ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2023, № 3, с. 14–37

КОМПЬЮТЕРНЫЕ МЕТОДЫ

УДК 004.93'14,004.021,519.177

УЛУЧШЕНИЕ НЕПОЛНОЙ МНОГОАСПЕКТНОЙ КЛАСТЕРИЗАЦИИ ТЕНЗОРНЫМ ВОССТАНОВЛЕНИЕМ ГРАФА СВЯЗНОСТИ[1]

© 2023 г. Х. Жанг[a], С. Чен[a,*], Ю. Жу[b], И. А. Матвеев[c,]**

[a]*Колледж математики Нанкинского ун-та аэронавтики и космонавтики, Нанкин, КНР*

[b]*Факультет экспериментального фундаментального обучения, Нанкинский ун-т аэронавтики и космонавтики, Нанкин, КНР*

[c]*ФИЦ ИУ РАН, Москва, Россия*

**e-mail: lyandcxh@nuaa.edu.cn*

***e-mail: matveev@ccas.ru*

Поступила в редакцию 21.12.2022 г.
После доработки 08.01.2023 г.
Принята к публикации 06.02.2023 г.

С развитием технологий сбора данных появляется множество многоаспектных данных, и их кластеризация стала актуальной темой. Большинство методов *многоаспектной кластеризации* (multi-view clustering) предполагают, что все аспекты полностью наблюдаемы. Но во многих случаях это не так. Для работы с неполными многоаспектными данными были предложены некоторые тензорные методы. Однако традиционная тензорная норма требует больших вычислительных затрат, и такие методы, как правило, не могут обрабатывать неполные выборки и дисбаланс различных аспектов. Предлагается новый метод кластеризации неполных многоаспектных данных. Определяется новая тензорная норма для восстановления графа связности, графы регуляризируются до согласованного низкоразмерного представления образцов. Затем веса итеративно обновляются для каждого аспекта. По сравнению с существующими, предложенный метод не только определяет согласованность между аспектами, но и получает низкоразмерное представление образцов с помощью полученной проекционной матрицы. Для решения разработан эффективный оптимизационный алгоритм на основе метода неопределенных множителей Лагранжа. Экспериментальные результаты на четырех наборах данных демонстрируют эффективность метода.

DOI: 10.31857/S0002338823030137, **EDN:** EVCFML

Введение. Многоаспектные данные могут дать больше информации, чем данные с одним аспектом, во многих практических задачах. Многоаспектная кластеризация (МАК) стала актуальной в связи с получением большого количества таких данных [1]. Большинство существующих методов хорошо работают в предположении, что данные являются полными, т.е. для каждого образца есть его представление в каждом аспекте. Однако это предположение не всегда справедливо на практике. Например, при диагностике болезни Альцгеймера [2] по разным причинам не все виды обследований пациента доступны, что приводит к неполным многоаспектным данным. Соответственно кластеризация таких данных в последнее время привлекает все больше внимания.

В последние годы были разработаны некоторые методы *многоаспектной кластеризации неполных данных* (также *неполная многоаспектная кластеризация*, НМАК) (incomplete multi-view clustering, IMVC), некоторые из них кратко описаны в разд. 1. НМАК на основе графов привлекла внимание многих исследователей благодаря тому, что граф является мощным инструментом для анализа взаимосвязей [3]. Неполная многоаспектная спектральная кластеризация с адаптивным обучением графов (incomplete multi-view spectral clustering with adaptive graph learning, IMSC_AGL) [4] использует графы, построенные из низкорангового подпространства каждого аспекта для формирования согласованного низкоразмерного представления каждого образца. Обобщенная НМАК с гибким распространением структуры локальности (generalized

[1] Работа выполнена при частичной финансовой поддержке РФФИ (грант № 21-51-53019); Государственного фонда естественных наук Китая (гранты № 11971231; 12111530001).

incomplete multi-view clustering with flexible locality structure diffusion, GIMC_FLSD) [5] комбинирует обучение графов и факторизацию матриц для получения унифицированного представления, которое сохраняет информацию графа. Неполное обучение неотрицательного представления с несколькими графами (incomplete multi-view non-negative representation learning, IMNRL) [6] выполняет факторизацию матриц нескольких неполных графов для получения согласованного неотрицательного представления с ограничениями на графы. Адаптивная кластеризация неполных графов на основе пополнения графов (adaptive graph completion based incomplete multi-view clustering, AGC_IMC) [7] выполняет последовательное обучение на представлениях, используя представление графа каждого аспекта в качестве регуляризации неполных данных других аспектов, и далее получает согласованное низкоразмерное представление. Эти методы применяют обучение графов для фиксации локальной топологии представления. Однако при этом учитывается только внутриаспектное сходство представленных образцов и игнорируется межаспектное сходство. Поэтому эта группа методов не может эффективно найти дополнительную информацию, скрытую в различных аспектах.

Корреляция – классическая мера, которую можно применить для получения дополнительной информации из многоаспектных данных. Для работы с неполными многоаспектными данными, содержащими информацию о корреляции между аспектами, предложено несколько тензорных методов. В работе [8] используются представления подпространства с тензорными ограничениями низкого ранга, чтобы исследовать внутри- и межаспектные связи между образцами и одновременно улавливать корреляции высокого порядка между несколькими аспектами. Неполная многоаспектная тензорная спектральная кластеризация с выводом отсутствующих представлений (incomplete multi-view tensor spectral clustering with missing view inferring, IMVTSC-MVI) [9] объединяет в единую структуру пространство признаков, пополненное отсутствующими представлениями, и пространство многообразий, основанное на обучении графов сходства. Стоит отметить, что IMVTSC-MVI рассматривает каждый аспект одинаково, т.е. с одинаковым весом. В [10] предлагается получать полный граф для каждого аспекта, учитывая сходство графов между аспектами и применяя метод дополнения тензоров. Хотя вышеупомянутые методы на основе тензоров достигли хороших результатов, все еще существуют некоторые проблемы. В большинстве известных работ ограничение низкого ранга действует для преобразованного графа, что значительно увеличивает вычислительные затраты. Это связано с тем, что необходимо выполнение сингулярного разложения на каждом фронтальном срезе тензора $n \times m \times n$, где n и m – количество образцов и аспектов соответственно. С этой целью в данной работе для уменьшения вычислительной сложности определяется ядерная норма тензора, использующая полуортогональную матрицу.

Предлагаемый метод называется *улучшение неполной многоаспектной кластеризации пополнением тензорного графа* (enhancing incomplete multi-view clustering via tensor graph completion, EIMC_TGC). Метод объединяет пополнение тензорного графа, адаптивное взвешивание аспектов и последовательную кластеризацию подпространств в единую структуру. Отличительная особенность метода – переопределение ядерной нормы тензора на основе полуортогональной матрицы для сокращения вычислений и определение оператора сжатия для получения аналитических решений подзадач. Вклад данной работы сводится к следующим выводам.

1. Предлагается метод HMAK, использующий тензорное пополнение. В отличие от существующего метода MAK на основе тензорной нормы, мы переопределяем тензорную ядерную норму через полуортогональную, что снижает вычислительные затраты, и адаптивно назначаем веса каждому аспекту.

2. Для получения аналитического решения подзадачи с переопределенной ядерной нормой предусмотрен оператор сжатия. Предлагается алгоритм оптимизации на основе расширенного метода множителей Лагранжа (augmented Lagrange multiplier, ALM).

3. Эксперименты на четырех наборах данных подтверждают, что этот метод эффективен в задачах кластеризации.

Структура данной работы следующая. В разд. 1 кратко представлены известные аналогичные работы. В разд. 2 изложен предлагаемый метод и дан алгоритм решения. В разд. 3 описана проверка эффективности метода численными экспериментами. В разд. 3.4 подводятся итоги работы.

1. Известные аналоги. В этом разделе представлены известные из литературы последних лет методы обработки неполных многоаспектных данных, а также некоторые обозначения и базовые определения. Методы HMAK можно разделить на пять категорий по основному используе-

мому подходу: кластеризация подпространств, факторизация неотрицательных матриц, обучение графов, глубокие нейронные сети, вероятностный подход.

Методы кластеризации подпространства [11–16] проецируют все существующие аспекты в общее низкоразмерное подпространство и выравнивают все представления в этом подпространстве для получения согласованного низкоразмерного представления. Например, в [15] используется разреженное низкоранговое представление подпространства для совместного измерения внутриаспектных и межаспектных отношений. В [12] предложен полупарный обобщенный корреляционный анализ с частичным обучением (semi-paired and semi-supervised generalized correlation analysis, S^2GCA) на основе канонического корреляционного анализа (canonical correlation analysis, CCA), который обрабатывает неполные данные и применим к сценарию с частичным привлечением учителя.

Методы на базе неотрицательной матричной факторизации (non-negative matrix factorization, NMF) [17–22] направлены на разложение матрицы данных каждого аспекта на произведение одной общей (базисной) матрицы и матрицы, соответствующей каждому из аспектов. Например, в [17] предложена частичная МАК на основе NMF для изучения общего представления для полных выборок и частного скрытого представления с той же матрицей базиса. Многоаспектное обучение с неполными аспектами (multi-view learning with incomplete views, MVL-IV) [18] факторизует матрицу неполных данных каждого аспекта как произведение общей матрицы и базисной матрицы соответствующего аспекта, имеющей низкий ранг.

Методы на основе обучения графа [23–27] строят общую матрицу графа связности (или матрицу ядра), так чтобы сохранить геометрическую структуру. Эти методы используют спектральную кластеризацию и многоядерное обучение. Например, НМАК адаптивным пополнением графа (adaptive graph completion based incomplete multi-view clustering, AGC_IMC) [7] получает общий граф с помощью специальной регуляризации и далее добивается согласованного низкоразмерного представления образцов. Метод неполных многоядерных k-средних с взаимным дополнением ядер (incomplete multiple kernel k-means with mutual kernel completion, MKKM-IK-MKC) [23] получает неполные матрицы ядер и осуществляет интерполяцию.

Методы на основе глубоких нейронных сетей [28–32] используют мощную способность нейронной сети к выделению признаков. Например, в [29] разработан новый метод НМАК, который проецирует все данные нескольких видов на полное и единое представление с ограничением присущей геометрической структуры. В [31, 32] предложена НМАК с применением генеративной состязательной сети [33], которая порождает недостающие данные на основе имеющихся. В [34] единое представление получается путем максимизации взаимной информации между различными аспектами посредством контрастного обучения, недостающие данные восстанавливаются путем минимизации условной энтропии различных аспектов.

Методы, основанные на вероятностном подходе [35–38], обычно используют байесовскую статистику, гауссовские модели, вариационный вывод и другие инструменты для моделирования многоаспектных данных с вероятностной точки зрения. Например, в [37] применяется модель латентных переменных общего гауссовского процесса для моделирования неполных многоаспектных данных для кластеризации. В [36] для обработки неполных многоаспектных данных предложено объединение байесовского канонического корреляционного анализа [39] и обучения с учителем [40].

Существует два подхода к использованию тензоров в многоаспектном обучении. Первый заключается в том, чтобы рассматривать различные аспекты как подпространства пространства высокой размерности и обрабатывать их соответственно. Представителем первого подхода является тензорный канонический корреляционный анализ (tensor canonical correlation analysis, TCCA) [41]. TCCA определяет ковариационный тензор для обработки многоаспектных данных с произвольным количеством аспектов, расширяя матрицу взаимной ковариации в каноническом корреляционном анализе. Глубокий TCCA (deep TCCA, DTCCA) [42] объединяет TCCA с глубоким обучением нейросетей. DTCCA может не только обрабатывать данные с произвольным количеством аспектов, но и не требует хранения всего тензора, что означает, что он может обрабатывать большое количество данных. В [43] рассматривались многоаспектные данные как тензор размерности 3, для вычисления низкоразмерного представления образцов использовано разложение Такера.

Второй подход предполагает, что графы различных представлений похожи, т.е. тензор данных имеет низкий ранг. Низкоранговая тензорная многоаспектная кластеризация подпространств (low-rank tensor constrained multi-view subspace clustering, LT-MSC) [44] рассматривает совокупность матриц данных различных аспектов как тензор, ранг которого должен быть ограничен, что

Таблица 1. Некоторые используемые обозначения

Обозначение	Описание
$\mathcal{X}_{(i,j,k)}$	Элемент i-й строки, j-го столбца, k-го слоя тензора $\mathcal{X}$
$\mathcal{X}^{(i)}$	i-й слой тензора $\mathcal{X}$
$\mathcal{X}_{(:,i,:)}$	i-й боковой срез тензора $\mathcal{X}$
$\mathcal{X}_{(i,j)}$	Элемент i-й строки, j-го столбца матрицы X
$X_{(:,i)}$	i-й столбец матрицы $\mathcal{X}$
$X^\top$	Транспонированная матрица X
$\|X\|_*$	Ядерная (следовая) норма матрицы X
$\|X\|_F$	Норма Фробениуса матрицы X, согласно (2.2)
$\|\mathcal{X}\|_F$	Норма Фробениуса тензора $\mathcal{X}$, согласно (2.3)
$\sigma_i(X)$	i-е наименьшее сингулярное значение матрицы X
$\langle \mathcal{X}, \mathcal{Y} \rangle$	Скалярное произведение тензоров $\mathcal{X}$ и $\mathcal{Y}$, т.е. $\sum_{i,j,k} \mathcal{X}_{(i,j,k)} \mathcal{Y}_{(i,j,k)}$
$\nabla f(x)$	Градиент функции $f(x)$
x_i	i-й элемент вектора x

уменьшает избыточность данных и повышает качество кластеризации. В [45] применен известный своей устойчивостью метод анализа главных компонент, сохраняющий низкий ранг тензора, построенного на основе вероятностых переходных матриц.

1.1. О б о з н а ч е н и я. Матрицы обозначаются курсивными прописными буквами, например $X \in \mathbb{R}^{n_1 \times n_2}$. Элемент матрицы задается двумя индексами (номерами строки и столбца), таким образом матрица имеет *размерность* 2. Несколько матриц размерности 2 одного размера можно объединить в матрицу размерности 3, далее называемую *тензором*[2]. Матрицы, составляющие тензор, называются *слоями*. Тензоры обозначаются заглавными каллиграфическими буквами, например $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ – тензор из n_1 строк, n_2 столбцов и n_3 слоев. Некоторые основные обозначения приведены в табл. 1.

О п р е д е л е н и е 1. Рассмотрим тензор $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Слой с номером k представляет собой матрицу, составленную из элементов тензора $\mathcal{X}$, с третьим индексом, равным k: $\mathcal{X}_{(:,:,k)} \in \mathbb{R}^{n_1 \times n_2}$, $k = \overline{1, n_3}$. Для краткости будем обозначать $X^{(k)} \equiv \mathcal{X}_{(:,:,k)}$. Также можно называть слой *фронтальным срезом*.

О п р е д е л е н и е 2. Аналогично j-м *боковым срезом* назовем матрицу, составленную из элементов тензора с фиксированным вторым индексом (номером столбца): $\mathcal{X}_{(:,j,:)} \in \mathbb{R}^{n_1 \times n_3}$, $j = \overline{1, n_2}$.

О п р е д е л е н и е 3. *Трубкой* тензора $\mathcal{X}$ назовем вектор, составленный из элементов тензора с двумя фиксированными индексами. Здесь используются трубки с фиксированными первыми двумя индексами: $x = \mathcal{X}_{(i,j,:)} \in \mathbb{R}^{n_3}$.

О п р е д е л е н и е 4. Для тензора $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ введем операцию *разворачивания*:

$$\operatorname{unfold}(\mathcal{X}) = \begin{bmatrix} X^{(1)} \\ X^{(2)} \\ \vdots \\ X^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}, \tag{1.1}$$

[2] Данное определение не совпадает с общепринятым в математике, поскольку не задана система координат и закон преобразования при ее изменении. Однако в информатике имеет место именно такое понимание термина.

которая переводит его в матрицу.

О п р е д е л е н и е 5. Обратная операция *сворачивания* переводит матрицу в тензор:

$$\text{fold}\left(\text{unfold}\left(\mathscr{X}\right)\right) = \mathscr{X}. \tag{1.2}$$

О п р е д е л е н и е 6. Для тензора $\mathscr{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ и матрицы $\Phi \in \mathbb{R}^{n_3 \times n_3}$ определим *трансформированный тензор* $\mathscr{X}_\Phi \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, составленный из боковых срезов исходного тензора, умноженных на матрицу Φ, так что $\mathscr{X}_{\Phi(:,i,:)} = \mathscr{X}_{(:,i,:)}\Phi$.

2. Предлагаемый метод. Канонический корреляционный анализ (KKA) является классическим и эффективным методом для многоаспектного обучения. KKA может обрабатывать только два аспекта. По этой причине для обработки данных с большим числом аспектов предлагается обобщенный KKA (generalized CCA, GCCA) [46]. В частности, дается выборка из n образцов в m аспектах $\left\{X_i \in \mathbb{R}^{d_i \times n}\right\}_{i=1}^m$, где d_i – размерность i-го аспекта. Тогда модель GCCA представляется как

$$\begin{gathered} \min_{A,\{W_i\}_{i=1}^m} \sum_{i=1}^m \left\|A - W_i^\top X_i\right\|_F^2 \\ \text{s.t.} \quad AA^\top = I, \end{gathered} \tag{2.1}$$

где запись s.t. обозначает “при условии”, I – единичная матрица, $\|A\|_F$ – норма Фробениуса матрицы A:

$$\|A\|_F = \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} A_{(i,j)}^2}. \tag{2.2}$$

Норма Фробениуса для тензора определяет аналогично:

$$\|\mathscr{A}\|_F = \sqrt{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} \mathscr{A}_{(i,j,k)}^2}. \tag{2.3}$$

Символы A и $\{W_i\}_{i=1}^m$ в уравнении (2.1) – это последовательное низкоразмерное представление образцов и матрица проекции каждого аспекта соответственно. Геометрическая структура среди образцов является достоверной априорной информацией, которая может быть эффективно закодирована графом. Следовательно, для получения согласованного низкоразмерного представления с сохранением структуры предлагается графовый многоаспектный KKA (graph multi-view CCA, GMCCA) [47], целевая функция приобретает следующий вид:

$$\begin{gathered} \min_{A,\{W_i\}_{i=1}^m} \sum_{i=1}^m \left\|A - W_i^\top X_i\right\|_F^2 + \lambda tr\left(AL_G A^\top\right) \\ \text{s.t.} \quad AA^\top = I, \end{gathered} \tag{2.4}$$

где L_G – матрица лапласиана предварительно построенного графа G, а $\lambda > 0$ – параметр регуляризации.

К сожалению, вышеупомянутый метод применим только к полным многоаспектным данным. Здесь на основе GMCCA и восстановления тензоров разработан новый метод для неполных многоаспектных данных. Для заданного неполного набора данных $\{X_i \in \mathbb{R}^{d_i} \times n\}_{i=1}^m$ с m аспектами и n образцами, недостающие экземпляры которых заполняются нулевыми векторами, т.е. если j-й экземпляр i-го представления отсутствует, то $X_{i(:,j)} = \vec{0}$. Недостающая информация каждого представления записывается в диагональную матрицу $\left\{P_i \in \mathbb{R}^{n \times n}\right\}_{i=1}^m$. P_i определяется следующим образом:

$$P_{i(j,j)} = \begin{cases} 1, & \text{если } j\text{-й образец описан в } i\text{-м аспекте,} \\ 0 & \text{иначе.} \end{cases} \tag{2.5}$$

Символ d_i обозначает размерность признака i-го представления. Наивная модель для обработки неполных многоаспектных данных, основанная на (2.4), выглядит следующим образом:

$$\min_{A,\{W_i\}_{i=1}^m} \sum_{i=1}^{m} \left(\left\| (A - W_i^\top X_i) P_i \right\|_F^2 + \lambda \operatorname{tr}(A L_{G_i} A^\top) \right) \quad \text{s.t.} \quad AA^\top = I, \tag{2.6}$$

где L_{G_i} – матрица лапласиана предварительно построенного графа G_i i-го представления. Стоит отметить, что модель (2.6) рассматривает каждое представление одинаково, что обычно нереалистично. По этой причине аналогично [48] вводятся веса $\{\overline{\delta}_i\}_{i=1}^m$ для взвешивания важности различных представлений:

$$\overline{\delta}_i = \frac{1}{\sqrt{\left\| (A - W_i^\top X_i) P_i \right\|_F^2 + \lambda \operatorname{tr}(A L_{G_i} A^\top)}}, \quad i = \overline{1, m}. \tag{2.7}$$

Интуитивно понятно, что если i-й аспект может предоставить больше полезной информации, то $\left\| \left(A - W_i^\top X_i\right) P_i \right\|_F^2$ и $\operatorname{tr}\left(A L_{G_i} A^\top\right)$ должны быть малы, и поэтому вес $\overline{\delta}_i$ для этого аспекта будет большим в соответствии с уравнением (2.7) и наоборот. Однако если представление имеет высокий процент пропусков, то вес этого представления также будет больше. Это происходит потому, что высокий процент пропусков приводит к тому, что P_i содержит больше нулей, что в свою очередь делает $\overline{\delta}_i$ больше. Чтобы избежать этого явления, вес изменен на

$$\delta_i = n_i \overline{\delta}_i \,, \tag{2.8}$$

где n_i – количество доступных экземпляров i-го аспекта.

Невозможно напрямую построить полный граф G_i для каждого аспекта из-за неполноты данных. Поэтому его необходимо дополнить [7, 23]. Согласованность между аспектами является важным свойством реальных данных. Соответственно графы различных аспектов должны быть очень похожими. Это свойство хорошо отражается низким рангом тензора, составленного из этих графов (матриц смежности). Следовательно, исходя из восстановления тензора, целевая функция для восстановления графов выражается следующим образом:

$$\min_{\mathcal{G}} \quad \|\mathcal{G}\|_{\Phi,w,S_p}^p + \frac{\gamma}{2} \left\| P_\Omega(\mathcal{G}) - P_\Omega(\mathcal{M}) \right\|_F^2 \quad \text{s.t.} \quad G_i \geq 0, \quad G_i 1 = 1, \tag{2.9}$$

где $\mathcal{G} \in \mathbb{R}^{n \times m \times n}$ и $\mathcal{G}_{(:,i,:)} = G_i$, 1 – вектор соответствующей размерности со всеми единичными компонентами, $\mathcal{M} \in \mathbb{R}^{n \times m \times n}$ – тензор предварительно построенного неполного графа с недостающими позициями, заполненными нулями, где $\mathcal{M}_{(:,i,:)} = M_i$ – предварительно построенный граф i-го аспекта, $P_\Omega(\cdot)$ – ортогональная проекция на линейное подпространство тензора, определенного на $\Omega = \{(i,j,k) \mid \mathcal{M}_{(i,j,k)} \text{не пропущено}\}$:

$$P_\Omega(\mathcal{M})_{(i,j,k)} = \begin{cases} \mathcal{M}_{(i,j,k)}, & \text{если} (i,j,k) \in \Omega, \\ 0 & \text{иначе.} \end{cases} \tag{2.10}$$

Тензорная норма $\|\cdot\|_{\Phi,w,S_p}$ определена следующим образом.

О п р е д е л е н и е 7. Пусть задан тензор $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $h = \min(n_1, n_3)$, вектор весов $w \in \mathbb{R}^h$, при этом $w_1 \geq w_2 \geq \cdots \geq w_l \geq 0$, $0 < p \leq 1$, и матрица преобразования $\Phi \in \mathbb{R}^{n_3 \times r}$, тогда

$$\|\mathcal{X}\|_{\Phi,w,S_p} = \left(\sum_{i=1}^{n_3} \left\| X_\Phi^{(i)} \right\|_{w,S_p}^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n_3} \sum_{j=1}^{h} w_j \sigma_j \left(X_\Phi^{(i)} \right)^p \right)^{\frac{1}{p}}, \tag{2.11}$$

где $X_\Phi^{(i)}$ – i-й фронтальный срез тензора $\mathcal{X}$, трансформированного матрицей Φ.

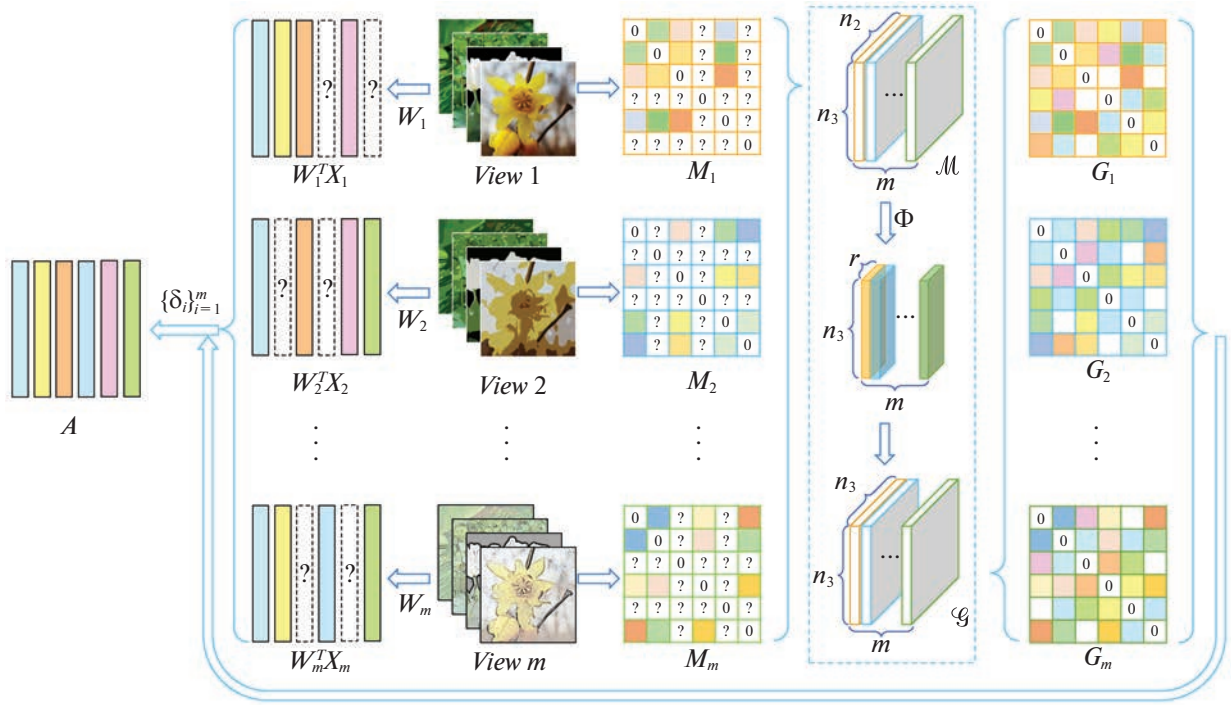

Рис. 1. Схема метода EIMC_TGC

Примечание 1. Заметим, что в литературе [9, 10] матрица преобразования обычно является ортогональной квадратной матрицей, т.е. $\Phi^\top \Phi = I$ и $r = n_3$. Наше эмпирическое исследование показывает, что в этом нет необходимости. Поэтому в данной работе в качестве матрицы преобразования используется полуортогональная матрица, т.е. $\Phi^\top \Phi = I$ и $r < n_3$. Это позволит сократить вычислительные затраты, т.е. если раньше требовалось n_3 разложения по сингулярным значениям, то теперь требуется только r. Экспериментальное исследование в разд. 3.4 показывает, что $r = n_3$ может быть не оптимальным.

Объединив модель восстановления графов (2.9) с (2.6) и введя веса $\{\delta_i\}_{i=1}^m$ согласно (2.7) и (2.8), получаем следующую модель:

$$\min_{A,\{W_i\}_{i=1}^m,\mathcal{G}} \left\{ \sum_{i=1}^m n_i \sqrt{\left\| (A - W_i^\top X_i) P_i \right\|_F^2 + \lambda \operatorname{tr}(A L_{G_i} A^\top)} + \right.$$
$$\left. + \mu \|\mathcal{G}\|_{\Phi,w,S_p}^p + \frac{\gamma}{2} \|P_\Omega(\mathcal{G}) - P_\Omega(\mathcal{M})\|_F^2 \right\} \tag{2.12}$$
$$\text{s.t.} \quad AA^\top = I, \quad G_i \geq 0, \quad G_i \mathbf{1} = \mathbf{1},$$

где $\mu > 0$ – параметр регуляризации. Схема предложенного метода дана на рис. 1.

Примечание 2. Известно [49], что количество нулевых собственных значений матрицы Лапласиана L_G должно быть равно количеству связных компонент графа G. Другими словами, для графа с c компонент связности

$$\sum_{i=1}^c \sigma_i(L_G) = 0. \tag{2.13}$$

Имеет место следующее уравнение:

$$\sum_{i=1}^c \sigma_i(L_G) = \min_{FF^\top = I, F \in \mathbb{R}^{c \times n}} \operatorname{tr}(F L_G F^\top). \tag{2.14}$$

Поскольку правая часть (2.14) формально идентична члену $\mathrm{tr}(AL_{G_i}A^\top)$ в целевой функции (2.12), можно ожидать, что решение задачи (2.12) породит разбиение на малое число компонент, т.е. хорошую кластеризацию. Примечательно, что наш метод может работать с неполными выборками с выученными $\{W_i\}_{i=1}^m$.

2.1. Р е ш е н и е. Для эффективного решения задачи (2.12) нам необходимо ввести вспомогательную переменную $\mathcal{Y}$, чтобы разделить взаимозависимые члены так, чтобы они могли быть решены независимо. Таким образом, можно переформулировать задачу (2.12) в следующую эквивалентную форму:

$$\begin{aligned}\min_{A,\{W_i\}_{i=1}^m,\mathcal{G}} &\left\{\sum_{i=1}^m\left(\delta_i\left\|(A-W_i^\top X_i)P_i\right\|_F^2+\lambda\delta_i\mathrm{tr}(AL_{G_i}A^\top)\right)+\right.\\ &\left.+\mu\|\mathcal{Y}\|_{\Phi,w,S_p}^p+\frac{\gamma}{2}\|P_\Omega(\mathcal{G})-P_\Omega(\mathcal{M})\|_F^2\right\}\\ \text{s.t.}\quad & AA^\top=I,\quad G_i\geq 0,\quad G_i\mathbf{1}=\mathbf{1},\quad \mathcal{G}=\mathcal{Y},\end{aligned}\tag{2.15}$$

где δ_i задается (2.7). Воспользовавшись недавними достижениями в области методов переменных направлений, предложим эффективный алгоритм, основанный на методе неопределенных множителей Лагранжа (inexact augmented Lagrange multipliers, IALM) для решения (2.15). Расширенная функция Лагранжа задается следующим образом:

$$\begin{aligned}L_\rho(A,\{W_i\}_{i=1}^m,\mathcal{G},\mathcal{Y},\mathcal{C})&=\sum_{i=1}^m(\delta_i\left\|(A-W_i^\top X_i)P_i\right\|_F^2+\lambda\delta_i\mathrm{tr}(AL_{G_i}A^\top))+\\ &+\mu\|\mathcal{Y}\|_{\Phi,w,S_p}^p+\frac{\gamma}{2}\|P_\Omega(\mathcal{G})-P_\Omega(\mathcal{M})\|_F^2+\langle\mathcal{C},\mathcal{G}-\mathcal{Y}\rangle+\frac{\rho}{2}\|\mathcal{G}-\mathcal{Y}\|_F^2,\end{aligned}\tag{2.16}$$

где $\rho>0$ – штрафной параметр, $\mathcal{C}\in\mathbb{R}^{n\times m\times n}$ – множители Лагранжа. В этом разделе используются надстрочные знаки для обозначения количества итераций, т.е. A^k или A^{k+1} и т.д.

2.1.1. П е р е с ч е т A^{k+1} и $\left\{W_i^{k+1}\right\}_{i=1}^m$ с ф и к с и р о в а н н ы м и $\mathcal{G}^k$ и $\left\{\delta_i^k\right\}_{i=1}^m$. Для обновления A^{k+1} и $\left\{W_i^{k+1}\right\}_{i=1}^m$ рассмотрим следующие задачи оптимизации:

$$\begin{aligned}&A^{k+1},\left\{W_i^{k+1}\right\}_{i=1}^m=\\ &=\arg\min_{A,\{W_i\}_{i=1}^m}\sum_{i=1}^m\left(\delta_i^k\left\|(A-W_i^\top X_i)P_i\right\|_F^2+\lambda\delta_i^k\mathrm{tr}(AL_{G_i^k}A^\top)\right)\\ &s.t.\quad AA^\top=I.\end{aligned}\tag{2.17}$$

После простых алгебраических операций задача (2.17) может быть преобразована в следующую задачу на собственные значения:

$$\left(\sum_{i=1}^m\delta_i^k\left(P_i-P_iX_i^\top\left(X_iP_iX_i^\top\right)^{-1}X_iP_i+\lambda L_{G_i^k}\right)\right)A^\top=A^\top\Sigma,\tag{2.18}$$

где Σ – диагональная матрица собственных значений. Тогда $A^\top$ можно получить из собственного вектора, соответствующего первым d наибольшим собственным значениям, где d – размерность последовательного низкоразмерного представления (т.е. $A\in\mathbb{R}^{d\times n}$). Имея A^{k+1}, можно получить $W_i^{k+1}=\left(X_iP_iX_i^\top\right)^{-1}X_iP_iA^{k+1\top}$ для $i=\overline{1,m}$. После получения A^{k+1} и $\left\{W_i^{k+1}\right\}_{i=1}^m$ веса $\left\{\delta_i^{k+1}\right\}_{i=1}^m$ могут быть обновлены в соответствии с определением (2.7).

2.1.2. Пересчет $\mathcal{G}^{k+1}$ с фиксированными A^{k+1}, $\left\{\delta_i^{k+1}\right\}_{i=1}^m$, $\mathcal{Y}^k$, $\mathcal{C}^k$ и ρ^k. Чтобы получить $\mathcal{G}^{k+1}$, фиксируем другую переменную и решаем следующую задачу оптимизации:

$$\begin{aligned}\mathcal{G}^{k+1} = \arg\min_{\mathcal{G}} \Bigg\{ & \lambda \sum_{i=1}^{m} \delta_i^{k+1} \operatorname{tr}\left(A^{k+1} L_{G_i} A^{k+1\top}\right) + \\ & + \frac{\gamma}{2}\left\|P_\Omega(\mathcal{G}) - P_\Omega(\mathcal{M})\right\|_F^2 + \frac{\rho^k}{2}\left\|\mathcal{G} - \mathcal{Y}^k + \frac{\mathcal{C}^k}{\rho^k}\right\|_F^2 \Bigg\} \\ & s.t. \quad G_i \geq 0, \quad G_i \mathbf{1} = \mathbf{1}.\end{aligned} \tag{2.19}$$

Задача (2.19) – это задача наименьших квадратов с ограничениями. Заметим, что каждая трубка тензора задачи (2.19) независима, поэтому она может быть преобразована в nm независимых подзадач. Без потери общности пусть $g \in \mathbb{R}^n$, $s \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, $\omega \in \mathbb{R}^n$, $c \in \mathbb{R}^n$ и $t \in \mathbb{R}^n$ – определенные трубки тензоров $\mathcal{G}$, $\mathcal{M}$, $\mathcal{Y}$, Ω, $\mathcal{C}$ и $\mathcal{T}$ соответственно, где $\mathcal{T}^k_{(i,j,l)} = \lambda\delta_l^{k+1}\left\|A^{k+1}_{(:,i)} - A^{k+1}_{(:,j)}\right\|_F^2 \Big/ 2$. Таким образом, можно получить следующую оптимизационную задачу:

$$\begin{aligned} g^{k+1} = \arg\min_{g} & \frac{1}{\rho^k} t^{k\top} g + \frac{\gamma}{2\rho^k}\left\|P_\omega(g) - P_\omega(z)\right\|_F^2 + \frac{1}{2}\left\|g - y^k + \frac{c^k}{\rho^k}\right\|_F^2 \\ & \text{s.t.} \quad g \geq 0, \quad \mathbf{1}^\top g = 1. \end{aligned} \tag{2.20}$$

Функцию Лагранжа для (2.20) запишем как

$$\begin{aligned} L(g,\alpha,\beta) = & \frac{1}{\rho^k} t^{k\top} g + \frac{\gamma}{2\rho^k}\left\|P_\omega(g) - P_\omega(z)\right\|_F^2 + \frac{1}{2}\left\|g - y^k + \frac{c^k}{\rho^k}\right\|_F^2 - \\ & - \alpha(\mathbf{1}^\top g - 1) - \beta^\top g, \end{aligned} \tag{2.21}$$

где $\alpha \in \mathbb{R}$ и $\beta \in \mathbb{R}^n$ являются множителями. Если взять производную функции Лагранжа по g и положить ее равной нулю, то получится уравнение

$$g - y^k + \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k}\left(P_\omega(g) - P_\omega(z)\right) + \frac{1}{\rho^k} t^k - \alpha\mathbf{1} - \beta = 0. \tag{2.22}$$

Согласно условию ККТ, т.е. $1^\top g = 1$, имеем

$$\alpha = \frac{1}{n}\left(1 - \mathbf{1}^\top y^k + \mathbf{1}^\top \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k}\mathbf{1}^\top\left(P_\omega(g) - P_\omega(z)\right) + \frac{1}{\rho^k}\mathbf{1}^\top t^k - \mathbf{1}^\top\beta\right). \tag{2.23}$$

Сочетая (2.22) и (2.23), получим

$$\begin{aligned} g + \frac{\gamma}{\rho^k} P_\omega(g) = & \; y^k - \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k} P_\omega(z) - \frac{1}{\rho^k} t^k - \\ & - \frac{1}{n}\mathbf{1}\mathbf{1}^\top\left(y^k - \frac{c^k}{\rho^k} + \frac{\gamma}{\rho^k} + P_\omega(z) - \frac{1}{\rho^k} t^k\right) + \frac{1}{n}\mathbf{1} + \\ & + \frac{1}{n}\mathbf{1}\mathbf{1}^\top\left(\frac{\gamma}{\rho^k} P_\omega(g) - \beta\right) + \beta. \end{aligned} \tag{2.24}$$

Далее, согласно дополнительной релаксации в условии ККТ, имеем

$$g_i^{k+1} + \frac{\gamma}{\rho^k} P_\omega(g^{k+1})_i = \left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+, \tag{2.25}$$

где

$$u^k = y^k - \frac{c^k + \gamma P_\omega(z) - t^k}{\rho^k},$$
$$v^k = \frac{1}{n}\mathbf{1}^\top\left(\frac{\gamma P_\omega(g^{k+1})}{\rho^k} - \beta\right) \tag{2.26}$$

и применяется обозначение $(\cdot)_+ = \max(0,\cdot)$. Итак, можно получить

$$g_i^{k+1} = \begin{cases} \left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+, & \text{если} \quad i \in \omega, \\ \frac{\rho^k}{\rho^k + \gamma}\left(u_i^k - \frac{1}{n}\mathbf{1}^\top u^k + \frac{1}{n} + v^k\right)_+ & \text{иначе.} \end{cases} \tag{2.27}$$

Поскольку $\mathbf{1}^\top g^{k+1} = 1$, значение v^k можно найти как корни следующей функции:

$$f(v) = \sum_{i=1}^{n} g_i^{k+1} - 1. \tag{2.28}$$

Уравнение (2.28) может быть решено методом Ньютона. Производная первого порядка от $f(v)$ имеет вид

$$\nabla f(v) = \sum_{i=1}^{n} \nabla g_i^{k+1}, \tag{2.29}$$

где

$$\nabla g_i^{k+1} = \begin{cases} 1, & \text{если } i \in \omega, \\ \frac{\rho^k}{\rho^k + \gamma} & \text{иначе.} \end{cases} \tag{2.30}$$

Решение задачи (2.20) (получение $\mathcal{G}$) представим алгоритмом 1.

Алгоритм 1.

Вход: t^k, ρ^k, ω, z, c^k, t^k.

Выход: Тензор $\mathcal{G}^{k+1}$.

Шаг 1. Инициализация: Номер шага $q = 0$, $v^0 = 0$.

Шаг 2. Вычислить u^k, согласно (2.26).

Шаг 3. Вычислить $f(v^q)$, согласно (2.28).

Шаг 4. Если $\left|f(v^q)\right| \le 10^{-10}$, то перейти к шагу 8.

Шаг 5. Вычислить градиент $f(v)$ в точке v^q, согласно (2.29).

Шаг 6. Пересчитать $v^{q+1} = v^q - f(v^q)/\nabla f(v^q)$.

Шаг 7. $q \leftarrow q + 1$ и перейти к шагу 3.

Шаг 8. Вычислить g^{k+1}, согласно (2.27).

Шаг 9. Составить тензор $\mathcal{G}^{k+1}$ из трубок g^{k+1}.

2.1.3. Пересчет $\mathcal{Y}^{k+1}$ с фиксированными $\mathcal{G}^{k+1}$, $\mathcal{C}^k$ и ρ^k. При сохранении всех остальных переменных $\mathcal{G}^{k+1}$ может быть получено путем решения следующей оптимизационной задачи:

$$\mathcal{G}^{k+1} = \arg\min_{\mathcal{Y}} \quad \frac{\mu}{\rho^k}\|\mathcal{Y}\|^p_{\Phi,w,S_p} + \frac{1}{2}\left\|\mathcal{G}^{k+1} + \frac{\mathcal{C}^k}{\rho^k} - \mathcal{Y}\right\|_F^2. \tag{2.31}$$

Чтобы решить ее, сначала введем теорему 1, которая доказывается в Приложении.

Т е о р е м а 1. Пусть $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $l = \min\{n_1, n_2\}$, $w_1 \geq w_2 \geq \cdots \geq w_l \geq 0$. Для модели

$$\min \tau \|\mathcal{X}\|^p_{\Phi,w,S_p} + \frac{1}{2}\|\mathcal{X} - \mathcal{A}\|^2_F \tag{2.32}$$

оптимальным решением является

$$\mathcal{X}^* = \left(\text{fold}\left(\text{unfold}\left(\mathcal{S}_{\tau,w,p}(\mathcal{A}_\Phi)\right); \text{unfold}\left(\mathcal{A}_{\Phi^c}\right)\right)\right)_{\overline{\Phi}^\top}, \tag{2.33}$$

где $\mathcal{S}_{\tau,w,p}(\mathcal{A}_\Phi)$ – тензор, i-й фронтальный срез которого $S_{\tau,w,p}\left(A^{(i)}_\Phi\right)$.

Поэтому оптимальным решением (2.31) будет

$$\mathcal{Y}^{k+1} = \left(\text{fold}\left(\text{unfold}\left(\mathcal{S}_{\frac{\mu}{\rho},w,p}(\mathcal{B}^k_\Phi)\right); \text{unfold}\left(\mathcal{B}^k_{\Phi^c}\right)\right)\right)_{\overline{\Phi}^\top}, \tag{2.34}$$

где $\mathcal{B}^k = \mathcal{G}^{k+1} + \frac{\mathcal{C}^k}{\rho^k}$, $\mathcal{S}_{\frac{\mu}{\rho},w,p}(\mathcal{B}^k_\Phi)$, как показано в доказательстве теоремы 1 в Приложении. На основе вышеприведенного описания псевдокод сведен в алгоритм 2 решения задачи (2.12).

А л г о р и т м 2.

Вход: $\{X_i\}_{i=1}^m$, $\{P_i\}_{i=1}^m$, w, $\mathcal{M}$, λ, μ, γ.

Выход: A, $\{W_i\}_{i=1}^m$.

Ш а г 1. Инициализация: Номер шага $k = 0$, $\eta = 1.1$, $\rho^0 = 10^{-4}$, $\left\{\delta_i^0 = \frac{1}{m}\right\}_{i=1}^m$, $\mathcal{C}^0 = 0$, $\mathcal{G}^0 = \mathcal{Y}^0 = \mathcal{M}$.

Ш а г 2. Расчитать A^{k+1}, согласно (2.18).

Ш а г 3. Расчитать W_i^{k+1} как $\left(X_i P_i X_i^\top\right)^{-1} X_i P_i A^{k+1\top}$ для всех i.

Ш а г 4. Расчитать δ_i^{k+1}, согласно определению для всех i.

Ш а г 5. Расчитать $\mathcal{G}^{k+1}$ алгоритмом 1.

Ш а г 6. Расчитать $\mathcal{Y}^{k+1}$, согласно (2.34).

Ш а г 7. Расчитать $\mathcal{C}^{k+1} = \mathcal{C}^k + \rho^k\left(\mathcal{G}^{k+1} - \mathcal{Y}^{k+1}\right)$.

Ш а г 8. Расчитать $\rho^{k+1} = \eta\rho^k$.

Ш а г 9. Если алгоритм сошелся, то выдать $A = A^{k+1}$, $W_i = W_i^{k+1}$ для всех i.

Ш а г 10. $k \leftarrow k + 1$ и перейти к шагу 2.

2.2. А н а л и з с х о д и м о с т и. Проанализируем сходимость алгоритма 2. Сначала покажем ограниченность последовательностей в лемме 1, которая доказана в Приложении. Затем сходимость будет исследована в теореме 2, которая также доказана в Приложении.

Л е м м а 1. Последовательности $\left\{\mathcal{C}^k\right\}$, $\left\{\mathcal{Y}^k\right\}$, $\left\{A^k\right\}$, $\left\{\left\{W_i^k\right\}_{i=1}^m\right\}$, $\left\{\mathcal{G}^k\right\}$ и расширенная функция Лагранжа (2.16), которые генерируются алгоритмом 2, ограничены.

Т е о р е м а 2. Последовательности $\left\{\mathcal{G}^k\right\}$ и $\left\{\mathcal{Y}^k\right\}$, порожденные алгоритмом 2, удовлетворяют следующим требованиям:

1) $\lim\limits_{k\to\infty}\left\|\mathcal{G}^k - \mathcal{Y}^k\right\|_F = 0$,

2) $\lim\limits_{k\to\infty}\left\|\mathcal{G}^{k+1} - \mathcal{G}^k\right\|_F = 0$,

3) $\lim\limits_{k\to\infty}\left\|\mathcal{Y}^{k+1} - \mathcal{Y}^k\right\|_F = 0$.

Таблица 2. Описание многоаспектных баз данных

	Количество			
База данных	классов	аспектов	образцов	признаков в аспектах
Цифры	10	5	2000	240/76/216/47/64
3_Sources	6	3	169	3560/3631/3036
BBC Sport	5	4	116	1991/2063/2113/2158
Orl	40	4	400	41/399/399/41

3. Вычислительные эксперименты. Наборы данных. Четыре набора данных, использованные в эксперименте, показаны в табл. 2 и описаны ниже.

1. *Цифры* [50]. Набор данных содержит 10 классов объектов, т.е. изображения рукописных цифр 0–9. В каждом классе есть 200 образцов. Известны шесть аспектов: средние значения пикселей, коэффициенты Фурье, корреляции профилей, момент Цернике, коэффициенты Карунена–Лоэва и морфологические. В данном эксперименте задействованы только первые пять.

2. *3_Sources* [51]. Набор данных содержит 948 историй из трех онлайн-ресурсов – BBC, Reuters и Guardian. В данном эксперименте для сравнения с другими методами используется подмножество, содержащее 169 историй, о которых сообщалось на всех ресурсах. Эти 169 историй разделены на шесть классов: бизнес, развлечения, здоровье, политика, спорт и технологии.

3. *BBC Sport* [52]. Набор данных содержит 737 документов о новостях спорта, которые выражены 2–4 мнениями и разделены на пять категорий. Для наших экспериментов мы выбрали подмножество образцов, описанных по 4 раза, в нем 116 элементов.

4. *Orl* [53]. Очень популярный набор данных лиц, содержащий 400 изображений лиц, предоставленных 40 лицами, каждое из которых состоит из 10 фотографий. Поскольку Orl – это набор данных с одним аспектом, мы извлекли LBP, GIST и пирамиду гистограммы ориентированных градиентов и объединили их с пиксельными характеристиками, чтобы сформировать набор данных с четырьмя аспектами.

Построение неполных многоаспектных данных. Для каждого из вышеперечисленных наборов данных случайным образом выбираем $p\%$ ($p \in \{90, 70, 50, 30\}$) образцов и превращаем их в неполные путем удаления данных из случайно выбранного аспекта.

Сравниваемые методы. Следующие методы, которые могут работать с неполными многоаспектными данными, использованы для сравнения.

1. BSV [54]. BSV выполняет кластеризацию k-средних и классификацию ближайших соседей на всех аспектах и сообщает оптимальные результаты кластеризации и классификации, где недостающие экземпляры дополняются средним экземпляром соответствующего аспекта.

2. Concat [54]. Объединение всех аспектов в один и применение k-средних и ближайших соседей для получения окончательных результатов кластеризации и классификации, где недостающие экземпляры также заполняются средним экземпляром, как BSV.

3. DAIMC [20]. Метод использует два способа для согласованного представления всех аспектов: матричное разложение на основе выравнивания информации об экземплярах и разреженную регрессию на основе базисных матриц.

4. IMSC_AGL [4]. Метод получает консенсусное низкоразмерное представление при помощи спектральной кластеризации и строит графы низкорангового представления.

5. TCIMC [10]. Метод получает полный граф путем тензорного восстановления для каждого вида и отличается от представленного здесь метода тензорной нормой.

Наш метод сначала получает низкоразмерное представление A всех образцов, а затем применяет метод k-средних для A. Поскольку метод чувствителен к начальному значению, вычисления производятся несколько раз и результат усредняется. Найденные центры служат для построения графа каждого представления:

$$M_{i(k,l)} = \begin{cases} \exp\left(-\dfrac{\left\|X_{i(k,:)} - X_{i(l,:)}\right\|_2^2}{2\sigma^2}\right), & \text{если} \quad X_{i(k,:)} \in k\text{-NN}\left(X_{i(l,:)}\right), \\ 0 \quad \text{иначе}, \end{cases} \tag{3.1}$$

где $k = 10$, $\sigma = 1$. Недостающие позиции в M_i заполняются нулями. Значения p, w и Φ в тензорной норме $\|\cdot\|_{\Phi,w,S_p}$ задаются 0.6, **1** и дискретным косинусным преобразованием соответственно.

Меры качества. Оценка качества производится тремя известными мерами: точностью (accuracy, ACC), нормализованной взаимной информацией (normalized mutual information, NMI), чистотой (purity).

3.1. Результаты. Таблица 3 показывает результаты кластеризации шести методов на четырех наборах данных. Из нее можно сделать следующие выводы.

1. Представленный метод имеет лучшие результаты в большинстве случаев по сравнению с другими пятью методами. В задаче кластеризации предложенный метод примерно на 5 и 10% лучше, чем другие, на данных Orl и Handwritten, соответственно. Неизбежны некоторые данные, для которых та или иная мера не является оптимальной, например задача кластеризации наборов данных Handwritten, 3Sources и BBC Sport.

2. Таблица 3 демонстрирует, что IMSC_AGL, TCIMC и наш метод имеют лучшее качество кластеризации, чем DAIMC на всех наборах данных. Это указывает на то, что графовые инструменты могут улучшить качество кластеризации. TCIMC и наш метод основаны на тензорном заполнении графа, в них отличается тензорная норма. Эксперименты показывают, что наш метод имеет лучшую производительность кластеризации, чем TCIMC, на большинстве наборов данных.

3. Из результатов эксперимента, приведенных в табл. 3, видно, что качество кластеризации снижается на большинстве наборов данных при увеличении доли пропущенных аспектов. Таким образом, обучение эффективному согласованному представлению на наборе данных с высоким уровнем пропусков затруднено. Отсутствующих экземпляров так много, что предварительной информации недостаточно, включая консенсус и взаимодополняемость между представлениями. Согласно результатам экспериментов на всех наборах данных, наш метод имеет более значительное преимущество в случае более высокого уровня пропусков.

3.2. Анализ чувствительности к параметрам. В этом разделе анализируется чувствительность трех параметров λ, μ и γ на наборах данных Handwritten, 3Sources, BBC Sport и Orl. Проверяются параметры λ со значениями {1, 4, 5, 7, 9} и μ со значениями {10, 30, 50, 70, 90}. Предложенный метод кластеризации выполняется с различными комбинациями двух параметров (λ & μ). Также выполняется кластеризация с различными γ из множества $\{10^p,\ p \in \overline{-2,6}\}$. Рисунки 2 и 3 показывают качество кластеризации в зависимости от параметра масштаба λ & μ и γ на четырех наборах данных с уровнем пропусков 70% соответственно. Согласно этим результатам, наш метод относительно стабилен для гиперпараметров λ, μ и γ, а производительность кластеризации обычно оптимальна для $\lambda = 3$, $\mu = 10$ и $\gamma = 100$.

3.3. Анализ сходимости. В разд. 2.3 показано, что алгоритм 2 сходится. Описаны численные эксперименты для проверки сходимости. Проведены эксперименты на наборах данных Handwritten, 3 Sources, BBC Sport и Orl с уровнем пропусков 70%. На рис. 4 показана сходимость алгоритма и качество кластеризации модели при увеличении числа итераций. Из рис. 4 видно, что предложенный метод оптимизации имеет хорошую сходимость, при которой $\left\|\mathcal{G}^{k+1} - \mathcal{G}^k\right\|_F$, $\left\|\mathcal{Y}^{k+1} - \mathcal{Y}^k\right\|_F$, и $\left\|\mathcal{G}^k - \mathcal{Y}^k\right\|_F$ может быстро уменьшаться и стремиться к нулю во всех четырех наборах данных. Результаты кластеризации также стабилизируются с увеличением числа итераций.

3.4. Влияние параметра r. Исследуем влияние параметра r на результат эксперимента. Вводим r от $n \times 10$ до $n \times 100\%$ для каждого эксперимента на наборах данных Handwritten, 3 Sources, BBC Sport и Orl с уровнем пропусков 70%, т.е. $r \in \{n \times 10\%, n \times 20\%, \ldots, n \times 100\%\}$.

Таблица 3. Средние и стандартные значения ACC, NMI и Purity (в процентах) семи методов для четырех наборов данных

Dataset	Rate/methods		BSV	Concat	DAIMC	IMSC_AGL	TCIMV	EIMC_TGC
Hand-written	ACC	0.9	47.15B±2.86	39.96B±3.37	84.51B±0.00	94.14B±0.00	86.79B±0.98	**99.60B±0.00**
		0.7	52.04B±24.95	44.37B±10.75	85.21B±0.06	94.32B±0.00	86.35B±1.26	**99.95B±0.00**
		0.5	62.51B±19.99	48.95B±2.88	87.13B±0.04	**95.40B±0.00**	85.90B±0.36	91.50B±0.00
		0.3	67.47B±35.51	53.12B±5.01	86.61B±0.05	**96.04B±0.00**	85.15B±1.01	88.65B±0.00
	NMI	0.9	38.79B±1.09	32.91B±1.51	73.23B±0.01	88.07B±0.00	88.95B±1.34	**98.94B±0.00**
		0.7	45.34B±9.94	37.91B±4.83	76.64B±0.01	88.12B±0.01	89.77B±0.54	**99.86B±0.00**
		0.5	54.09B±8.35	43.27B±1.73	77.17B±0.03	90.01B±0.00	89.81B±2.36	**93.97B±0.00**
		0.3	60.78B±11.11	49.02B±1.12	77.24B±0.02	91.34B±0.00	89.58B±0.95	**93.57B±0.00**
	Purity	0.9	47.15B±2.83	40.65B±1.71	84.51B±0.01	94.14B±0.00	88.01B±1.32	**99.60B±0.00**
		0.7	52.77B±17.79	46.22B±4.08	86.75B±0.01	94.34B±0.00	88.40B±2.01	**99.95B±0.00**
		0.5	62.51B±17.69	51.27B±1.21	87.18B±0.02	**95.40B±0.00**	88.40B±0.66	91.50B±0.00
		0.3	68.80B±23.08	57.06B±2.66	86.73B±0.23	**96.04B±0.00**	88.10B±1.32	89.40B±0.00
3_Sources	ACC	0.9	35.50B±3.11	37.33B±6.96	46.86B±0.00	60.95B±1.55	68.64B±0.00	**74.41B±0.16**
		0.7	35.74B±12.46	38.04B±30.50	53.37B±1.93	57.45B±1.20	69.23B±0.00	**78.70B±0.14**
		0.5	36.09B±8.01	40.35B±47.45	50.29B±0.00	**65.62B±5.94**	56.80B±0.00	59.76B±0.06
		0.3	35.62B±15.77	46.39B±35.41	49.76B±0.04	68.11B±2.83	**69.82B±0.00**	59.16B±0.08
	NMI	0.9	8.53B±11.81	7.96B±14.37	44.29B±0.00	59.17B±0.19	63.85B±0.00	**63.97B±0.22**
		0.7	9.29B±12.87	11.36B±69.01	46.03B±3.02	60.55B±0.10	54.34B±0.02	**66.29B±0.11**
		0.5	8.56B±11.76	13.87B±75.81	45.38B±0.16	**57.89B±6.65**	52.50B±0.00	52.47B±1.14
		0.3	7.92B±6.29	22.97B±98.79	47.53B±0.35	**61.18B±0.86**	57.98B±0.00	60.20B±0.20
	Purity	0.9	38.99B±5.24	39.40B±18.61	62.13B±0.00	77.75B±0.09	79.23B±0.00	**79.28B±0.09**
		0.7	39.88B±12.93	42.48B±61.37	68.52B±0.68	77.28B±0.09	75.14B±0.00	**82.25B±0.14**
		0.5	39.64B±16.10	45.32B±74.47	68.28B±0.10	**76.56B±6.08**	72.19B±0.00	70.41B±0.06
		0.3	39.17B±12.20	52.60B±61.97	67.57B±0.06	76.69B±1.26	**78.10B±0.00**	74.56B±0.00
BBC Sport	ACC	0.9	33.88B±4.46	33.36B±4.79	60.69B±0.19	66.21B±0.13	81.03B±0.00	**83.62B±0.00**
		0.7	34.13B±7.63	34.91B±21.35	66.38B±1.98	78.45B±0.00	81.62B±0.00	**81.90B±0.00**
		0.5	35.26B±4.86	33.62B±3.30	61.81B±1.33	69.66B±0.79	79.31B±0.00	**82.76B±0.00**
		0.3	34.65B±14.00	37.58B±61.13	62.16B±5.85	75.00B±0.00	**81.89B±0.00**	80.17B±0.00
	NMI	0.9	5.82B±3.72	5.81B±2.24	46.45B±0.95	51.17B±0.83	78.20B±0.00	**83.03B±0.00**
		0.7	6.23B±5.19	6.27B±7.76	55.52B±0.55	69.55B±0.00	**83.58B±0.00**	79.91B±0.00
		0.5	6.11B±10.06	5.37B±5.26	45.28B±2.29	57.68B±0.31	71.07B±0.00	**77.31B±0.00**
		0.3	6.54B±19.41	11.01B±63.76	37.05B±19.07	68.60B±0.40	**76.19B±0.00**	74.50B±0.00
	Purity	0.9	35.00B±3.34	34.65B±2.61	68.10B±0.83	76.55B±0.13	91.03B±0.00	**93.11B±0.00**
		0.7	35.17B±6.73	36.55B±19.19	75.26B±1.49	87.93B±0.00	93.82B±0.00	**93.96B±0.00**
		0.5	36.21B±5.11	35.29B±4.03	68.62B±1.51	82.58B±0.79	88.79B±0.00	**92.24B±0.00**
		0.3	35.60B±13.71	39.14B±55.19	65.00B±7.46	87.84B±0.07	90.21B±0.00	**90.52B±0.00**
Orl	ACC	0.9	43.10B±2.98	16.08B±0.88	56.18B±6.47	71.20B±4.95	86.50B±1.06	**92.10B±2.04**
		0.7	40.70B±1.22	15.22B±2.23	58.71B±4.36	75.58B±4.16	84.95B±0.85	**88.72B±4.83**

Таблица 3. Окончание

Dataset	Rate/methods		BSV	Concat	DAIMC	IMSC_AGL	TCIMV	EIMC_TGC
	NMI	0.5	42.27B±1.92	16.87B±2.88	61.08B±1.89	71.45B±2.02	85.45B±0.54	**87.87B±2.37**
		0.3	42.40B±1.60	21.50B±135.6	65.05B±5.19	73.85B±4.44	84.00B±2.04	**89.45B±1.76**
		0.9	47.81B±0.20	24.28B±2.91	70.98B±1.91	83.22B±1.39	91.06B±0.09	**95.02B±0.24**
		0.7	47.93B±0.80	24.72B±5.56	73.70B±2.07	86.79B±0.36	92.94B±0.07	**95.25B±0.36**
		0.5	48.58B±0.79	29.04B±9.52	76.18B±1.89	84.57B±0.57	92.27B±0.17	**94.98B±0.16**
		0.3	48.62B±0.13	35.14B±178.1	79.90B±2.60	86.10B±0.57	91.24B±0.41	**96.23B±0.28**
	Purity	0.9	46.18B±1.20	19.10B±0.57	58.82B±4.76	74.13B±3.80	87.40B±0.09	**92.65B±1.07**
		0.7	44.47B±0.42	18.25B±2.09	61.05B±4.71	77.90B±1.86	88.35B±0.03	**90.63B±1.88**
		0.5	45.92B±1.04	19.82B±2.97	63.70B±3.15	74.12B±2.18	87.35B±0.17	**90.07B±0.75**
		0.3	45.72B±0.17	24.90B±146.8	68.15B±4.96	76.75B±2.19	86.60B±0.41	**92.30B±1.27**

Производительность кластеризации нашего метода на наборах данных Handwritten и Orl уменьшается при увеличении r/n (рис. 5). На наборах данных 3 Sources и BBC Sport производительность предложенной модели существенно не меняется при увеличении r/n. Рисунок 6 показывает, что время работы для решения $\mathcal{Y}$ становится больше с увеличением значения r/n на четырех наборах данных. Отметим, что $r = n$ не является оптимальным выбором, в данном методе имеет смысл использовать $r < n$.

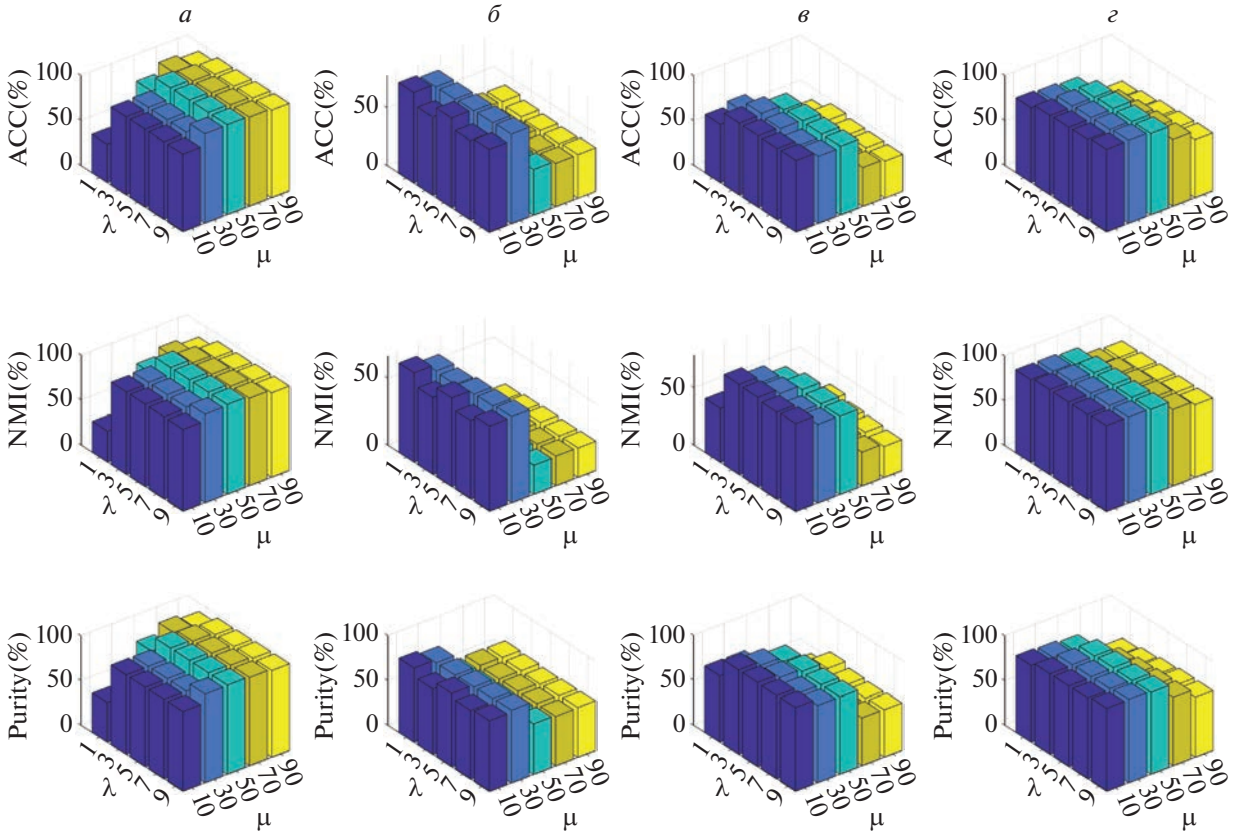

Рис. 2. Качество кластеризации при различных комбинациях параметров λ и μ на наборах данных Handwritten (*а*), 3 Sources (*б*), BBC Sport (*в*) и Orl (*г*) с уровнем пропусков 70%

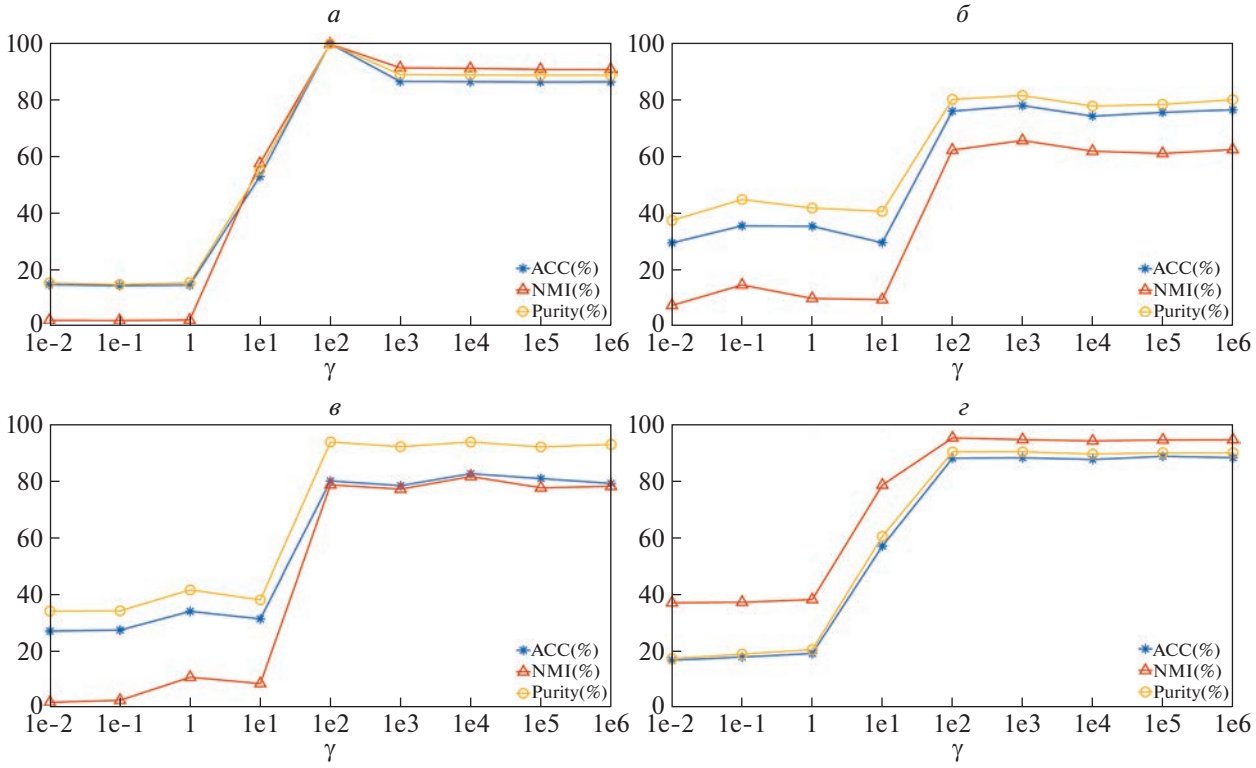

Рис. 3. Качество кластеризации при различных значениях параметра γ на наборах данных Handwritten (*а*), 3 Sources (*б*), BBC Sport (*в*) и Orl (*г*) с уровнем пропусков 70%

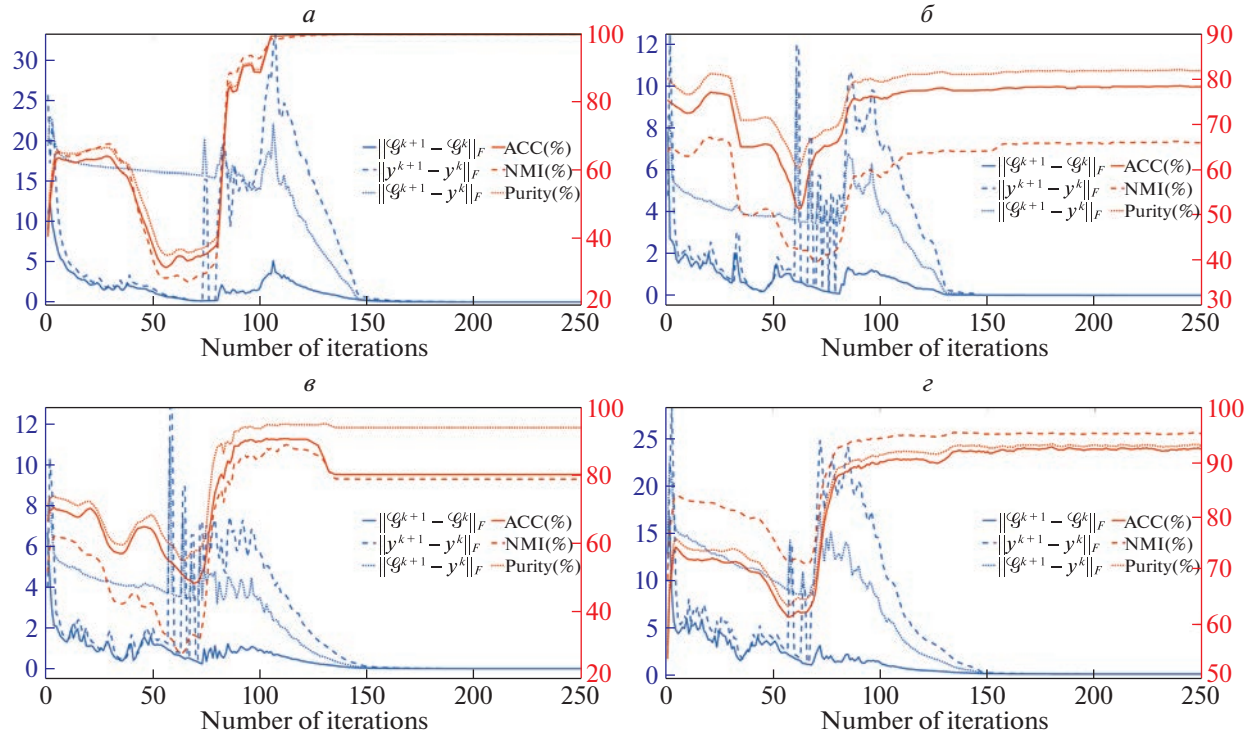

Рис. 4. Зависимость качества кластеризации от количества итераций на наборах данных Handwritten (*а*), 3 Sources (*б*), BBC Sport (*в*) и Orl (*г*) с уровнем пропусков 70%

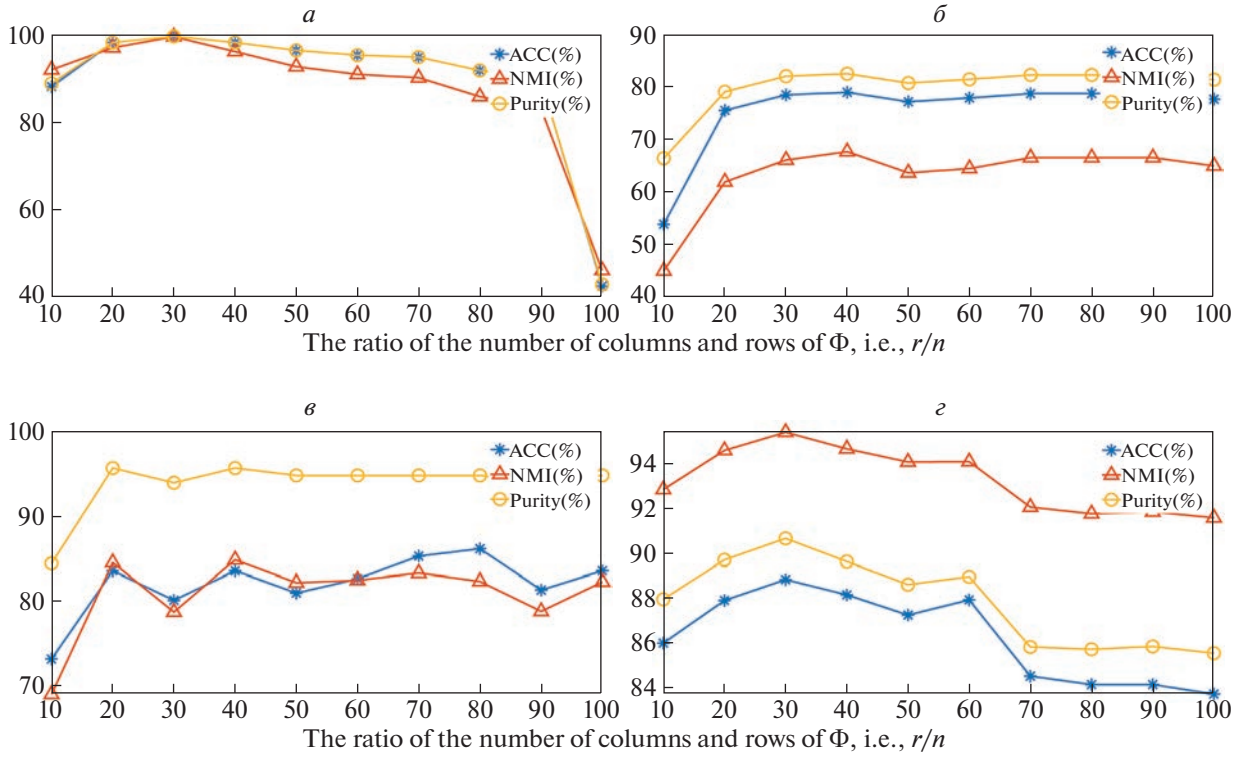

Рис. 5. Качество кластеризации для различных пропорций параметра r к n на наборах данных Handwritten (*а*), 3 Sources (*б*), BBC Sport (*в*) и Orl (*г*) с уровнем пропусков 70%

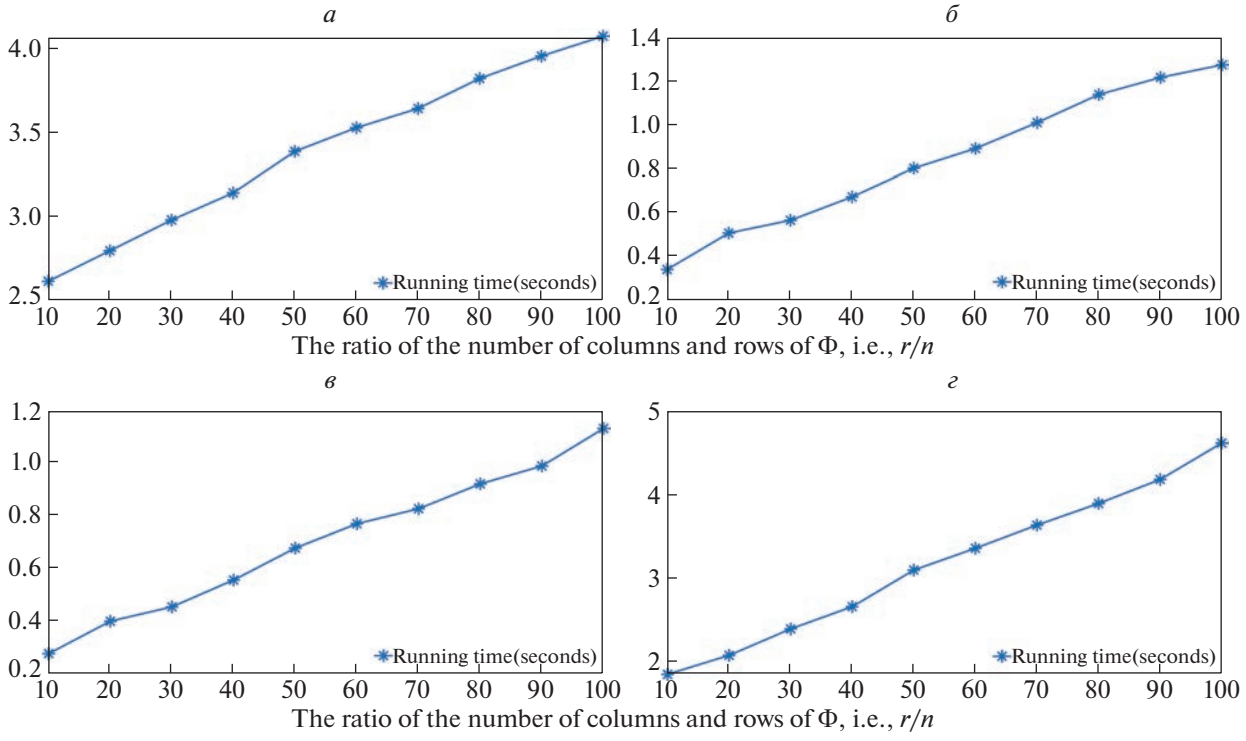

Рис. 6. Время работы для различных отношений параметра r к n на наборах данных Handwritten (*а*), 3 Sources (*б*), BBC Sport (*в*) и Orl (*г*) с уровнем пропусков 70%

Заключение. Представлен метод кластеризации неполных данных с несколькими аспектами, который применяет графы связности, дополненные путем тензорного восстановления. Это позволяет найти согласованное низкоразмерное подпространство, сохраняющее топологию (взаимное расположение) образцов. Построенная модель решается на основе метода неопределенных множителей Лагранжа, анализируется сходимость. Качество работы предложенного метода проверено на реальных наборах данных.

ПРИЛОЖЕНИЕ

Доказательство теорем 1, 2 и леммы 1. Перед доказательством теоремы 1, приведем леммы 2 и 3, представленные в [55].

Лемма 2. Пусть для оптимизационной задачи

$$\min_{x>0} \frac{1}{2}(x-\sigma)^2 + wx^p \tag{П.1}$$

при заданных p и w определен порог

$$\tau_p(w) = (2w(1-p))^{\frac{1}{2-p}} + wp(2w(1-p))^{\frac{p-1}{2-p}}. \tag{П.2}$$

Можно заключить, что:

1) если $\sigma \le \tau_p(w)$, то оптимальное решение $x_p(\sigma, w)$ из (П.1) равно нулю;

2) если $\sigma > \tau_p(w)$, то оптимальным решением является $x_p(\sigma, w) = \operatorname{sign}(\sigma) S_p(\sigma, w)$, где $S_p(\sigma, w)$ можно получить из $S_p(\sigma, w) - \sigma + wp(S_p(\sigma, w))^{p-1} = 0$.

Лемма 3. Пусть $Y = U_Y \Sigma_Y V_Y^\top$ – сингулярное разложение $Y \in \mathbb{R}^{m\times n}$, $\tau > 0$, $l = \min(m, n)$, $w_1 \ge w_2 \ge \cdots \ge w_l \ge 0$. Тогда глобальным оптимумом задачи

$$\min_X \quad \tau\|X\|_{w,S_p}^p + \frac{1}{2}\|X-Y\|_F^2 \tag{П.3}$$

будет

$$S_{\tau,w,p}(Y) = U_Y P_{\tau,w,p}(Y) V_Y^\top, \tag{П.4}$$

где $P_{\tau,w,p}(Y) = \operatorname{diag}(\gamma_1, \gamma_2, \ldots, \gamma_l)$ и $\gamma_i = x_p(\sigma_i(Y), \tau w_i)$ могут быть получены, согласно лемме 2.

Доказательство теоремы 1. Для заданной полуортогональной матрицы $\Phi \in \mathbb{R}^{n_3\times r}$ существует полуортогональная матрица $\Phi^c \in \mathbb{R}^{n_3\times(n_3-r)}$, удовлетворяющая $\overline{\Phi}^\top \overline{\Phi} = I$, где $\overline{\Phi} = \left[\Phi, \Phi^c\right] \in \mathbb{R}^{n_3\times n_3}$. Согласно определению, имеем

$$\begin{aligned} \mathscr{X}^* &= \arg\min_{\mathscr{X}} \tau\|\mathscr{X}\|_{\Phi,w,S_p}^p + \frac{1}{2}\|\mathscr{X}-\mathscr{A}\|_F^2 = \arg\min_{\mathscr{X}} \tau\sum_{i=1}^r \left\|X_\Phi^{(i)}\right\|_{w,S_p}^p + \frac{1}{2}\|\mathscr{X}_{\overline{\Phi}} - \mathscr{A}_{\overline{\Phi}}\|_F^2 = \\ &= \arg\min_{\mathscr{X}} \tau\sum_{i=1}^r \left\|X_\Phi^{(i)}\right\|_{w,S_p}^p + \frac{1}{2}\sum_{i=1}^r \left\|X_\Phi^{(i)} - A_\Phi^{(i)}\right\|_F^2 + \frac{1}{2}\sum_{j=1}^{n_3-r} \left\|X_{\Phi^c}^{(j)} - A_{\Phi^c}^{(j)}\right\|_F^2. \end{aligned} \tag{П.5}$$

В (П.5) переменные $X_\Phi^{(i)}$ и $X_{\Phi^c}^{(j)}$ независимы. Таким образом, задача может быть разделена на n_3 независимых подзадач:

$$\min_{X_\Phi^{(i)}} \tau\left\|X_\Phi^{(i)}\right\|_{w,S_p}^p + \frac{1}{2}\left\|X_\Phi^{(i)} - A_\Phi^{(i)}\right\|_F^2, \quad i = \overline{1,r}, \tag{П.6}$$

и

$$\min_{X_{\Phi^c}^{(j)}} \quad \frac{1}{2}\left\|X_{\Phi^c}^{(j)} - A_{\Phi^c}^{(j)}\right\|_F^2, \quad j = \overline{1, n_3 - r}. \tag{П.7}$$

Согласно лемме 3, оптимальными решениями (П.6) и (П.7) являются $X_{\Phi}^{(i)*} = U_{A_{\Phi}^{(i)}} P_{\tau,w,p}\left(A_{\Phi}^{(i)}\right) V_{A_{\Phi}^{(i)}}^{\top}$ и $X_{\Phi^c}^{(j)*} = A_{\Phi^c}^{(j)}$ соответственно. Таким образом, получаем оптимальные решения

$$\begin{aligned}\mathcal{X}_{\overline{\Phi}}^{*} &= \operatorname{fold}\left(X_{\Phi}^{(1)*}; X_{\Phi}^{(2)*}; \cdots; X_{\Phi}^{(r)*}; X_{\Phi^c}^{(1)*}; X_{\Phi^c}^{(2)*}; \cdots; X_{\Phi^c}^{(n_3-r)*}\right) = \\ &= \operatorname{fold}\left(S_{\tau,w,p}\left(A_{\Phi}^{(1)}\right); \cdots; S_{\tau,w,p}\left(A_{\Phi}^{(r)}\right); A_{\Phi^c}^{(1)}; \cdots; A_{\Phi^c}^{(n_3-r)}\right) = \\ &= \operatorname{fold}\left(\operatorname{unfold}\left(\mathcal{S}_{\tau,w,p}\left(\mathcal{A}_{\Phi}\right)\right); \operatorname{unfold}\left(\mathcal{A}_{\Phi^c}\right)\right).\end{aligned} \tag{П.8}$$

Далее оптимальным решением (2.32) станет $\mathcal{X}^* = \left(\mathcal{X}_{\overline{\Phi}}^{*}\right)_{\overline{\Phi}^{\top}}$.

Доказательство леммы 1. Из ограничения на $\mathcal{G}$ в (2.12) ясно, что $\left\{\mathcal{G}^k\right\}$ должно быть ограничено. Далее, согласно (2.18), известно, что $\{A^k\}$ и $\left\{\left\{W_i^k\right\}_{i=1}^m\right\}$ также ограничены. На k-й итерации в алгоритме 2 имеем

$$\begin{aligned}\left\|\mathcal{C}^{k+1}\right\|_F &= \left\|\mathcal{C}^k + \rho^k\left(\mathcal{G}^{k+1} - \mathcal{Y}^{k+1}\right)\right\|_F = \rho^k\left\|\mathcal{G}^{k+1} + \frac{\mathcal{C}^k}{\rho^k} - \mathcal{Y}^{k+1}\right\|_F = \\ &= \rho^k\left\|\left(\mathcal{G}^{k+1} + \frac{\mathcal{C}^k}{\rho^k} - \mathcal{Y}^{k+1}\right)_{\overline{\Phi}}\right\|_F = \rho^k\left\|\mathcal{B}_{\overline{\Phi}}^k - \mathcal{Y}_{\overline{\Phi}}^{k+1}\right\|_F.\end{aligned} \tag{П.9}$$

Из (2.34) получаем

$$\mathcal{Y}_{\overline{\Phi}}^{k+1} = \operatorname{fold}\left(\operatorname{unfold}\left(\mathcal{S}_{\frac{\mu}{\rho^k},w,p}\left(\mathcal{B}_{\Phi}^k\right)\right); \operatorname{unfold}\left(\mathcal{B}_{\Phi^c}^k\right)\right). \tag{П.10}$$

Подставляя (П.10) в (П.9), имеем

$$\begin{aligned}\left\|\mathcal{C}^{k+1}\right\|_F &= \rho^k\left\|\mathcal{B}_{\overline{\Phi}}^k - \operatorname{fold}\left(\operatorname{unfold}\left(\mathcal{S}_{\frac{\mu}{\rho^k},w,p}\left(\mathcal{B}_{\Phi}^k\right)\right); \operatorname{unfold}\left(\mathcal{B}_{\Phi^c}^k\right)\right)\right\|_F = \\ = \rho^k\left\|\mathcal{B}_{\Phi}^k - \mathcal{S}_{\frac{\mu}{\rho^k},w,p}\left(\mathcal{B}_{\Phi}^k\right)\right\|_F &\le \rho^k\sqrt{\sum_{i=1}^{r}\sum_{j=1}^{\min(n_1,n_2)}\left(\frac{Jw_j\mu}{\rho^k}\right)^2} = \sqrt{\sum_{i=1}^{r}\sum_{j=1}^{\min(n_1,n_2)}\left(Jw_j\mu\right)^2},\end{aligned} \tag{П.11}$$

где J – число итераций при решении задачи (П.1). Таким образом, $\left\{\mathcal{C}^k\right\}$ ограничено. Далее, согласно (2.34), видим, что

$$\begin{aligned}\left\|\mathcal{Y}^{k+1}\right\|_F &= \left\|\operatorname{fold}\left(\operatorname{unfold}\left(\mathcal{S}_{\frac{\mu}{\rho},w,p}\left(\mathcal{B}_{\Phi}^k\right)\right); \operatorname{unfold}\left(\mathcal{B}_{\Phi^c}^k\right)\right)\right\|_F \le \\ \le \left\|\mathcal{S}_{\frac{\mu}{\rho},w,p}\left(\mathcal{B}_{\Phi}^k\right)\right\|_F + \left\|\mathcal{B}_{\Phi^c}^k\right\|_F &\le \sum_{i=1}^{r}\left\|P_{\tau,w,p}\left(\mathcal{B}_{\Phi}^{k\,(i)}\right)\right\|_F + \left\|\mathcal{B}_{\Phi^c}^k\right\|_F.\end{aligned} \tag{П.12}$$

Уже упоминалось, что $\left\{\mathcal{C}^k\right\}$ и $\left\{\mathcal{G}^k\right\}$ ограничены, поэтому $\left\{\mathcal{B}^k\right\}$ ограничено, а значит, $\left\{\mathcal{Y}^k\right\}$ также ограничено.

Ниже доказывается, что расширенная функция Лагранжа (2.16) ограничена. Заметим, что нельзя утверждать, что (2.16) ограничена на основании того, что все переменные ограничены, так как $\{\rho^k\}$ не ограничена. Сначала докажем, что

$$L_{\rho^k}\left(A^{k+1}, \left\{W_i^{k+1}\right\}_{i=1}^m, \mathcal{G}^k, \mathcal{Y}^k, \mathcal{C}^k\right) \le L_{\rho^k}\left(A^k, \left\{W_i^k\right\}_{i=1}^m, \mathcal{G}^k, \mathcal{Y}^k, \mathcal{C}^k\right). \tag{П.13}$$

Вспомним (2.17), пусть $f(A,W_i)=\left\|\left(A-W_i^{\top}X_i\right)P_i\right\|_F^2+\lambda\mathrm{tr}\left(AL_{G_i^k}A^{\top}\right)$ и, сочетая с $\delta_i^k=n_i\Big/\sqrt{f(A^k,W_i^k)}$, можно вывести

$$\sum_{i=1}^{m}\frac{n_i f\left(A^{k+1},W_i^{k+1}\right)}{2\sqrt{f\left(A^k,W_i^k\right)}}\le\sum_{i=1}^{m}\frac{n_i f\left(A^k,W_i^k\right)}{2\sqrt{f\left(A^k,W_i^k\right)}}. \tag{П.14}$$

Для любых положительных чисел a и b выполняется следующее неравенство:

$$a-\frac{a^2}{2b}\le b-\frac{b^2}{2b}. \tag{П.15}$$

А так как $f(A,W_i)\ge 0$, то имеем

$$\sum_{i=1}^{m}n_i\left(\sqrt{f\left(A^{k+1},W_i^{k+1}\right)}-\frac{f\left(A^{k+1},W_i^{k+1}\right)}{2\sqrt{f\left(A^k,W_i^k\right)}}\right)\le\sum_{i=1}^{m}n_i\left(\sqrt{f\left(A^k,W_i^k\right)}-\frac{f\left(A^k,W_i^k\right)}{2\sqrt{f\left(A^k,W_i^k\right)}}\right). \tag{П.16}$$

Комбинируя (П.14) и (П.16), получаем

$$\sum_{i=1}^{m}n_i\sqrt{f\left(A^{k+1},W_i^{k+1}\right)}\le\sum_{i=1}^{m}n_i\sqrt{f\left(A^k,W_i^k\right)}. \tag{П.17}$$

Таким образом, выводится уравнение (П.13). В разд. 2.2.2 и 2.2.3 так как подзадачи $\mathcal{G}$ и $\mathcal{Y}$ имеют оптимальное решение, то очевидно, что

$$\begin{aligned}L_{\rho^k}\left(A^{k+1},\left\{W_i^{k+1}\right\}_{i=1}^m,\mathcal{G}^{k+1},\mathcal{Y}^{k+1},\mathcal{C}^k\right)\le L_{\rho^k}\left(A^{k+1},\left\{W_i^{k+1}\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^k\right)\le\\ \le L_{\rho^k}\left(A^k,\left\{W_i^k\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^k\right).\end{aligned} \tag{П.18}$$

Тогда

$$\begin{aligned}L_{\rho^k}\left(A^k,\left\{W_i^k\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^k\right)&=L_{\rho^{k-1}}\left(A^k,\left\{W_i^k\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^{k-1}\right)+\frac{\rho^k-\rho^{k-1}}{2}\left\|\mathcal{G}^k-\mathcal{Y}^k\right\|_F^2+\\&+\left\langle\mathcal{C}^k-\mathcal{C}^{k-1},\mathcal{G}^k-\mathcal{Y}^k\right\rangle=L_{\rho^{k-1}}\left(A^k,\left\{W_i^k\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^{k-1}\right)+\\&+\frac{\rho^k-\rho^{k-1}}{2}\left\|\frac{\mathcal{C}^k-\mathcal{C}^{k-1}}{\rho^{k-1}}\right\|_F^2+\left\langle\mathcal{C}^k-\mathcal{C}^{k-1},\frac{\mathcal{C}^k-\mathcal{C}^{k-1}}{\rho^{k-1}}\right\rangle=\\&=L_{\rho^{k-1}}\left(A^k,\left\{W_i^k\right\}_{i=1}^m,\mathcal{G}^k,\mathcal{Y}^k,\mathcal{C}^{k-1}\right)+\frac{\rho^k+\rho^{k-1}}{2(\rho^{k-1})^2}\left\|\mathcal{C}^k-\mathcal{C}^{k-1}\right\|_F^2.\end{aligned} \tag{П.19}$$

Пусть b_c является границей $\left\|\mathcal{C}^k-\mathcal{C}^{k-1}\right\|_F^2$. Сочетая уравнения (П.18), получаем

$$L_{\rho^k}\left(A^{k+1},\left\{W_i^{k+1}\right\}_{i=1}^m,\mathcal{G}^{k+1},\mathcal{Y}^{k+1},\mathcal{C}^k\right)\le L_{\rho^0}\left(A^1,\left\{W_i^1\right\}_{i=1}^m,\mathcal{G}^1,\mathcal{Y}^1,\mathcal{C}^0,\right)+b_c\sum_{k=1}^{\infty}\frac{\rho^k+\rho^{k-1}}{2(\rho^{k-1})^2}. \tag{П.20}$$

Так как $\rho>1$, то имеет место следующее неравенство:

$$\sum_{k=1}^{\infty}\frac{\rho^k+\rho^{k-1}}{2(\rho^{k-1})^2}\le\sum_{k=1}^{\infty}\frac{\rho^k}{(\rho^{k-1})^2}=\eta\sum_{k=1}^{\infty}\frac{1}{\rho^{k-1}}<+\infty. \tag{П.21}$$

Комбинируя (П.19)–(П.21), расширенная функция Лагранжа (2.16) ограничена.

Доказательство теоремы 2. Согласно лемме 1, знаем, что $\left\{\mathcal{C}^k\right\}$, $\left\{\mathcal{G}^k\right\}$ и $\left\{\mathcal{Y}^k\right\}$ все ограничены. Теорема Больцано–Вейерштрасса гласит, что каждая ограниченная последовательность вещественных чисел имеет сходящуюся подпоследовательность. Поэтому существует по

крайней мере одна точка накопления для $\{\mathscr{C}^k\}$, $\{\mathscr{G}^k\}$ и $\{\mathscr{Y}^k\}$ соответственно. В частности, можно получить

$$\lim_{k\to\infty}\left\|\mathscr{G}^k-\mathscr{Y}^k\right\|_F=\lim_{k\to\infty}\frac{1}{\rho^k}\left\|\mathscr{C}^k-\mathscr{C}^{k-1}\right\|_F=0. \tag{П.22}$$

Как и в разд. 2.2.2, каждая трубка из $\mathscr{G}^k$ может быть проанализирована независимо. Если вспомнить, что g^{k+1} обновляется способом (2.27), то выполняется следующее неравенство:

$$\begin{aligned}\left\|g^{k+1}-g^k\right\|_F&=\left\|\begin{array}{l}\dfrac{\rho^k}{\rho^k+\gamma}\max\left(P_w\left(u^k-\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\dfrac{1}{n}\mathbf{1}+v\mathbf{1}\right),0\right)\\ +\max\left(P_{w^c}\left(u^k-\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\dfrac{1}{n}\mathbf{1}+v\mathbf{1}\right),0\right)-g^k\end{array}\right\|_F\le\\ &\le\left\|\frac{\rho^k}{\rho^k+\gamma}\left(u^k-\frac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\frac{1}{n}\mathbf{1}+v\mathbf{1}-\frac{\rho^k+\gamma}{\rho^k}g^k\right)\right\|_F+\left\|u^k-\frac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\frac{1}{n}\mathbf{1}+v\mathbf{1}-g^k\right\|_F.\end{aligned} \tag{П.23}$$

Проанализируем первый и второй члены (П.23) отдельно:

$$\begin{aligned}&\left\|\frac{\rho^k}{\rho^k+\gamma}\left(u^k-\frac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\frac{1}{n}\mathbf{1}+v\mathbf{1}-\frac{\rho^k+\gamma}{\rho^k}g^k\right)\right\|_F=\\ &=\left\|\frac{\rho^k}{\rho^k+\gamma}\begin{pmatrix}y^k-\dfrac{c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\\ -\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top\left(y^k-\dfrac{c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\right)\\ +\dfrac{1}{n}\mathbf{1}-g^k-\dfrac{\gamma}{\rho^k}g^k+v1\end{pmatrix}\right\|_F=\\ &=\left\|\frac{\rho^k}{\rho^k+\gamma}\begin{pmatrix}\dfrac{c^{k-1}-2c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\\ -\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top\left(\dfrac{c^{k-1}-2c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\right)\\ -\dfrac{\gamma}{\rho^k}g^k+v\mathbf{1}\end{pmatrix}\right\|_F\end{aligned} \tag{П.24}$$

и

$$\begin{aligned}&\left\|u^k-\frac{1}{n}\mathbf{1}\mathbf{1}^\top u^k+\frac{1}{n}\mathbf{1}+v\mathbf{1}-g^k\right\|_F=\\ &=\left\|\begin{array}{l}y^k-\dfrac{c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k+\dfrac{1}{n}\mathbf{1}+v\mathbf{1}-g^k\\ -\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top\left(y^k-\dfrac{c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\right)\end{array}\right\|_F=\\ &=\left\|\begin{array}{l}\dfrac{c^{k-1}-c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k+v\mathbf{1}\\ -\dfrac{1}{n}\mathbf{1}\mathbf{1}^\top\left(\dfrac{c^{k-1}-c^k}{\rho^k}+\dfrac{\gamma}{\rho^k}P_w(m)-\dfrac{1}{\rho^k}t^k\right)\end{array}\right\|_F.\end{aligned} \tag{П.25}$$

Ниже показано, что v стремится к 0, когда k стремится к бесконечности. В соответствии с определением u^k имеет место следующее уравнение:

$$\lim_{k\to\infty}\left\|u^k-g^k\right\|_2=\lim_{k\to\infty}\left\|g^k+\frac{c^{k-1}-2c^k}{\rho^k}+\frac{\gamma}{\rho^k}P_w(m)-\frac{1}{\rho^k}t^k-g^k\right\|_2=0. \tag{П.26}$$

Сопоставление (2.27) и (2.28), (П.26) влечет $f(0) = 0$, когда k стремится к бесконечности, т.е. v стремится к 0. В итоге, $\lim\limits_{k\to\infty}\left\|g^{k+1} - g^{k}\right\|_2 = 0$ и, следовательно,

$$\lim_{k\to\infty}\left\|\mathcal{G}^{k+1} - \mathcal{G}^{k}\right\|_F = 0. \tag{П.27}$$

Для $\mathcal{Y}^k$ справедливо следующее неравенство:

$$\begin{aligned}\left\|\mathcal{Y}^{k+1} - \mathcal{Y}^{k}\right\|_F &= \left\|\mathcal{Y}^{k+1} - \mathcal{G}^{k+1} - \mathcal{Y}^{k} + \mathcal{G}^{k} + \mathcal{G}^{k+1} - \mathcal{G}^{k}\right\|_F \le \\ &\le \left\|\mathcal{Y}^{k+1} - \mathcal{G}^{k+1}\right\|_F + \left\|\mathcal{Y}^{k} - \mathcal{G}^{k}\right\|_F + \left\|\mathcal{G}^{k+1} - \mathcal{G}^{k}\right\|_F.\end{aligned} \tag{П.28}$$

Таким образом, комбинируя (П.22) и (П.27), имеем

$$\lim_{k\to\infty}\left\|\mathcal{Y}^{k+1} - \mathcal{Y}^{k}\right\|_F = 0. \tag{П.29}$$

СПИСОК ЛИТЕРАТУРЫ

1. *Zhao J., Xie X., Xu X., Sun S.* Multi-view Learning Overview: Recent Progress and New Challenges // Information Fusion. 2017. V. 38. P. 43–54.
2. *Liu Y., Fan L., Zhang C., Zhou T., Xiao Z., Geng L., Shen D.* Incomplete Multi-modal Representation Learning for Alzheimer's Disease Diagnosis // Medical Image Analysis. 2021. V. 69. P. 101953.
3. *Qiao L., Zhang L., Chen S., Shen D.* Data-driven Graph Construction and Graph Learning: A Review // Neurocomputing. 2018. V. 312. P. 336–351.
4. *Wen J., Xu Y., Liu H.* Incomplete Multiview Spectral Clustering with Adaptive Graph Learning // IEEE Trans. Cybernetics. 2020. V. 50. № 4. P. 1418–1429.
5. *Wen J., Zhang Zheng, Zhang Zhao, Fei L.K., Wang M.* Generalized Incomplete Multiview Clustering with Flexible Locality Structure Diffusion // IEEE Trans. Cybernetics. 2021. V. 51. № 1. P. 101–114.
6. *Zhang N., Sun S.* Incomplete Multiview Nonnegative Representation Learning with Multiple Graphs // Pattern Recognition. 2022. V. 123. P. 108412.
7. *Wen J., Yan K., Zhang Z., Xu Y., Wang J.Q., Fei L.K., Zhang B.* Adaptive Graph Completion Based Incomplete Multiview Clustering // IEEE Trans. Multimedia. 2021. V. 23. P. 2493–2504.
8. *Liu J., Teng S., Zhang W., Fang X., Fei L., Zhang Z.* Incomplete Multiview Subspace Clustering with Low-rank Tensor // Proc. IEEE Intern. Conf. Acoustics, Speech and Signal Processing. Toronto, Canada, 2021. P. 3180–3184.
9. *Wen J., Zhang Zheng, Zhang Zhao, Zhu L., Fei L.K., Zhang B., Xu Y.* Unified Tensor Framework for Incomplete Multiview Clustering and Missing-view Inferring // Proc. 35th AAAI Conf. Artificial Intelligence. AAAI Press: Palo Alto, CA, USA. 2021. V. 35. P. 10273–10281.
10. *Xia W., Gao Q., Wang Q., Gao X.* Tensor Completion-based Incomplete Multiview Clustering // IEEE Trans. Cybernetics. 2022. V. 52. № 12. P. 13635–13644.
11. *Blaschko M.B., Lampert C.H., Gretton A.* Semi-supervised Laplacian Regularization of Kernel Canonical Correlation Analysis // Proc. Joint Europ. Conf. Machine Learning and Knowledge Discovery in Databases. Antwerp, Belgium, 2008. P. 133–145.
12. *Chen X., Chen S., Xue H., Zhou X.* A Unified Dimensionality Reduction Framework for Semi-paired and Semi-supervised Multiview Data // Pattern Recognition. 2012. V. 45. № 5. P. 2005–2018.
13. *Zhou X., Chen X., Chen S.* Neighborhood Correlation Analysis for Semi-paired Two-view Data // Neural Processing Letters. 2013. V. 37. № 3. P. 335–354.
14. *Yuan Y., Wu Z., Li Y., Qiang J., Gou J., Zhu Y.* Regularized Multiset Neighborhood Correlation Analysis for Semi-paired Multiview Learning // Intern. Conf. Neural Information Processing. Vancouver, Canada, 2020. P. 616–625.
15. *Yang W., Shi Y., Gao Y., Wang L., Yang M.* Incomplete Data Oriented Multiview Dimension Reduction via Sparse Low-rank Representation // IEEE Trans. Neural Networks and Learning Systems. 2018. V. 29. № 12. P. 6276–6291.
16. *Zhu C., Chen C., Zhou R., Wei L., Zhang X.* A New Multiview Learning Machine with Incomplete Data // Pattern Analysis and Applications. 2020. V. 23. № 3. P. 1085–1116.
17. *Li S., Jiang Y., Zhou Z.* Partial Multiview Clustering // Proc. AAAI Conf. artificial intelligence. Québec City, Canada, 2014. V. 28. № 1.
18. *Xu C., Tao D., Xu C.* Multiview Learning with Incomplete Views // IEEE Trans. Image Processing. 2015. V. 24. № 12. P. 5812–5825.

19. *Wen J., Zhang Z., Xu Y., Zhong Z.* Incomplete Multiview Clustering via Graph Regularized Matrix Factorization // Proc. European Conf. Computer Vision Workshops. Munich, Germany, 2018. P. 1–16.
20. *Hu M., Chen S.* Doubly Aligned Incomplete Multiview Clustering // Proc. Intern. Joint Conf. Artificial Intelligence. Stockholm, Sweden, 2018. P. 2262–2268.
21. *Hu M., Chen S.* One-pass Incomplete Multiview Clustering // Proc. AAAI Conf. Artificial Intelligence. Honolulu, Hawaii, USA, 2019. V. 33. P. 3838–3845.
22. *Liu J., Teng S., Fei L., Zhang W., Fang X., Zhang Z., Wu N.* A Novel Consensus Learning Approach to Incomplete Multiview Clustering // Pattern Recognition. 2021. V. 115. P. 107890.
23. *Liu X., Zhu X., Li M., Wang L., Zhu E., Liu T., Kloft M., Shen D., Yin J., Gao W.* Multiple Kernel k-means with Incomplete Kernels // IEEE Trans. Pattern Analysis and Machine Intelligence. 2019. V. 42. № 5. P. 1191–1204.
24. *Wen J., Sun H., Fei L., Li J., Zhang Z., Zhang B.* Consensus Guided Incomplete Multiview Spectral Clustering // Neural Networks. 2021. V. 133. P. 207–219.
25. *Zhuge W., Luo T., Tao H., Hou C., Yi D.* Multiview Spectral Clustering with Incomplete Graphs // IEEE Access. 2020. V. 8. P. 99820–99831.
26. *Liu X., Zhu X., Li M., Wang L., Tang C., Yin J., Shen D., Wang H., Gao W.* Late Fusion Incomplete Multiview Clustering // IEEE Trans. Pattern Analysis and Machine Intelligence. 2018. V. 41. № 10. P. 2410–2423.
27. *Zheng X., Liu X., Chen J., Zhu E.* Adaptive Partial Graph Learning and Fusion for Incomplete Multiview Clustering // Intern. J. Intelligent Systems. 2022. V. 37. № 1. P. 991–1009.
28. *Xie M., Ye Z., Pan G., Liu X.* Incomplete Multiview Subspace Clustering with Adaptive Instance Sample Mapping and Deep Feature Fusion // Applied Intelligence. 2021. V. 51. № 8. P. 5584–5597.
29. *Zhao L., Chen Z., Yang Y., Wang Z.J., Leung V.C.* Incomplete Multiview Clustering via Deep Semantic Mapping // Neurocomputing. 2018. V. 275. P. 1053–1062.
30. *Zhang C., Han Z., Fu H., Zhou J.T., Hu Q.* CPM-nets: Cross Partial Multiview Networks // Advances in Neural Information Processing Systems. 2019. V. 32.
31. *Wang Q., Ding Z., Tao Z., Gao Q., Fu Y.* Partial Multiview Clustering via Consistent GAN // Proc. IEEE Intern. Conf. Data Mining. Singapore, 2018. P. 1290–1295.
32. *Xu C., Liu H., Guan Z., Wu X., Tan J., Ling B.* Adversarial Incomplete Multiview Subspace Clustering Networks // IEEE Trans. Cybernetics. 2022. V. 52. № 10. P. 10490–10503.
33. *Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y.* Generative Adversarial Networks // Comm. ACM. 2020. V. 63. № 11. P. 139–144.
34. *Lin Y., Gou Y., Liu Z., Li B., Lv J., Peng X.* Completer: Incomplete Multiview Clustering via Contrastive Prediction // Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition. Nashville, TN, USA, 2021. P. 11174–11183.
35. *Zhang B., Hao J., Ma G., Yue J., Shi Z.* Semi-paired Probabilistic Canonical Correlation Analysis // Intelligent Information Processing VII. IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer, 2014. V. 432.
36. *Matsuura T., Saito K., Ushiku Y., Harada T.* Generalized Bayesian Canonical Correlation Analysis with Missing Modalities // 15th Europ. Conf. Computer Vision (ECCV). Munich, Germany, 2018. V. 11134. P. 641–656.
37. *Li P., Chen S.* Shared Gaussian Process Latent Variable Model for Incomplete Multiview Clustering // IEEE Trans. Cybernetics. 2018. V. 50. № 1. P. 61–73.
38. *Kamada C., Kanezaki A., Harada T.* Probabilistic Semi-canonical Correlation Analysis // Proc. 23rd ACM Intern. Conf. Multimedia. Brisbane, Australia, 2015. P. 1131–1134.
39. *Wang C.* Variational Bayesian Approach to Canonical Correlation Analysis // IEEE Trans. Neural Networks. 2007. V. 18. № 3. P. 905–910.
40. *Kimura A., Sugiyama M., Nakano T., Kameoka H., Sakano H., Maeda E., Ishiguro K.* SemiCCA: Efficient Semi-supervised Learning of Canonical Correlations // Information and Media Technologies. 2013. V. 8. № 2. P. 311–318.
41. *Luo Y., Tao D., Ramamohanarao K., Xu C., Wen Y.* Tensor Canonical Correlation Analysis for Multiview Dimension Reduction // IEEE Trans. Knowledge and Data Engineering. 2015. V. 27. № 11. P. 3111–3124.
42. *Wong H., Wang L., Chan R., Zeng T.* Deep Tensor CCA for Multiview Learning // IEEE Trans. Big Data. 2021. V. 8. P. 1664–1677.
43. *Cheng M., Jing L., Ng M.K.* Tensor-based Low-dimensional Representation Learning for Multiview Clustering // IEEE Trans. Image Processing. 2018. V. 28. № 5. P. 2399–2414.
44. *Zhang C., Fu H., Liu S., Liu G., Cao X.* Low-rank Tensor Constrained Multiview Subspace Clustering // Proc. IEEE Intern. Conf. Computer Vision. Santiago, Chile, 2015. P. 1582–1590.
45. *Wu J., Lin Z., Zha H.* Essential Tensor Learning for Multiview Spectral Clustering // IEEE Trans. Image Processing. 2019. V. 28. № 12. P. 5910–5922.
46. *Carroll J.* Generalization of Canonical Correlation Analysis to Three or More Sets of Variables // Proc. 76th Annual Convention of the American Psychological Association. 1968. V. 3. P. 227–228.

47. *Chen J., Wang G., Giannakis G.B.* Graph Multiview Canonical Correlation Analysis // IEEE Trans. Signal Processing. 2019. V. 67. № 11 P. 2826–2838.
48. *Nie F., Li J., Li X.* Self-weighted Multiview Clustering with Multiple Graphs // Intern. Joint Conf. Artificial Intelligence. Melbourne, Australia, 2017. P. 2564–2570.
49. *Fan K.* On a Theorem of Weyl Concerning Eigenvalues of Linear Transformations // Proc. National Academy of Sciences. 1949. V. 35. № 11. P. 652–655.
50. *van Breukelen M., Duin R.P.W., Tax D.M.J., den Hartog J.E.* Handwritten Digit Recognition by Combined Classifiers // Kybernetika. 1998. V. 34. № 4. P. 381–386.
51. *Greene D.* 3 Sources Dataset // Электронный ресурс: http://erdos.ucd.ie/datasets/3sources.html. Дата доступа: 7 января 2023 г.
52. *Greene D., Cunningham P.* Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering // Proc. 23rd Intern. Conf. Machine Learning. Pittsburgh, PA, USA, 2006. P. 377–384.
53. *Samaria F.S., Harter A.C.* Parameterisation of a Stochastic Model for Human Face Identification // Proc. IEEE Workshop on Applications of Computer Vision. Sarasota, FL, USA, 1994. P. 138–142.
54. *Zhao H., Liu H., Fu Y.* Incomplete Multi-modal Visual Data Grouping // Proc. Intern. Joint Conf. Artificial Intelligence. N.Y., USA, 2016. P. 2392–2398.
55. *Xie Y., Gu S., Liu Y., Zuo W., Zhang W., Zhang L.* Weighted Schatten p-norm Minimization for Image Denoising and Background Subtraction // IEEE Trans. Image Processing. 2016. V. 25. № 10. P. 4842–4857.