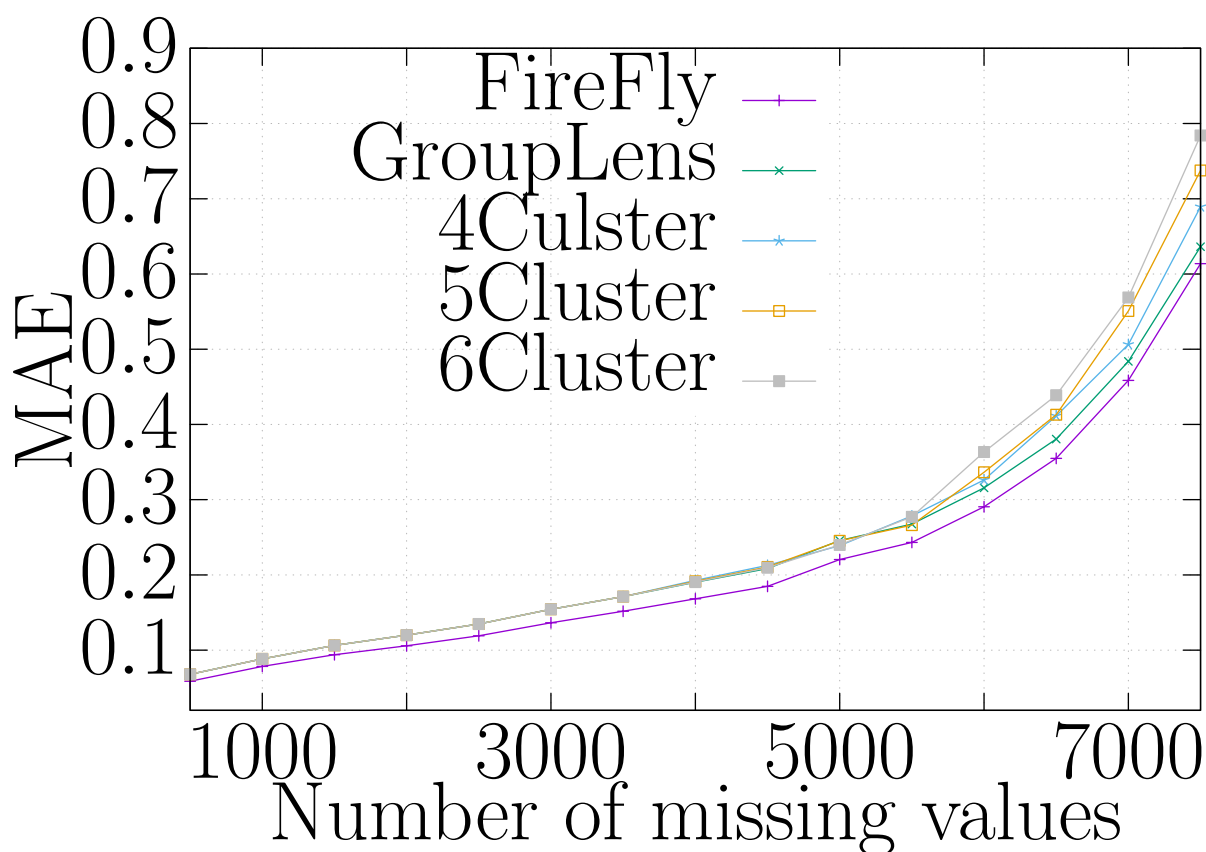
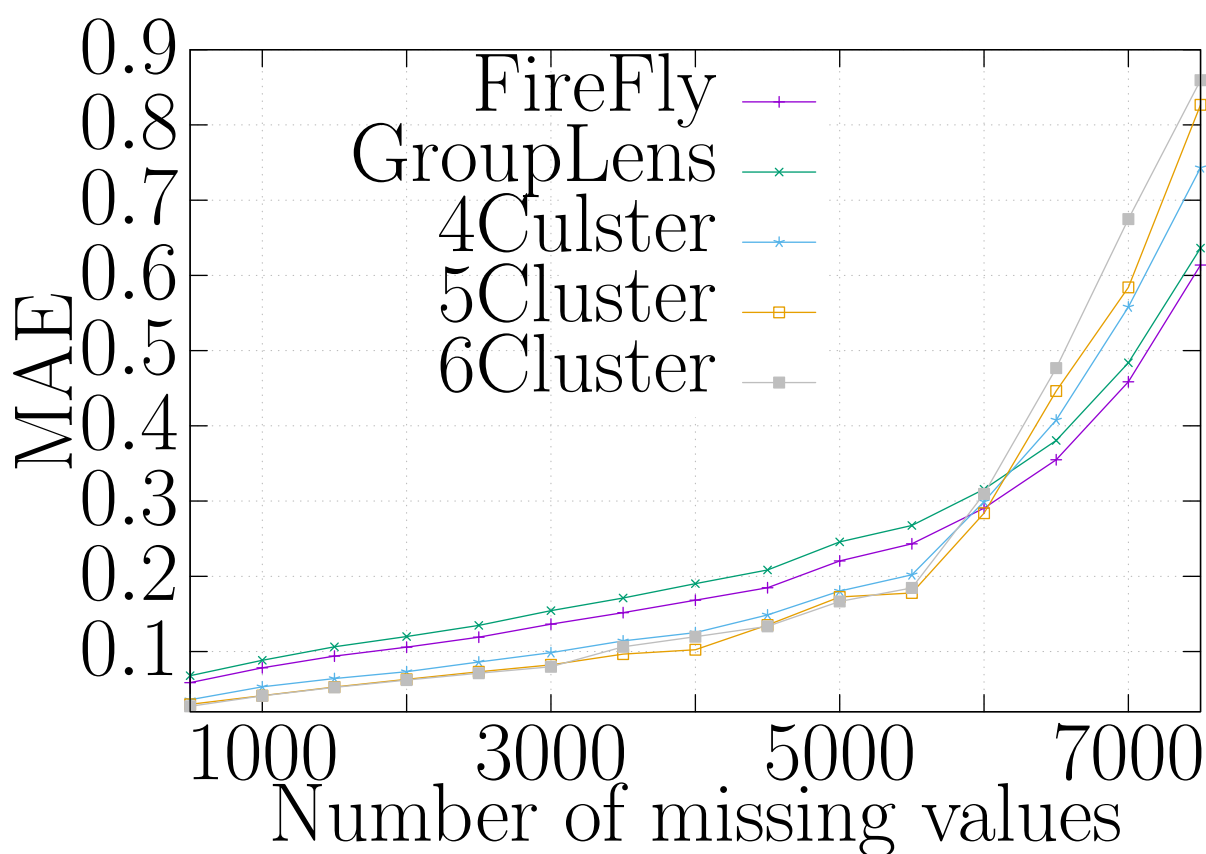
5.2.2 結果と考察

欠損数 5,500 における 10 通りの欠損パターンの平均 MAE が最小のパラメータを採用し、そのパラメータでの欠損数 500 から 7,500 における平均 MAE の遷移を各手法ごとに調べた。正解のクラスタ数である $C = 5$ とその前後の $C = \{4, 6\}$ の結果を図 5.1 から図 5.9 に示す。図 5.1 は、KLFCS におけるパラメータを $\lambda = 35$ とした平均 MAE の結果である。KLFCS を適用した提案法は、欠損数 5,500 以下では従来法の二手法より高精度だが、欠損数 6,000 以上では FireFly より精度が劣り、欠損数 6,500 以上では従来法の二手法より精度が劣った。また欠損数が 6,000 のときを境に正解のクラスタ数である $C = 5$ より $C = 4$ と設定した方が平均 MAE が小さくなった。図 5.2 は、BFCS におけるパラメータを $m = 1.5$ とした平均 MAE の結果である。BFCS を適用した提案法は、欠損数 2,000 以下では従来法の二手法より高精度だが、欠損数 2,500 以上では FireFly より精度が劣り、欠損数 3,000 以上からは GroupLens と精度が同等となった。図 5.3 は、QFCS におけるパラメータを $(\lambda, m) = (35, 1.01)$ とした平均 MAE の結果である。QFCS を適用した提案法は、KLFCS と同様に欠損数 5,500 以下のとき従来法より高精度であった。図 5.4 は、KLFCCM におけるパラメータを $\lambda = 14$ とした平均 MAE の結果である。KLFCCM を適用した提案法は、欠損数 3,500 から 5,000 の間に着目すると、 $C = 6$ と設定したときの MAE と $C = \{4, 5\}$ と設定したときに MAE の差が他の手法より高いことを確認した。図 5.5 は、BFCCM におけるパラメータを $m = 1.01$ とした平均 MAE の結果である。BFCCM を適用した提案法は、欠損数 5,500 以下のとき従来法の二手法より高精度であった。図 5.6 は、QFCCM におけるパラメータを $(\lambda, m) = (100, 1.01)$ とした平均 MAE の結果である。QFCCM を適用した提案法は、BFCCM と同様に欠損数 5,500 以下のとき従来法より高精度であった。図 5.7 は、KLFCCMM におけるパラメータを $(\lambda, t) = (10, 0.001)$ とした平均 MAE の結果である。KLFCCMM を適用した提案法は、GroupLens と同等かそれ未満の精度となった。図 5.8 は、BFCCMM におけるパラメータを $(m, t) = (1.01, 0.1)$ とした平均 MAE の結果である。BFCCMM を適用した提案法は、欠損数 5,500 以下のとき従来法の二手法より高精度であった。図 5.9 は、QFCCMM におけるパラメータを $(\lambda, m, t) = (100, 1.01, 0.01)$ とした平均 MAE の結果である。QFCCMM を適用した提案法は、欠損数 6,000 以下のとき従来法より高精度であった。図 5.10 は、欠損数 5,500 におけるクラスタ数を変化させたときの 10 通りの欠損パターンの平均 MAE の遷移を示す。クラスタ数を BFCS では $C = 9$ 、QFCS では $C = 6$ より大きく設定すると各変数が収束する前に計算限界を迎えクラスタリングが不可となった。BFCS, KLFCCM, KLFCCMM 以外のファジィクラスタリング手法を適用した提案法は、クラスタ数 $C = 5$ に近づくにつれて MAE が低く高精度であり、 $C = 6$ より大きく設定すると MAE が高く精度が下がる傾向が見られた。

図 5.1: KLFCS($\lambda = 35$) における 10 通り欠損パターンの平均 MAE図 5.2: BFCS($m = 1.5$) における 10 通り欠損パターンの平均 MAE

図 5.3: QFCS($\lambda = 35, m = 1.01$) における 10 通り欠損パターンの平均 MAE図 5.4: KLFCCM($\lambda = 14$) における 10 通り欠損パターンの平均 MAE

図 5.5: BFCCM($m = 1.01$) における 10 通り欠損パターンの平均 MAE図 5.6: QFCCM($\lambda = 100, m = 1.01$) における 10 通り欠損パターンの平均 MAE

図 5.7: KLFCCMM($\lambda = 10, t = 0.001$) における 10 通り欠損パターンの平均 MAE図 5.8: BFCCMM($m = 1.01, t = 0.1$) における 10 通り欠損パターンの平均 MAE

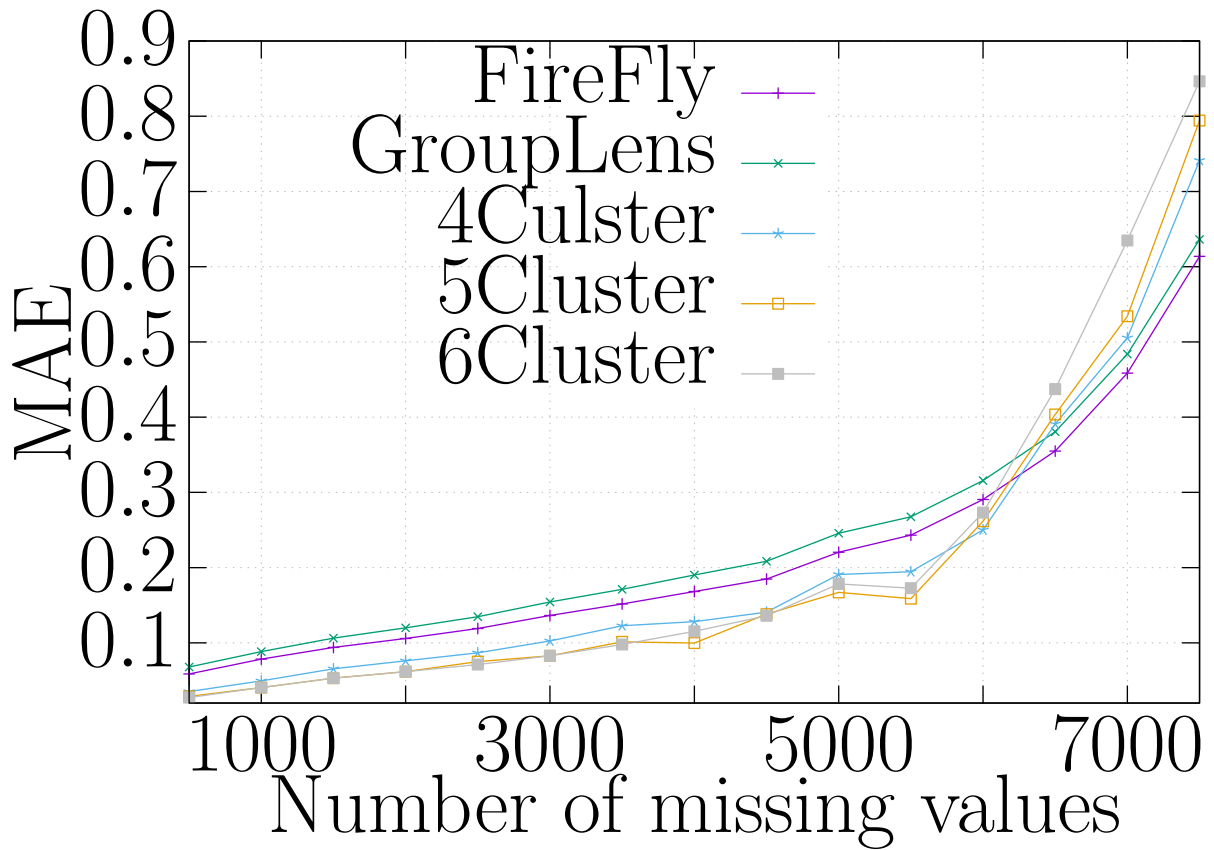


図 5.9: QFCCMM($\lambda = 100, m = 1.01, t = 0.01$) における 10 通り欠損パターンの平均 MAE

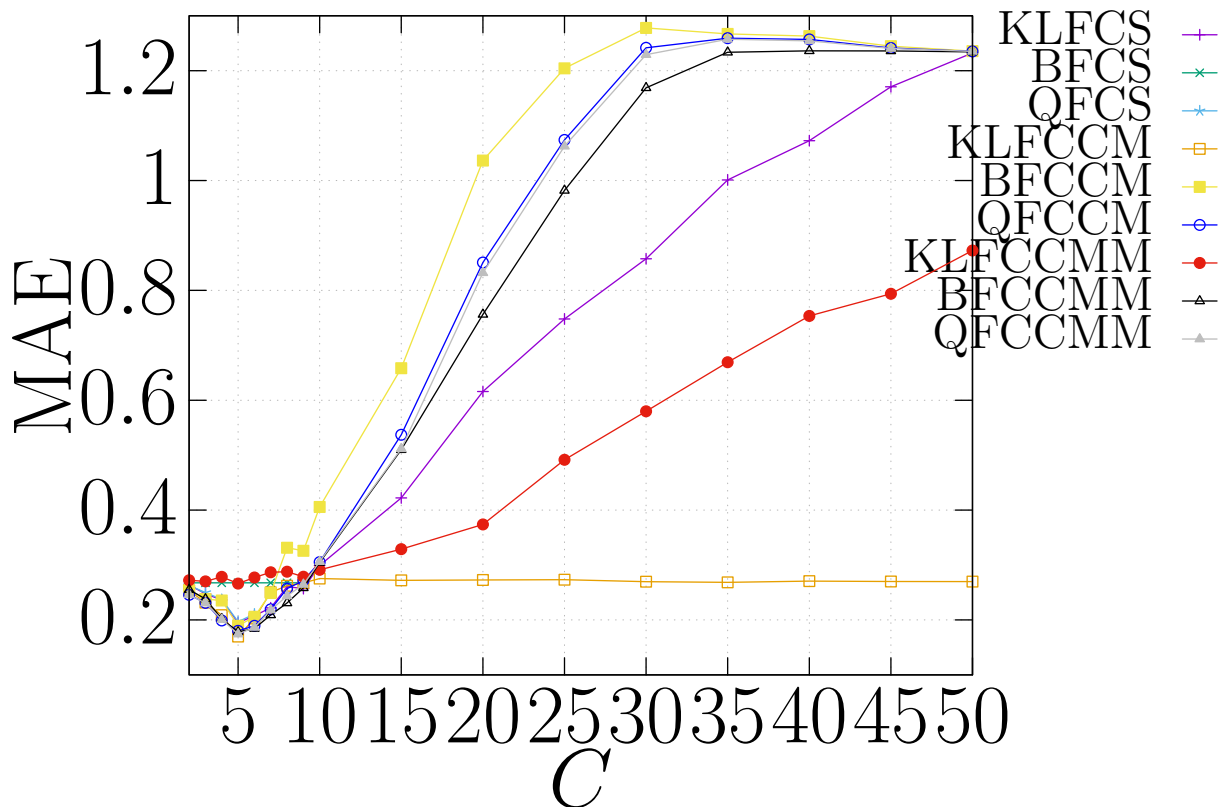


図 5.10: 欠損数 5,500 におけるクラスタ数を変化させたときの 10 通りの欠損パターンの平均 MAE

表 5.4 にクラスタ数 $C = 5$ と設定したときの欠損数 500, 5,500, 7,500 における 10 通りの欠損パターンの平均 MAE を示す。赤色は、各欠損数における最小の MAE を表す。パラメータを選択する際の基準とした欠損数 5,500 のとき、QFCCMM を適用した提案法は他のどの手法より高精度であった。しかし、欠損数 7,500 と設定したとき、提案法は従来法の二手法より精度が劣った。また、欠損数 7,500 と設定したとき、FCCM と FCCMM の六手法より FCS の三手法の方が MAE が低く高精度であった。これらの人工データ数値実験の結果から、欠損数が 5,500 以下で正解のクラスタ数を設定したとき、提案法の有効性を確認し、中でも QFCCMM を適用した提案法が高精度であることが示された。一方、欠損数が 5,500 より増えたとき、ファジィクラスタリングを応用することの有効性は失われたが、初期値の与え方やパラメータの採用方法によって改善を見込める。

表 5.4: $C = 5$ としたときの 10 通りの欠損パターンにおける平均 MAE

| 欠損数 手法 | 500 | 5,500 | 7,500 |
|-----------|----------|----------|----------|
| FireFly | 0.058658 | 0.243212 | 0.613757 |
| GroupLens | 0.067850 | 0.267617 | 0.636294 |
| KLFCS | 0.030742 | 0.194801 | 0.699196 |
| BFCS | 0.027216 | 0.267617 | 0.636294 |
| QFCS | 0.030462 | 0.196837 | 0.700160 |
| KLFCCM | 0.028480 | 0.169717 | 0.809289 |
| BFCCM | 0.029736 | 0.189351 | 0.852983 |
| QFCCM | 0.029305 | 0.180207 | 0.834473 |
| KLFCCMM | 0.067850 | 0.266237 | 0.737585 |
| BFCCMM | 0.029818 | 0.177908 | 0.826636 |
| QFCCMM | 0.028904 | 0.158812 | 0.794224 |

5.3 実データ数値実験

5.3.1 実データ

協調フィルタリング性能評価に六つの実データを用いた。bookcrossing は、Cai-Nicolas Ziegler によって <https://www.bookcrossing.com/> から収集されたデータである [41]。278,858 人のユーザが、271,379 冊の本を 1 から 10 段階で評価することで計 1,149,780 の評価値を得ている。epinions は、Paolo Massa によって <http://www.epinions.com/> から収集された商品の評価データである [42]。49,290 人のユーザが、139,738 のアイテムを 1 から 5 段階で評価することで計 664,824 の評価値を得ている。jester は、Ken Goldberg によって Jester Online Joke から収集されたジョークの評価データである [43]。59,132 人のユーザが、150 のジョークを -10.0 から +10.0 までの実数で評価することで約 1.7 百万の評価値を得ている。libimseti は、Vaclav Petricek によって <http://moje.libimseti.cz/> から収集されたユーザのプロファイルを評価したデータである [44]。135,359 人のユーザが、168,791 のプロファイルを 1 から 10 段階で評価することで計 17,359,346 の評価値を得ている。movie は、GroupLens によって作成されたデータである [45]。6,040 人のユーザが、3,900 の映画を 1 から 5 段階で評価することで計 1,000,209 の評価値を得ている。sushi は、Toshihiro Kamishima によって作

成されたデータである [46]. 5,000 人のユーザが, 100 の寿司ネタを 1 から 5 段階で評価することで計 50,000 の評価値を得ている.

協調フィルタリングの精度向上のため, これらの元データから既評価数が余りにも少ないユーザやアイテムを前処理を行った. `bookcrossing` では, 15 冊以上評価したユーザかつ 8 人以上に評価された本以外の情報を削ることで 1,090 ユーザ 2,248 アイテムによる 35,179 の評価データに整形した. `epinions` では, 24 以上評価したユーザかつ 24 人以上に評価されたアイテム以外の情報を削ることで 1,022 ユーザ 835 アイテムによる 42,808 の評価データに整形した. `jester` では, 128 以上評価したユーザ以外の情報を削ることで 2,916 ユーザ 140 アイテムによる 373,338 の評価データに整形した. また, 負の評価値を正にするため全体に +11 をすることで評価値の底上げをした. `libimseti` では, 229 以上評価したユーザかつ 229 人以上に評価されたプロフィール以外の情報を削ることで 866 ユーザ 1,156 アイテムによる 400,955 の評価データに整形した. `movie` では, 200 以上評価したユーザかつ 240 人以上に評価された映画以外の情報を削ることで 905 ユーザ 684 アイテムによる 277,546 の評価データに整形した. `sushi` は, ユーザが 10 個のアイテムを必ず評価するという制限のついたデータであったため, そのまま用いた. 表 5.5 に整形済みの評価データを示す.

表 5.5: 実データ数値実験に使用する整形済みの評価データ

| データ | ユーザ数 | アイテム数 | 実要素数 | スパース性 |
|---------------------------|------|-------|--------|----------|
| <code>bookcrossing</code> | 1091 | 2248 | 35179 | 98.5656% |
| <code>epinions</code> | 1022 | 835 | 42808 | 94.9837% |
| <code>jester</code> | 2916 | 140 | 373338 | 8.54938% |
| <code>libimseti</code> | 866 | 1156 | 400955 | 59.9484% |
| <code>movielens</code> | 905 | 684 | 277546 | 55.1636% |
| <code>sushi</code> | 5000 | 100 | 50000 | 90% |

5.3.2 設定

`bookcrossing` と `sushi` では 10,000, 他四つのデータでは 20,000 個の既評価値を欠損させ, 一つのデータセットに対して 5 通りの欠損パターンで実行する. 欠損箇所を決める乱数シード値は, 最大欠損数に至るまでインクリメントすることで変化させ, 新たな欠損パターンは, これまで加算されたシード値によって決定する. F-measure は, 閾値を各データの最大評価値の 7 割に設定することで算出する. ROC 曲線は, 0 からデータごとの最大評価値まで 0.1 刻みで閾値を変化させることで描く. ファジィクラスタリングのパラメータの範囲は, 基本的に $\lambda \in \{1, 10, 10^2, 10^3\}$, $m \in \{1.0001, 1.0004, 1.0007, 1.001\}$, $t \in \{10^{-4}, 10^{-5}, 10^{-6}\}$, $C \in \{2, 3, \dots, 8\}$ に従うが, パラメータの端点で平均 AUC が最高精度とならない値まで変化させる. 最終的に定まったパラメータ範囲を表 5.6 と表 5.7 に示す. 一組のパラメータに対し `k-means++` によるアルゴリズムで 10 通りクラスタ中心とクラスタ混合比による初期設定を行い, 目的関数値最小のクラスタリング結果を採用する. また, 10 回以上計算限界を迎えた, もしくは更新回数を 10^3 を超えても収束しない場合, そのパラメータでの処理を終えてクラスタリング不可であったとみなす. 以下に協調フィルタリングにおける実データ数値実験の流れを示す:

実験の流れ 5 (協調フィルタリングにおける実データ数値実験).

STEP 1. 表 5.5 に示すデータの読み込み.

- STEP 2. 表 5.6 の範囲でパラメータを設定する.
- STEP 3. 既評価値を欠損する.
- STEP 4. $\{B, KL, Q\}$ FCS では式 (3.8) に従い球面化し, $\{B, KL, Q\}$ FC $\{CM, CMM\}$ では式 (3.9) に従い単体化する.
- STEP 5. k -means++ における平方ユークリッド距離の代わりにオブジェクト-クラス間非類似度を $\{B, KL, Q\}$ FCS では式 (2.78), $\{KL, Q, B\}$ FCCM では式 (2.128), $\{KL, B, Q\}$ FCCMM では式 (2.172) に置き換えたアルゴリズム 3 に従い初期クラスタ中心を設定する.
- STEP 6. 初期クラスタサイズ調整変数を $1/C$ に設定する.
- STEP 7. $\{B, KL, Q\}$ FCS では式 (2.78), $\{KL, Q, B\}$ FCCM では式 (2.128), $\{KL, B, Q\}$ FCCMM では式 (2.172) に従いオブジェクト-クラス間非類似度を更新する.
- STEP 8. BFCS では式 (2.88), KLFCFS では式 (2.102), QFCS では式 (2.116), KLFFCCM では式 (2.138), BFCCM では式 (2.166), QFCCM では式 (2.152), KLFFCCMM では式 (2.182), BFCCMM では式 (2.196), QFCCMM では式 (2.210) に従い帰属度を更新する.
- STEP 9. BFCS では式 (2.90), KLFCFS では式 (2.104), QFCS では式 (2.118), KLFFCCM では式 (2.140), BFCCM では式 (2.168), QFCCM では式 (2.154), KLFFCCMM では式 (2.184), BFCCMM では式 (2.198), QFCCMM では式 (2.212) に従いクラスタ中心を更新する.
- STEP 10. BFCS では式 (2.92), KLFCFS では式 (2.106), QFCS では式 (2.120), KLFFCCM では式 (2.142), BFCCM では式 (2.170), QFCCM では式 (2.156), KLFFCCMM では式 (2.186), BFCCMM では式 (2.200), QFCCMM では式 (2.214) に従いクラスタサイズ調整変数を更新する.
- STEP 11. 更新前と更新後の帰属度, クラスタ中心, クラスタサイズ調整変数の最大値ノルムを計算し, それらの和が 10^{-10} 以上であれば STEP 7. へ, 計算限界を迎えたら初期値設定乱数シードをインクリメントし STEP 5. へ, 10 回計算限界を迎えたもしくは更新数が 10^3 を超えたらそのパラメータでのクラスタリングを不可とする.
- STEP 12. 収束した帰属度をクリスプ化する.
- STEP 13. アルゴリズム 10 に従って欠損された評価値を予測する. クラスタリング不可と見なされたパラメータは式 (4.8) によって予測し, そのときの目的関数値を 1.79769×10^{308} とする.
- STEP 14. 予測値と正解の評価値を照らし合わせ式 (4.4.1) に従い MAE, 設定した閾値にから表 (4.1) に基づいて F-measure と AUC を算出する.
- STEP 15. 10 通りの初期設定を行えたら, そのうち目的関数値が最小の結果を保持する. そうでなければ STEP 5. へ.
- STEP 16. 5 通りの欠損パターンで結果を得たら, その平均を算出する. そうでなければ欠損パターン発生乱数シード値をインクリメントし STEP 3. へ.
- STEP 17. 次に設定すべきパラメータがあれば STEP 2. へ.
- STEP 18. 次に処理すべきデータがあれば STEP 1. へ, そうでなければ終了する. ■

表 5.6: 実データ数値実験で採用したパラメータ範囲 $\{KL, B, Q\}FC\{CM, CS\}$

| 手法 \ データ | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|----------|--|---|---|---|--|---|
| KLFCM | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8, 10, 15, \dots, 40\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8\}$ |
| BFCS | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ |
| QFCS | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01\}$, $C \in \{2, 3, \dots, 8, 10, 15, \dots, 40\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20, 25\}$ | $\lambda \in \{1, 10, \dots, 10^6\}$, $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ |
| KLFCM | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $C \in \{2, 3, \dots, 8\}$ |
| BFCCM | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001, 1.01\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-7}, 1 + 10^{-6}, 1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ |
| QFCCM | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15, \dots, 40\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^5\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20, 25\}$ | $\lambda \in \{1, 10, \dots, 10^6\}$, $m \in \{1 + 10^{-8}, 1 + 10^{-7}, \dots, 1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001, 1.01\}$, $C \in \{2, 3, \dots, 8\}$ |

表 5.7: 実データ数値実験で採用したパラメータ範囲 $\{KL, B, Q\}FCCMM$

| データ 手法 | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|-----------|--|---|---|--|--|---|
| KLFCMM | $\lambda \in \{1, 10, \dots, 10^6\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-2}\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8, 10\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-1}\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-1}\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8\}$ |
| BFCMM | $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-2}\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-4}, 1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $t \in \{10^{-8}, 10^{-7}, \dots, 10^{-4}\}$, $C \in \{2, 3, \dots, 8, 10, 15\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01, 1.1\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-1}\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $m \in \{1.0001, 1.0004, \dots, 1.001, 1.01\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-1}\}$, $C \in \{2, 3, \dots, 8\}$ | $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, 10^{-4}\}$, $C \in \{2, 3, \dots, 8\}$ |
| QFCCMM | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-7}, 10^{-6}, \dots, 10^{-4}\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8, 10, 15, 20\}$ | $\lambda \in \{1, 10, \dots, 10^4\}$, $m \in \{1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-2}\}$, $C \in \{2, 3, \dots, 8\}$ | $\lambda \in \{1, 10, \dots, 10^3\}$, $m \in \{1 + 10^{-5}, 1.0001, 1.0004, \dots, 1.001\}$, $t \in \{10^{-6}, 10^{-5}, \dots, 10^{-3}\}$, $C \in \{2, 3, \dots, 8\}$ |

5.3.3 結果と考察

表 5.8 に、最高平均 AUC を達成したファジィクラスタリングにおけるパラメータを示す。5 通りの欠損パターンによる最高平均 AUC を達成したパラメータの平均 MAE, 平均 F-measure を表 5.9, 表 5.10 に、最高平均 AUC を表 5.11 に示す。各手法において最高精度な値を赤色で、従来法より高精度な値を青色で強調した。bookcrossing を適用した場合、三つの精度評価方法で QFCCM を適用した提案法が最高精度であった。QFCCMM と比べると QFCCM の最高平均 AUC を達成したクラスタ数が大きいことに着目する。QFCCMM が $C = 15$ に対して、QFCCM は $C = 35$ で最高平均 AUC を達成している。実データ数値実験では、設定したクラスタ数の端点で最高平均 AUC を達成した場合、まずファジィ化パラメータの範囲を $\lambda \in \{1, 10, 10^2, 10^3\}$, $m \in \{1.0001, 1.0004, 1.0007, 1.001\}$, $t \in \{10^{-4}, 10^{-5}, 10^{-6}\}$ に固定しクラスタ数を増加させ、クラスタ数が定まった後、追加でファジィ化パラメータの推定を行う。したがって、bookcrossing を適用した QFCCMM による提案法において、 $C = 35$ と固定して包括的にパラメータを変化させることで QFCCM の精度へ近づくことが期待できる。epinions と sushi を適用した場合、 $B\{FCS, FCCM, FCCMM\}$ の三手法を除き、提案法は従来法である GroupLens と同等の精度となった。この二つのデータにおいて、最高平均 AUC を達成したクラスタ数に着目すると、全手法で $C = 2$ である。帰属度を確認すると、全ユーザが一つのクラスタに属すクラスタリング結果になっており、ファジィクラスタリングを行いユーザをフィルタリングしたにも関わらず全ユーザを対象にしてピアソン相関係数を算出していた。epinions の元データは、他の元データと比べ、スパース性が強いこと最適なユーザクラスタを形成することができなかったと予想できる。また、表 5.5 に示した整形済みのデータにおいて、epinions は bookcrossing の次にスパース性が高いが、アイテム数の割合は bookcrossing と比べて小さい。bookcrossing を適用した場合、クラスタリングによるユーザフィルタリングが有効であったことを踏まえると、epinions のデータ整形の基準を変更しアイテム数を増やすことで epinions を適用した提案法は、従来法より高精度になることが期待できる。sushi も同様にユーザ数を削るデータ整形を行うことで、クラスタリングによる最適なユーザクラスタを形成できる見込みがある。

jester と libimseti を適用した場合、三つの精度評価方法で QFCS を適用した提案法が最高精度であった。movielens を適用した場合、MAE と AUC による精度評価において、BFCCMM を適用した提案法が最高精度であり、F-measure による精度評価では、QFCS を適用した提案法が最高精度であった。MAE と AUC では六つデータのうち四つのデータ、F-measure では六つデータのうち五つデータで QFCS が最高精度であった。全データにおいて提案したファジィクラスタリングと GroupLens を組み合わせた協調フィルタリングは、大半のデータで従来の協調フィルタリングより高い精度であった。実データ数値実験からファジィクラスタリングを協調フィルタリングに応用する有効性を確認した。

表 5.8: 5 通りの欠損パターンにおける最高平均 AUC を達成したパラメータ

| データ 手法 | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|----------------------------------|---------------------------------|------------------------------|---------------------------------|------------------------------|--------------------------------|------------------------------|
| KLFCs (λ, C) | ($10^2, 35$) | (1.0, 2) | ($10^3, 15$) | ($10^3, 15$) | ($10^2, 7$) | (10, 2) |
| BFCS (m, C) | (1.0007, 15) | (1.0007, 2) | (1.01, 6) | (1.001, 6) | (1.0004, 5) | (1.0007, 2) |
| QFCS (λ, m, C) | ($10^2, 1.001, 35$) | (10, 1.0001, 2) | ($10^3, 1.0001, 10$) | ($10^3, 1.0004, 20$) | ($10^5, 1.01, 5$) | (10, 1.0001, 2) |
| KLFCMM (λ, C) | (10, 20) | (1.0, 2) | ($10^3, 10$) | ($10^3, 3$) | ($10^3, 3$) | (1.0, 2) |
| BFCCM (m, C) | (1.001, 7) | (1.0007, 2) | (1.0001, 7) | (1.0001, 3) | ($1 + 10^{-6}, 3$) | (1.0001, 2) |
| QFCCM (λ, m, C) | (10, 1.0001, 35) | (1.0, 1.0001, 2) | ($10^4, 1.0001, 6$) | ($10^3, 1.0001, 4$) | ($10^4, 1 + 10^{-7}, 5$) | (1.0, 1.0001, 2) |
| KLFCMM (λ, t, C) | ($10^5, 10^{-3}, 2$) | (1.0, $10^{-4}, 2$) | ($10^3, 10^{-4}, 10$) | (10, $10^{-2}, 6$) | ($10^2, 10^{-2}, 3$) | (1.0, $10^{-4}, 2$) |
| BFCCMM (m, t, C) | (1.0004, $10^{-3}, 15$) | (1.1, $10^{-4}, 2$) | (1.0001, $10^{-7}, 10$) | (1.01, 0.1, 15) | (1.001, 0.1, 5) | (1.0001, $10^{-5}, 2$) |
| QFCCMM (λ, m, t, C) | ($10^2, 1.0004, 10^{-6}, 15$) | (1.0, 1.0001, $10^{-4}, 2$) | ($10^3, 1.0001, 10^{-5}, 10$) | (10, 1.0007, $10^{-4}, 15$) | ($10^3, 1.0007, 10^{-3}, 7$) | (1.0, 1.0001, $10^{-4}, 2$) |

表 5.9: 5 通りの欠損パターンにおける最高平均 MAE

| データ 手法 | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|-----------|--------------|----------|----------|-----------|-----------|----------|
| FireFly | 4.201491 | 1.303765 | 3.845269 | 1.246787 | 0.738720 | 1.069868 |
| GroupLens | 1.338443 | 0.970077 | 2.993373 | 1.260037 | 0.676261 | 0.924108 |
| KLFCs | 1.204653 | 0.970077 | 2.903557 | 1.144415 | 0.673406 | 0.924108 |
| BFCS | 1.243085 | 0.988594 | 2.907166 | 1.158393 | 0.672924 | 0.975704 |
| QFCS | 1.204228 | 0.970077 | 2.899465 | 1.141816 | 0.672205 | 0.924108 |
| KLFCcM | 1.199717 | 0.970077 | 2.909731 | 1.192282 | 0.672128 | 0.924108 |
| BFCCM | 1.222694 | 0.988408 | 2.899691 | 1.194279 | 0.672047 | 0.945013 |
| QFCCM | 1.190245 | 0.970077 | 2.904109 | 1.174198 | 0.672786 | 0.924108 |
| KLFCcMM | 1.289728 | 0.970077 | 2.906338 | 1.168268 | 0.671907 | 0.924108 |
| BFCCMM | 1.221883 | 0.970077 | 2.899471 | 1.148633 | 0.671722 | 0.946472 |
| QFCCMM | 1.222128 | 0.970077 | 2.909329 | 1.159731 | 0.672735 | 0.924108 |

表 5.10: 5 通りの欠損パターンにおける最高平均 F-measure

| データ 手法 | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|-----------|--------------|----------|----------|-----------|-----------|----------|
| FireFly | 0.626389 | 0.685638 | 0.387420 | 0.785615 | 0.744940 | 0.712558 |
| GroupLens | 0.861541 | 0.763546 | 0.648815 | 0.805824 | 0.779160 | 0.740651 |
| KLFCcM | 0.886557 | 0.763546 | 0.659023 | 0.815706 | 0.781725 | 0.740651 |
| KLFCs | 0.886088 | 0.763546 | 0.661554 | 0.820136 | 0.781443 | 0.740651 |
| BFCS | 0.879429 | 0.757989 | 0.660596 | 0.819739 | 0.780730 | 0.721228 |
| QFCS | 0.886138 | 0.763546 | 0.662192 | 0.821045 | 0.782027 | 0.740651 |
| BFCCM | 0.883195 | 0.758648 | 0.659229 | 0.815031 | 0.781874 | 0.735289 |
| QFCCM | 0.888401 | 0.763546 | 0.660166 | 0.817669 | 0.780334 | 0.740651 |
| KLFCcMM | 0.870252 | 0.763546 | 0.658691 | 0.815361 | 0.781593 | 0.740651 |
| BFCCMM | 0.883492 | 0.763546 | 0.660230 | 0.819178 | 0.781490 | 0.731400 |
| QFCCMM | 0.883371 | 0.763546 | 0.658948 | 0.818024 | 0.781638 | 0.740651 |

5.4 まとめ

人工データの数値実験では、欠損数 5,500 以下においてファジィクラスタリングを適用した提案法の有効性が示され、欠損数 6,000 のとき QFCCMM を適用した提案法は、全ての手法の中で最高精度を達成した。一方、欠損数が 7,500 において QFCCMM と比べて {B,KL,Q}FCS の精度が高い。表 5.12 は、人工データにおいて、欠損数 5,500 の最小平均 MAE を達成したファジィ化パラメータを採用し、 $C = 5$ と設定したときの 10 通りの欠損パターンにおける QFCS と QFCCMM の平均 ARI である。欠損数 5,500 では、QFCCMM の平均 ARI は QFCS より高く、逆に欠損数 7,500 では、QFCCMM の平均 ARI は QFCS より明らかに低い。したがって、QFCS は QFCCMM に比べて欠損したデータに

表 5.11: 5 通りの欠損パターンにおける最高平均 AUC

| データ 手法 | bookCrossing | epinions | jester | libimseti | movielens | sushi |
|-----------|--------------|----------|----------|-----------|-----------|----------|
| FireFly | 0.536845 | 0.674627 | 0.346844 | 0.909979 | 0.730208 | 0.669754 |
| GroupLens | 0.704500 | 0.733662 | 0.830666 | 0.913046 | 0.787796 | 0.723708 |
| KLFCS | 0.723359 | 0.733662 | 0.841467 | 0.929506 | 0.792133 | 0.723708 |
| BFCS | 0.717226 | 0.722371 | 0.840939 | 0.927999 | 0.792348 | 0.703232 |
| QFCS | 0.723660 | 0.733662 | 0.841978 | 0.929594 | 0.792885 | 0.723708 |
| KLFCCM | 0.724648 | 0.733662 | 0.840526 | 0.923149 | 0.792359 | 0.723708 |
| BFCCM | 0.719910 | 0.722559 | 0.841379 | 0.922787 | 0.792518 | 0.716043 |
| QFCCM | 0.726737 | 0.733662 | 0.840862 | 0.925907 | 0.792160 | 0.723708 |
| KLFCCMM | 0.710181 | 0.733662 | 0.840747 | 0.926597 | 0.792534 | 0.723708 |
| BFCCMM | 0.720024 | 0.733662 | 0.841609 | 0.928940 | 0.793009 | 0.712292 |
| QFCCMM | 0.720019 | 0.733662 | 0.840505 | 0.927429 | 0.792885 | 0.723708 |

対して適切にクラスタリングが可能であると考えられる。

表 5.12: $C = 5$ としたときの 10 通りの欠損パターンにおける平均 ARI (人工データ)

| 欠損数 手法 | 5,500 | 7,500 |
|-----------|----------|----------|
| QFCS | 0.886958 | 0.227328 |
| QFCCMM | 0.899558 | 0.091440 |

実データの数値実験では、六つの評価データのうち四つの評価データで QFCS を適用した提案法が最高平均 AUC を達成した。表 5.11 から QFCS と QFCCMM の結果において、最高平均 AUC の差が最も大きいデータのうち AUC の差が最も大きい欠損パターンに着目すると libimseti の第 2 欠損パターンが当てはまった。図 5.11 は、libimseti の第 3 欠損パターンにおける $Q\{FCS, FCCMM\}$ を適用した提案法の TP/FP である。TP/FP は高いほど高精度であることから、閾値が 1 から 4, 9.5 から 10 においては QFCS が高精度である。対して、閾値が 8.1 から 8.6 においては QFCCMM が高精度である。次に $Q\{FCS, FCCMM\}$ においてクラスタ数 $C = 20$ に固定し、5 通りの欠損パターンの最高平均 AUC を達成したファジィ化パラメータを採用し、そのうち AUC の差が最も大きい第 0 欠損パターンに着目した。表 5.13 は、libimseti の第 0 欠損パターンにおける $Q\{FCS, FCCMM\}(C = 20)$ の Contingency Table である。全ユーザ 866 に対して、QFCCMM では第 0 クラスタから第 8 クラスタまでに 856 のユーザが属している。一方、QFCS は QFCCMM と比べて第 0 クラスタから第 19 クラスタまで満遍なくユーザが分布している。次に全クラスタを適用した提案法の AUC と特定のクラスタのみ適用した提案法の AUC を比較する。QFCCMM と QFCS でユーザのばらつきが大きい第 2 クラスタに着目した。図 5.12 は、libimseti の第 0 欠損パターンにおける $Q\{FCS, FCCMM\}(C = 20)$ を適用した提案法の TP/FP である。一方、図 5.13 は、libimseti の第 0 欠損パターンにおける $Q\{FCS, FCCMM\}(C = 20)$ の第 1 クラスタに着目した適用した提案法の TP/FP である。図 5.12 と図 5.13 を比較すると第 1 クラスタに着目した結果における QFCS と QFCCMM の TP/FP の差が大

きいことが分かる．特に閾値が1から5の間において，QFCSとQFCCMMの精度差が大きいことを確認した．表5.14は，libimsetiの第0欠損パターンにおいて $Q\{FCS, FCCMM\}(C=20)$ を適用し，全クラスタを用いたAUCと第1クラスタのみ用いたAUCとその差を示す．第1クラスタのみを適用した提案法のAUCの差が大きいことが確認できる．これらの結果から，QFCCMMの第1クラスタに属す200のユーザのうち，QFCSの第1クラスタに属す66ユーザが一つのクラスタとされるべきである．したがって，QFCSによるユーザ分割の精度はQFCCMMより高いことが示された．

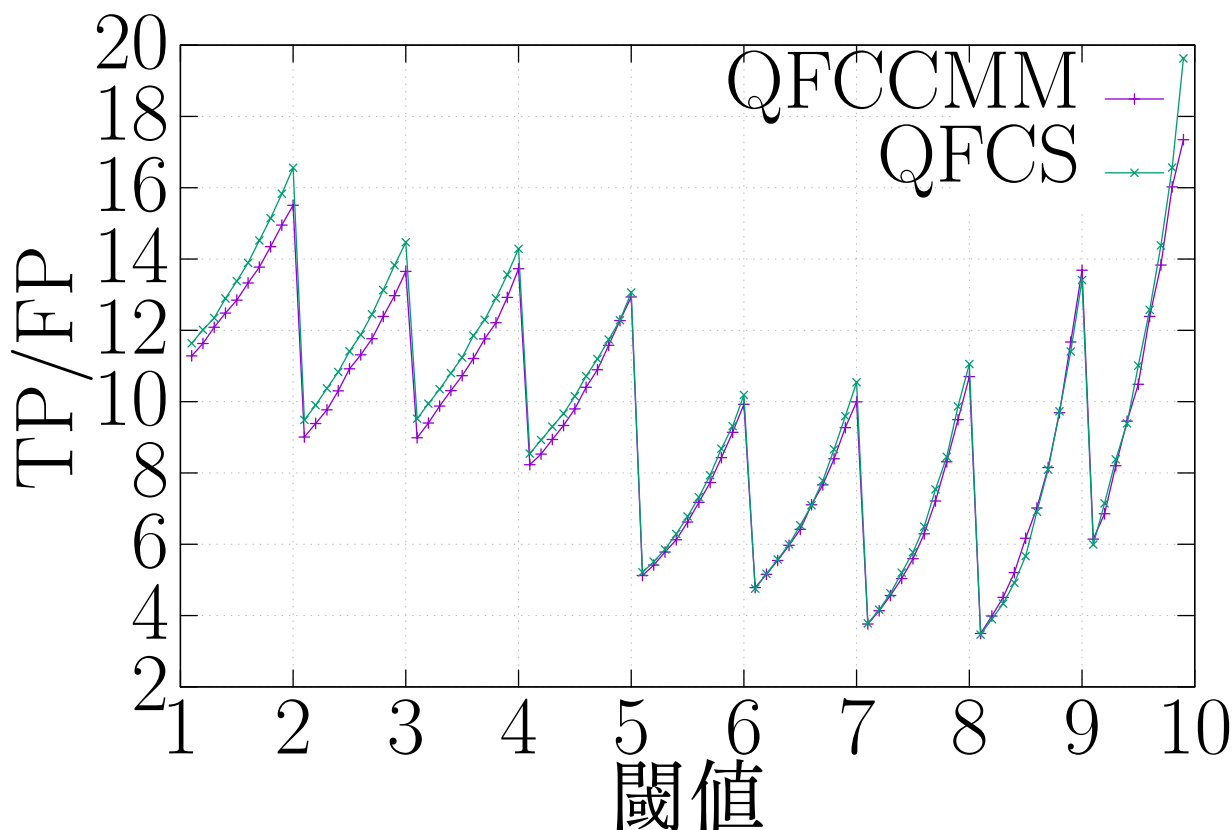


図 5.11: libimseti の第3欠損パターンにおける QFCS($C=20$), QFCCMM($C=15$) を適用した提案法の TP/FP

5.5 おわりに

評価データを使用した協調フィルタリングにおける数値実験について記述した．まず第5.2節で人工データを使用した数値実験の設定とその結果および考察について記述した．人工データによる数値実験では，欠損数6,000以下のときQFCCMMを適用した提案法が最高精度を達成し，欠損数5,500以下のとき $\{B, KL, Q\}FC\{S, CM, CMM\}$ を適用した提案法の有効性を示したが，欠損数が7,500のとき提案法は従来法より劣り，そのときQFCCMMを適用した提案法はQFCSより精度が低くなった．次に第5.3節で六つ実データと整形方法と数値実験の設定と結果および考察について示した．六つのデータのうち四つのデータでQFCSが最高AUCを達成し，bookcrossing, jester, libimseti, movielensを適用した提案法は，どのファジィクラスタリング手法を応用した場合であっても従来法より高精度であった．最後に第5.4節で人工データと実データの協調フィルタリング数値実験の結果を踏まえまとめた．人工データの既存評価数のうち7.5割欠損させた場合，QFCCMMよりQFCSの方がクラスタリング精度が高いことが示され，libimsetiの結果から，QFCSがQFCCMMよりユーザ分割の精度が

表 5.13: libimseti の第 0 欠損パターンにおける $Q\{FCS,FCCMM\}(C = 20)$ の Contingency Table

| QFCS \ QFCCMM | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 合計 |
|---------------|-----|-----|----|-----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|-----|
| 0 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 101 |
| 1 | 0 | 66 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 66 |
| 2 | 0 | 0 | 65 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 66 |
| 3 | 0 | 0 | 0 | 61 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 61 |
| 4 | 0 | 0 | 0 | 3 | 47 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 50 |
| 5 | 0 | 0 | 0 | 0 | 0 | 32 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 35 |
| 6 | 0 | 6 | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 38 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 40 |
| 8 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 38 |
| 9 | 24 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 |
| 10 | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 36 |
| 11 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 12 | 0 | 0 | 0 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 46 |
| 13 | 0 | 37 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 37 |
| 14 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 |
| 15 | 0 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 54 |
| 16 | 0 | 1 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 |
| 17 | 0 | 6 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 37 |
| 18 | 68 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 71 |
| 19 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| 合計 | 203 | 200 | 71 | 157 | 47 | 63 | 49 | 46 | 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 866 |

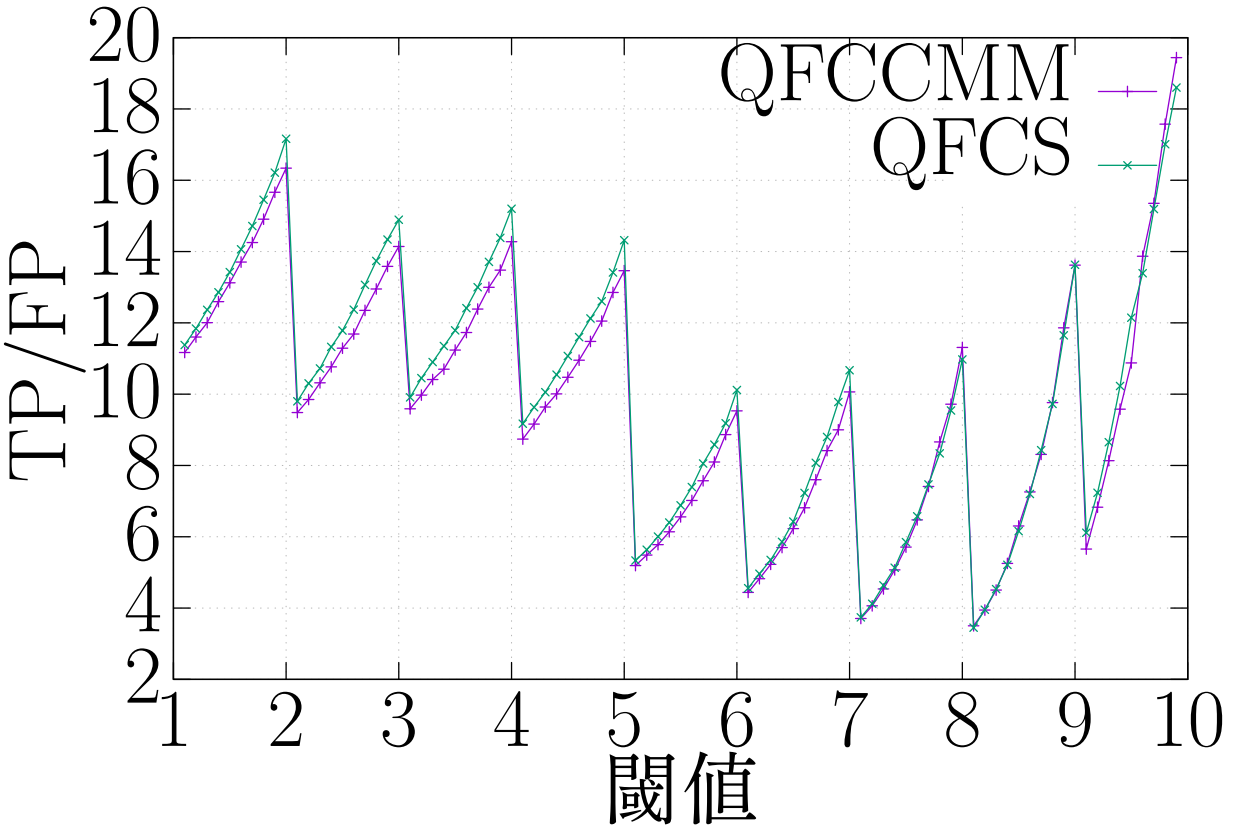


図 5.12: libimseti の第 0 欠損パターンにおける $Q\{FCS,FCCMM\}(C = 20)$ を適用した提案法の TP/FP

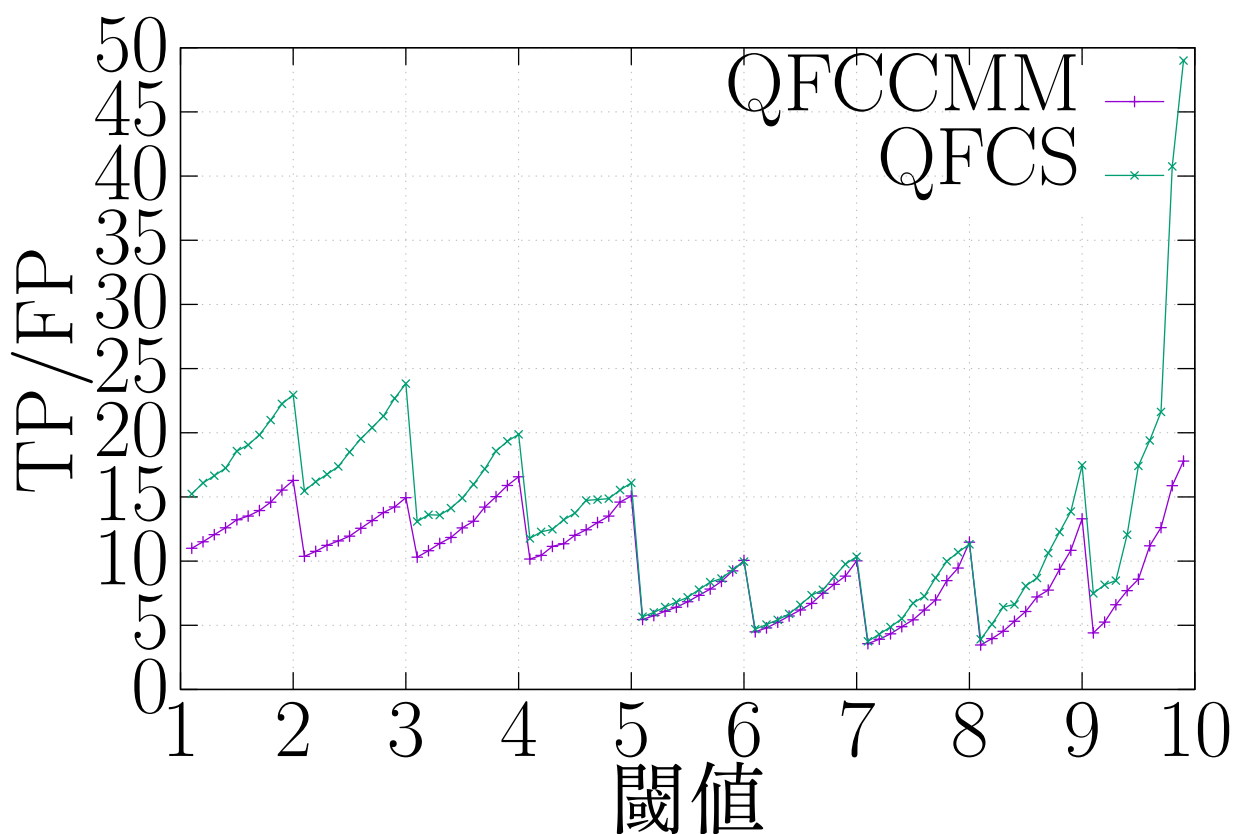


図 5.13: libimseti の第 0 欠損パターンにおける $Q\{FCS, FCCMM\}(C = 20)$ の第 1 クラスタに着目した提案法の TP/FP

表 5.14: libimseti の第 0 欠損パターンにおける $Q\{FCS, FCCMM\}(C = 20)$ を適用した AUC の差

| 手法 欠損 | QFCS | QFCCMM | QFCS-QFCCMM |
|----------|----------|----------|-------------|
| 全クラスタ | 0.930995 | 0.927696 | 0.00329818 |
| 第 1 クラスタ | 0.938487 | 0.930549 | 0.00793722 |

高く、協調フィルタリングへの応用に適した手法であることを確認した。

今後の展望として、欠損数や欠損パターン、ファジィクラスタリングにおけるパラメータや初期値設定数を増やすことによって、提案法の有効性をさらに強調することができるだろう。また、混合多項分布に基づくファジィクラスタリングと関連のあるユーザ間類似度を協調フィルタリングに組み込むことや、可能性クラスタリングを適用することや、関係性データのためのファジィクラスタリングを適用することで推薦精度向上を期待する。さらに、提案法に、トピックモデルを導入したり、Matrix Factorization のアイデアを融合することによって、さらなる高精度化や高性能化が期待できる。

第 6 章

結論

第 2 章では、ファジィクラスタリングについて述べた。混合多項分布に基づく q -divergence 正則化ファジィクラスタリング (QFCCMM) を提案し、最適化問題から更新式を導出し、アルゴリズムを示した。第 3 章では、文書データを使用したファジィクラスタリングにおける数値実験について記述した。3 通りの数値実験によって提案法である QFCCMM の優位性を示した。クラスタリング不可とみなしたデータに対してパラメータ範囲や収束条件などの実験設定を詳細に見直して、提案アルゴリズムをより堅牢にすることによって、クラスタリングに関する知識のないユーザにも使えるようになるだろう。第 4 章では、従来の協調フィルタリングである FireFly と GroupLens について示し、ファジィクラスタリング手法を GroupLens に応用した協調フィルタリング手法を提案した。提案法では、予めファジィクラスタリングによってユーザを分割することで、潜在的に似たユーザのみに対してピアソン相関係数を算出し、GroupLens による予測の精度向上を期待した。第 5 章では、評価データを使用した協調フィルタリングにおける数値実験について記述した。人工データにおける数値実験では、データの既評価数のうち 5.5 割以下欠損させたときファジィクラスタリングを適用した提案法の有効性が示され、6 割欠損させたとき QFCCMM を適用した提案法は最高精度を達成し、6.5 割以上欠損させたとき提案法の精度は従来法より劣った。ここで 7.5 割欠損させたとき、QFCS は QFCCMM よりユーザ分割の精度が高いことが示された。実データを使用した協調フィルタリングにおける数値実験では、六つのデータのうちの四つのデータで QFCS を適用した提案法が最高精度を達成し、大半のデータにおいて $\{B, KL, Q\}FC\{S, CM, CMM\}$ を応用した提案法は、従来の協調フィルタリングより高精度であることが示された。libimseti の詳細な数値実験結果から QFCS は QFCCMM よりユーザ分割の精度が高く、協調フィルタリングへの応用に適した手法であることが示された。欠損数や欠損パターン、ファジィクラスタリングにおけるパラメータや初期値設定数を増やすことによって、提案法の有効性をさらに強調することができるだろう。

文書データを使用したファジィクラスタリングにおける数値実験 2, 3 と評価データを使用した協調フィルタリングにおける数値実験から q -divergence の優位性を確認したが、文書クラスタリングでは QFCCMM の精度が高い一方、協調フィルタリングへの応用における実データ数値実験や欠損数増加時の人工データ数値実験では QFCS が高精度であった。これは、球面化されたデータをクラスタリングする FCS と球面性のあるピアソン相関係数の親和性や、FCS が欠損されたデータに対して適切なユーザ分割が可能であることが要因だと考えられる。

今後の展望として、まず、混合多項分布に基づく q -divergence 正則化ファジィクラスタリングに、可能性理論的アプローチやカーネル法を導入することによってクラスタリングにおけるさらなる高精度化や高性能化が期待できる。また、混合多項分布に基づくファジィクラスタリングと関連のあるユーザ間類似度を協調フィルタリングに組み込むことや、可能性クラスタリングを適用することや、関係性デー

タのためのファジィクラスタリングを適用することで推薦精度向上を期待する [47, 48, 49]. さらに, 協調フィルタリングに, トピックモデルを導入したり, Matrix Factorization のアイデアを融合することによって, さらに高精度化や高性能化が期待できる.

参考文献

- [1] Jan-Willem van Dam, and Michel van de Velden.: “Online profiling and clustering of Facebook users,” *Decision Support System*, Vol. 70, pp.60–72, (2015).
- [2] Tsallis, C.: “Possible Generalization of Boltzmann-Gibbs Statistics,” *J. Statist. Phys.*, Vol. 52, pp. 479–487, (1988).
- [3] Tsallis, C.: “Generalized Entropy-based Criterion for Consistent Testing,” *Phys. Rev. E*, Vol. 58, pp. 1442–1445, (1998).
- [4] Higashi, M., and Kanzawa, Y.: “Fuzzy Clustering Method for Spherical Data Based on q -Divergence,” *JACIII* (to appear in 2019).
- [5] Honda, K.: “Fuzzy Co-Clustering and Application to Collaborative Filtering,” *Integrated Uncertainty in Knowledge Modelling and Decision Making*, LNAI 9978, pp. 16–23, (2016).
- [6] Nakano, T., Honda, K., Ubukata, S., and Notsu, A.: “A Study on Recommendation Ability in Collaborative Filtering by Fuzzy Co-clustering with Exclusive Item Partition,” *2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS)*, pp. 686–689, (2016).
- [7] Honda, K., Oshio, S., and Notsu, A.: “FCM-type fuzzy co-clustering by K-L information regularization,” *Proc. FUZZ-IEEE*, pp. 2505–2510, (2014).
- [8] Honda, K., Oshio, S., and Notsu, A.: “Fuzzy Co-Clustering Induced by Multinomial Mixture Models,” *JACIII*, Vol. 19, No. 6, pp. 717–726, (2015).
- [9] Paul, R., Neophytos, I., Mitesh, S., Peter, S., and Jhon, R.: “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” *Proc. Computer supported cooperative work of the ACM*, pp. 175–186, (1994).
- [10] Sarwar, B., Karypis, G., and Riedl, J.: “Item-Based Collaborative Filtering Recommendation Algorithms,” *Proc. the 10th international conference on World Wide Web*, pp. 285–295, (2001).
- [11] Herlocker, J.L., Konstan, J.A., Borchers, A., and Riedl, J.: “An algorithmic framework for performing collaborative filtering,” *Proc. the 22nd annual international ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 230–237, (1999).
- [12] Func, S.: “Netflix Update: Try This at Home,” Available from: <https://sifter.org/~simon/journal/20061211.html>
- [13] Wang, C and Blei, D. M.: “Collaborative topic modeling for recommending scientific articles,” *Proc. the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 448–456, (2011).
- [14] MacQueen, J. B.: “Some Methods for Classification and Analysis of Multivariate Observations,” *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–

- 297, (1967).
- [15] Katsavounidis, I. C. C, Kuo, J and Zhang, Z.: “A New Initialization Technique for Generalized Lloyd Iteration,” IEEE Signal Processing Letters, Vol .1, No. 10 pp. 144–146, (1994). 144–146, 1994.
 - [16] Arthur, D., and Vassilvitskii, S.: “k-means++: the advantages of careful seeding,” Proc. The 18th annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035, (2007).
 - [17] L. A. Zadeh: “Fuzzy sets, Information and Control,” Vol. 8, No. 3, pp. 338–353, (1965).
 - [18] Dunn, J. C. “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters,” Journal of Cybernetics, Vol. 3, pp. 32–57 (1974).
 - [19] Bezdek, J. C.: “Pattern Recognition with Fuzzy Objective Function Algorithms,” Plenum Press, New York (1981).
 - [20] Miyamoto, S., Kurosawa, N.: “Controlling cluster volume sizes in fuzzy c-means clustering,” Proc. SCIS&ISIS2004, pp. 1–4, (2004).
 - [21] Miyamoto, S., and Mukaidono, M.: “Fuzzy c-Means as a Regularization and Maximum Entropy Approach,” Proc. 7th International Fuzzy Systems Association World Congress (IFSA’ 97), Vol. 2, pp. 86–92, (1997).
 - [22] Mènard, M., Courboulay, V., and Dardignac, P.: “Possibilistic and Probabilistic Fuzzy Clustering: Unification within the Framework of the Nonextensive Thermostatistics,” Pattern Recognition, Vol. 36, pp. 1325–1342, (2003).
 - [23] Yasuda, M.: “Tsallis Entropy Based Fuzzy Cmeans Clustering with Parameter Adjustment,” Proc. SCIS&ISIS2012, pp. 1534–1539, (2012).
 - [24] Kanzawa, Y.: “On Fuzzy Clustering based on Tsallis Entropy-regularization,” Proc. 30th Fuzzy System Symposium, pp. 452–457, (2014).
 - [25] Dhillon, I. S., and Modha, D. S.: “Concept Decompositions for Large Sparse Text Data Using Clustering,” Machine Learning, Vol. 42, pp. 143–175 (2001).
 - [26] Miyamoto, S., and Mizutani, K.: “Fuzzy Multiset Model and Methods of Nonlinear Document Clustering for Information Retrieval”, Lecture Notes in Computer Science, Vol. 3131, pp. 273–283, (2004).
 - [27] G, Karypis.: Karypis Lab, CLUTO - Software for Clustering High-Dimensional Datasets, Available from: <http://glaros.dtc.umn.edu/>
 - [28] Available from: <http://www.dataminingresearch.com/>
 - [29] Lise’s Inquisitive Students, Machine Learning Research Group @UMD, Available from: <http://www.cs.umd.edu/~sen/lbc-proj/LBC.html>
 - [30] Text REtrieval Conf.(TREC),<http://trec.nist.gov>
 - [31] TREC: Text REtrieval Conference Relevance Judgments, http://trec.nist.gov/data/qrels_eng/index.html
 - [32] J. Moore, E. Han, D. Boley, M. Gini, R. Gross, K. Hastings, G. Karypis, V. Kumar, and B. Mobasher.: “Web page categorization and feature selection using association rule and principal component clustering,” In 7th Workshop on Information Technologies and Systems, (1997).
 - [33] W. Hersh, C. Buckley, T.J. Leone, and D. Hickam.: “OHSUMED: An interactive retrieval evaluation and new large test collection for research,” In SIGIR-94, pp. 192–201, (1994).

- [34] D. D. Lewis.: “Reuters-21578 text categorization test collection distribution 1.0,” <http://www.research.att.com/~lewis>, (1999).
- [35] Manning, D. C, Raghavan, P, and Schütze, H.: “Introduction to Information Retrieval,” Cambridge University Press, pp. 109–110, (2008).
- [36] Santos, J. M., and Embrechts, M.: “On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification,” Springer Berlin Heidelberg, pp. 175–184, (2009).
- [37] Aggarwal, CC., Wolf, JL., Wu, WL., and Yu, PS.: “Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering,” Proc. the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 201–212, (1999).
- [38] Toby Segaran, 富山 仁建, 鴨澤 眞夫.: “集合知プログラミング,” 株式会社オライリージャパン, 初版第 8 刷発行, pp. 342, https://resources.oreilly.com/examples/9780596529321/blob/master/PCI_Code%20Folder/chapter2/recommendations.py, (2012).
- [39] Swets, JA.: “ROC Analysis Applied to the Evaluation of Medical Imaging Techniques,” Proc. Investigative Radiology, Vol. 14, pp. 109–121, (1979).
- [40] Hanley, J. A., and McNeil, B. J.: “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” Proc. Radiology, Vol. 143, pp. 29–36, (1982).
- [41] Available from: <http://www2.informatik.uni-freiburg.de/~ciegler/BX/>
- [42] Available from: <http://www.trustlet.org/epinions.html>
- [43] Available from: <http://goldberg.berkeley.edu/jester-data/>
- [44] Available from: <http://www.occamlab.com/petricek/data/>
- [45] Available from: <http://grouplens.org/datasets/movielens/>
- [46] Available from: <http://www.kamishima.net/sushi/>
- [47] Patra, B. K., Launonen, R., Ollikainen, V., and Nandi, S.: “A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data,” Knowledge-Based Systems, Vol. 82, pp. 163–177, (2015).
- [48] Kanzawa, Y.: “On Possibilistic Clustering Methods Based on Shannon/Tsallis-Entropy for Spherical Data and Categorical Multivariate Data,” Lecture Notes in Artificial Intelligence, Vol. 9321, pp. 115–128, (2015).
- [49] Kanzawa, Y.: “ q -Divergence-based Relational Fuzzy c -Means Clustering,” JACIII, Vol. 22, No. 1, pp. 34–43, (2018).

代表的な研究業績

- (1) Kondo, T., and Kanzawa, Y.: “Fuzzy Clustering Methods for Categorical Multivariate Data Based on q -Divergence,” JACIII, Vol. 22, No. 4, pp. 524–536, (2018).
- (2) Kondo, T., and Kanzawa, Y.: “Comparison of Fuzzy Co-Clustering Methods in Collaborative Filtering-based Recommender System,” Lecture Notes in Computer Science, Vol. 10571, pp. 103–116, (2017).
- (3) Kondo, T., and Kanzawa, Y.: “Performance Comparison of Collaborative Filtering Using Fuzzy Clustering for Spherical Data,” Proc. SCIS&ISIS2018, pp. 644–647, (2018).
- (4) Kondo, T., and Kanzawa, Y.: “Collaborative Filtering Using Fuzzy Clustering for Categorical Multivariate Data Based on q -Divergence,” (to appear in 2019).
- (5) 近藤 唯史, 神澤 雄智.: “二項 Latent Block Model に基づく KL ダイバージェンス正則化ファジィ共クラスタリング,” 第 33 回ファジィシステムシンポジウム講演論文集, pp. 141–144, (2017).

感想

学部3年次に研究室配属がそれぞれの希望によって決まるわけですが、自分の場合は、やりたいことが分からなくなるとなくでこの研究室を第一志望にしました。正直、飲み会がよくあるとか先生が飲み代を奮発してくれるとかしょうもない噂を聞いて選択した記憶がありますが、何の研究をしているかくらいは調べてました。最近流行っている人工知能の根本にあるような技術のクラスタリングなどデータ分析の基礎を学べる研究室がなぜ通信工学科(2018年度から情報通信工学科)にあるのか、情報工学科の分野ではないのか誰もが疑問に思っていたでしょう。この学科は、通信工学で必要なハードウェアとソフトウェアの技術をどちらも学べると謳っています。それだけ通信工学は、技術の幅が広いということなのでしょう。もしくはこの研究室は、入学3年目にして期待とは裏腹に無線とか光とかネットワークとか通信の王道が面白くなかったと思う学生やハードウェアの講義の成績が悪い学生の受け皿なのかもしれません。どちらかという自分もそういうタイプの学生でした(かと言ってソフトウェアが抜群にできたわけでもないですが...). 研究室配属が決定しゼミが始まってからは、好奇心を刺激され続けました。他研究室の学生から見ても異色だと言われる長ったらしい方程式を解くのも、C++でオブジェクト指向的なプログラミングをするのも、ベンチマークデータを使って実際に分析してみるのも、 \LaTeX で四苦八苦しながら文書を書くのも、Gnuplotで綺麗に図を描画するのに凝るのも、Linux環境を自分の思い通りにカスタマイズするのも楽しかったです。この好奇心が後々大学院進学したいという気持ちに繋がりました。学部4年になってからは、ゼミで学んだ知識を土台に研究することになります。ゼミの課題が終わり次第随時研究テーマが決定されました。学生個人がやりたいことを提示することもできますが、大半は先生から研究テーマを割り振られることになります。自分は、先生からいくつか提示された中からクラスタリングを推薦システムに応用するテーマを選択しました。比較的身近にある推薦アルゴリズムを自分で実装できることに魅力を感じたのがこのテーマを選んだ理由です。研究を始めて、類似研究の論文や参考書籍を読み漁ってどのように数値実験していくのか少しずつ理解していきました。最初のうちは、コーディングしたプログラムがエラー吐きまくりコアダンプしまくりで苦勞もしましたが、その分それらを修正しともに実行できたときは爽快でした。提案した手法が従来のものより精度の高い予測結果を出したときには大きな達成感を得ました。修士になってからは、研究だけすればよいというわけにもなくなります。大学院の授業やその課題、学会に投稿する論文や発表資料作成、週1のTA、学部3年生のゼミと4年生の卒研補佐、オープンキャンパスの手伝い、M2になれば就職活動など面倒事が増え純粋に研究する時間が減りました。初めての国際学会では、緊張で体の調子がおかしくなりましたが、その分野に詳しい学外の研究者や学生に発表を聞いてもらうのは良い経験になりましたし、校費でちょっとした旅行感覚を味わえてお得でした。後期にある1限からの大宮キャンパスTAは朝5時30分起きで地獄でしたが、4年で卒業した同期は皆社会人として働いていると考えたら些細な悩みだと思えましたし、TAやゼミで後輩へ指導することでより自分の理解を深めることもできました。研究はもちろん研究以外の活動全てに関しても自分の糧になったと思っています。

謝辞

神澤研究室でともに研究に励んだ仲間たちに感謝します。共に学会など経験した同期の東くん、ひとつ下の代で少ないメンバーのなか共に研究に励んだ井部くん、大関くん、加津間くん、渡邊くん、ふたつ下の代で共に就職活動を乗り越つつ研究に励んだ岡本くん、小川くん、上川くん、駒場くん、西田くん、中宮くん、中村くん、宮崎くん、池田くんに感謝します。また、3年間手厚く研究指導をしていただいた神澤先生へ深く感謝いたします。大変お世話になり、ときにはご迷惑をおかけいたしました。ご恩は一生忘れません。最後に大学4年間だけでなく修士へ進学を許していただき、さらに学費まで支援していただいた両親と私生活面で面倒をみてくれた祖母に感謝の言葉を述べてしめます。ありがとうございました。

付録

ソースコードは, github.com で管理していたのでその URL を掲載します.

クラスタリング手法や協調フィルタリングなどのソースコード:

<https://github.com/ntyaaan/clustering>

整形したデータなど:

<https://github.com/ntyaaan/some-datasets>

https://github.com/ntyaaan/clustering_datasets_format

図 6.1 にクラス図を示します.

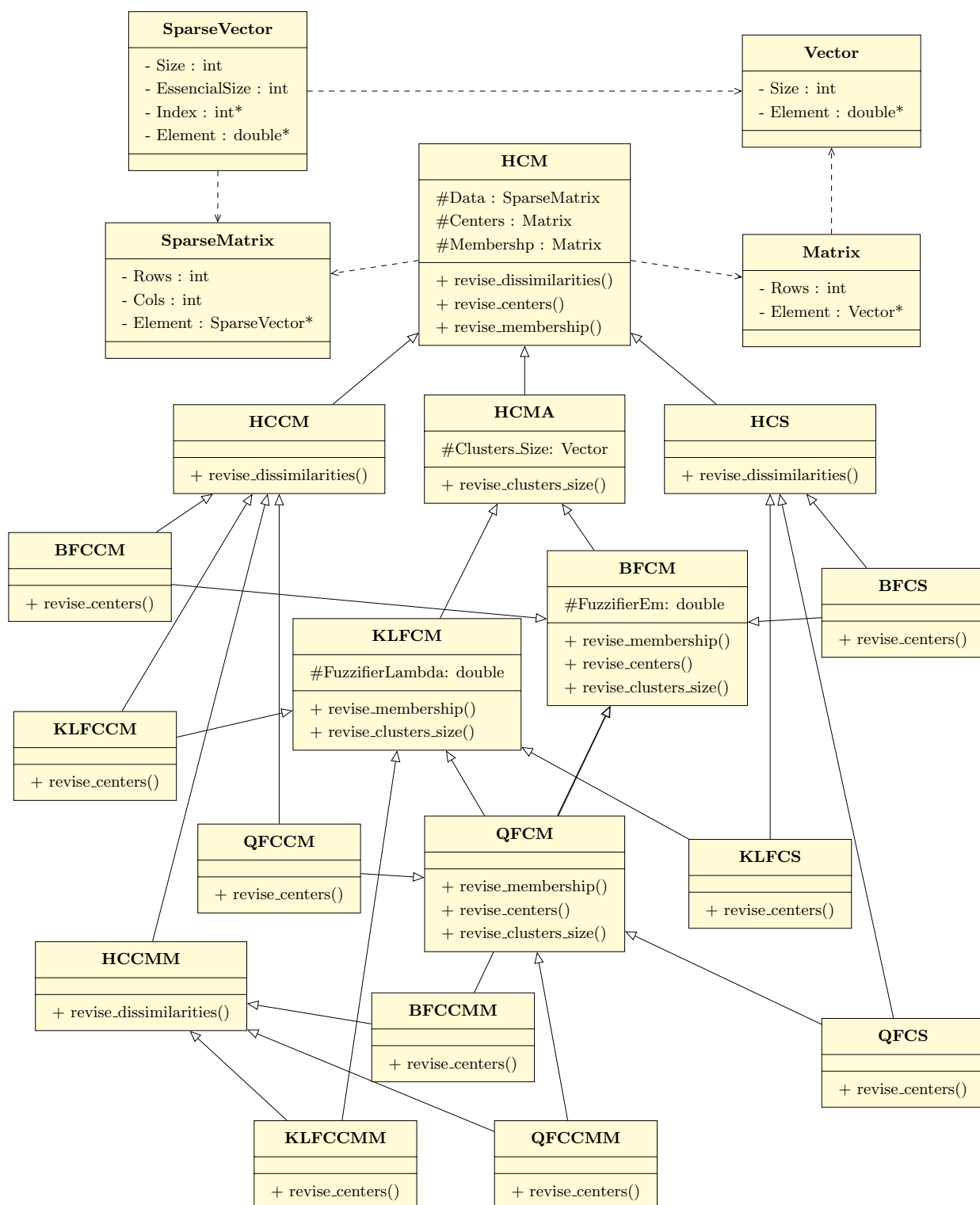


図 6.1: クラス図