Итак, опишем наш эксперимент.

Была дана обрабатываемая нами матрица числовых исходных данных размерностью примерно 261 на 56. Из матрицы вырезались случайным образом 10, 20,…,90 % числовых ячеек данных. Восстановление вырезанных значений проводилось двумя алгоритмами: “алгоритм восстановления анкетных данных” – старая версия алгоритма (та в которой использовалось 10 итераций), “модифицированный алгоритм восстановления анкетных данных” – новая версия алгоритма , та в которой уже используется 25 итераций, сглаживание выбросов значений восстанавливаемых параметров, для каждого восстанавливаемого значения используется критерий остановки восстановления, остановка происходит при пересечении порога в 10 % отклонения среднеарифметического сглаженного значения от максимально зафиксированного на предыдущих итерациях.

Положим –исходное значение восстанавливаемой числовой ячейки матрицы, восстанавливаемое значение экземпляра параметра. – восстановленное значение параметра. и – минимальное и максимальное значения параметра по всем обрабатываемым объектам. Для оценки погрешности одного восстановленного значения нами используется приведённая погрешность, которая вычисляется по формуле: . Приведённая погрешность указывается в процентах от размаха значений параметра. В таблице 3427fh6g467n28 приведены интервалы приведённой погрешность 0% - 10%, 10%-20%, … 80% - 90%, количество восстанавливаемых значений и процент от общего числа восстанавливаемых ячеек, попавших в интервал, также накопленное количество и накопленный процент для нового алгоритма и 10% вырезанных значений.

Таблица №3427fh6g467n28.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Интервал погрешности | Кол-во параметров | Накопленное кол-во параметров | Процент | Накопленный процент |
| 0-10% | 289 | 289 | 20.26648 | 20.26648 |
| 10-20% | 233 | 522 | 16.33941 | 36.60589 |
| 20-30% | 193 | 715 | 13.53436 | 50.14025 |
| 30-40% | 216 | 931 | 15.14727 | 65.28752 |
| 40-50% | 147 | 1078 | 10.30856 | 75.59607 |
| 50-60% | 125 | 1203 | 8.765778 | 84.36185 |
| 60-70% | 117 | 1320 | 8.204769 | 92.56662 |
| 70-80% | 70 | 1390 | 4.908836 | 97.47546 |
| 80-90% | 30 | 1420 | 2.103787 | 99.57924 |
| 90-100% | 6 | 1426 | 0.420757 | 100 |

Такой способ представления слегка не удобен для описания результатов эксперимента по восстановления 10, 20,… 90% вырезанных значений. По этому мы будем для описания одного эксперимента использовать число – площадь под графиком накопленного процента.

Приведём график на коренного процента для нового алгоритма для 10% вырнзанных данных (рис. T43fh8723j94378)

Рисунок №T43fh8723j94378. График приведённой погрешности, модифицированный алгоритм, 10% вырезанных данных.

Если плошать под графиком больше, то это означает, что больший объём данных восстановлен с меньшей приведённой погрешностью. Теперь, оперируя площадью под графиком накопленного процента приведённой погрешности, проиллюстрируем результаты наших экспериментов.

Проводилось восстановление числовых ячеек массива данных размерностью примерно 261 на 56, при использовании алгоритма восстановления анкетных данных и его модификации, при удалении от 10 до 90 процентов вырезанных данных.

Теперь приведём площади под графиками накопленных приведённых погрешностей. На рисунке 6gr65dg4d677b6t: пунктирный график это площади под графиками накопленных приведённых погрешностей алгоритм восстановления , непрерывный график – модифицированный алгоритм восстановления анкетах данных. Большим погрешностям соответствует ниже пролелегающий график

Рисунок №6gr65dg4d677b6t. Площади под графиками накопленных приведённых погрешностей.

Если судить только по бинарным данным (рис. 6gr65d43756387277b6t) или только по числовым (Рисунок №665783465782687b6t), то наблюдается примерно та же картина.

Рисунок №6gr65d43756387277b6t. Площади под графиками накопленных приведённых погрешностей.

Рисунок №665783465782687b6t. Площади под графиками накопленных приведённых погрешностей.д

Выдвигается составная гипотеза о том что, наши восстанавливаемые параметры условно можно разделить на три группы. Первая группа – числовые ячейки, чьи значения сходятся и достигают критерия остановки восстановления значения. Вторая группа – значения обладают более слабой сходимостью, возможно зашумлены и критерий остановки не достижим за отведённое число итераций. Третья группа – критерий остановки не достижим в принципе в обозримом вычислительном горизонте. Первая группа восстанавливаемых ячеек многочисленна – около 90 процентов от общего числа. Вторая и третья группа параметров не столь многочисленны – всего около 10 процентов, но они трудно различимы. Если пока что нет возможности уменьшить численность третьей группы, то можно путём дальнейших модификаций алгоритма свести на нет мощность второй группы.

По мимо этого у нас остались нерассмотренные другие общепризнанные оценки качества. Приведём их в таблице №ghf76hr76hf7nv69.

Таблица №ghf76hr76hf7nv69. Оценки качества результатов экспериментов

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Модифицированный алгоритм | | | | | | Старый алгоритм | | | | | |
| % вырезанных данных | Все данные | | Бинарные данные | | Не бинарные данные | | Все данные | | Бинарные данные | | Не бинарные данные | |
|  | MAE | ME | MAE | ME | MAE | ME | MAE | ME | MAE | ME | MAE | ME |
| 10% | 2.41999 | 1.282415 | 0.362073 | 0.094039 | 7.413533 | 4.173203 | 2.08349 | 0.046927 | 0.340964 | 0.027788 | 6.707497 | 0.096113 |
| 20% | 2.269957 | 0.612593 | 0.342503 | 0.089218 | 6.946719 | 1.882743 | 2.378919 | -0.74161 | 0.355773 | -0.01554 | 7.750254 | -2.67674 |
| 30% | 2.433276 | -0.47331 | 0.336859 | -0.00328 | 7.522239 | -1.6188 | 2.18847 | -0.11721 | 0.362035 | 0.007889 | 7.034829 | -0.45134 |
| 40% | 2.278205 | 0.758322 | 0.349654 | 0.039922 | 6.957194 | 2.506925 | 2.36545 | -0.59189 | 0.402915 | 0.01909 | 7.572015 | -2.22246 |
| 50% | 1.979791 | 0.043624 | 0.310934 | 0.018199 | 6.028195 | 0.104458 | 2.224646 | -0.00568 | 0.358561 | 0.013436 | 7.176969 | -0.05754 |
| 60% | 1.996714 | -0.02144 | 0.315146 | -0.03852 | 6.07584 | 0.022596 | 2.408981 | -0.08285 | 0.383509 | 0.003744 | 7.784673 | -0.31402 |
| 70% | 2.292136 | 0.480111 | 0.333123 | -0.0048 | 6.987596 | 1.62122 | 2.263752 | 0.16635 | 0.390972 | -0.02585 | 7.229717 | 0.682653 |
| 80% | 2.324838 | 0.511754 | 0.346116 | 0.019086 | 7.076946 | 1.696641 | 2.342878 | -0.20006 | 0.400524 | 0.066601 | 7.495787 | -0.91561 |
| 90% | 2.429837 | 0.691606 | 0.355257 | 0.017486 | 7.41601 | 2.3084 | 2.282456 | -0.43643 | 0.347292 | 0.024713 | 7.417399 | -1.66545 |