

# МЕТОДЫ ИНФОРМАТИКИ В МИКРОСКОПИИ МАТЕРИАЛОВ

## METHODS OF COMPUTER SCIENCE IN MICROSCOPY OF MATERIALS

Z. Temirgaliyev

*Summary.* The article deals with the methods of computer science and data science in the field of electron microscopy of materials. Attention is paid to the problems of noise, image distortion, spectral mixing. The ways of solving the drift problem are shown. The conclusion is made about the expediency of combining several methods.

*Keywords:* microscopy of materials; Poisson noise; Gaussian noise; Gaussian distributions; principal component analysis; distortion correction.

**Темиргалиев Жолдасбек Елубаевич**

Веб-разработчик, соискатель, Университет КИМЭП,  
Алма-Аты, Казахстан  
tzhe@yandex.ru

*Аннотация.* В статье рассмотрены методы информатики и науки о данных в области электронной микроскопии материалов. Уделено внимание проблемам шума, искажения изображения, спектрального смешения. Показаны пути решения проблемы дрейфа. Сделан вывод о целесообразности сочетания нескольких методов.

*Ключевые слова:* микроскопия материалов; шум Пуассона; гауссовский шум; распределения Гаусса; метод главных компонент; преобразование Анскомба; коррекция искажений.

**И**ntenсивное развитие компьютерной техники высокой вычислительной мощности, связанное с ростом объема и сложности данных и сопряженное с достаточной степенью доступа к ней, повлекло за собой разработку соответствующих математических методов и компьютерных алгоритмов [1, 2].

Разработанные математические модели и алгоритмы для таких областей, как машинное обучение, медицинская визуализация и компьютерное зрение, адаптируются к данным о характеристиках материалов.

Для высокоскоростной визуализации динамических процессов от быстрых пиксельных детекторов собираются большие, плотно сэмплированные наборы дифракционных данных и спектральных изображений в сложных экспериментах с измерением целых многомерных данных.

Одной из возможных концептуальных основ роли информатики в характеристике материалов является разделение процесса характеризации на три этапа: получение данных, формирование из этих данных информации о материалах, и формирование обобщенных знаний.

Например, в схеме: данные–информация–знания при рассмотрении отображения различных фаз, присутствующих в композите сначала, получают набор чувствительных к композиции спектральных изображений, состоящий из спектра энергетической дисперсии характерных рентгеновских лучей в сетке позиций на образце, затем анализируются данные для создания информации о выборке, такой как пространственное распределение элементов или состав фаз, а результаты измерений распределения фаз вносятся в историю обработки образца или

его механических свойств для создания обобщенных знаний и продвижения уровня техники в этой области.

Инструменты информатики применяются в первую очередь на первых двух этапах этой цепочки. Различные инструменты могут улучшить качество данных как в процессе их сбора, например, путем регистрации и коррекции погрешности, так и после сбора, например, путем шумоподавления.

В сложных наборах данных, инструменты информатики, могут включать в себя пиковую подгонку, деконволюцию реакции инструмента или подходы к компьютерному обучению, такие как анализ компонентов или кластеризация для определения важных функций.

Улучшение данных предполагает приближение реальных, несовершенных экспериментальных результатов к идеальному эксперименту

Различные подходы к достижению этих задач сводятся к двум типам проблем: снижение шума в экспериментальных данных с особым упором на шум Пуассона, который часто доминирует в экспериментах по подсчету частиц, таких как электронная микроскопия, и коррекция дрейфа и искажения, особенно изображений, полученных путем сканирования зонда через образец.

При снижении шума, если экспериментальные данные обозначить  $y_i$  при  $i=1 \dots N$ , а данные без шума обозначить  $x_i$ , то математическая задача состоит в том, чтобы получить из  $y_i$  значение  $x_i^*$  наиболее близкое к значению  $x_i$ .

Примеры возможных типов шума в микроскопии материалов включают детектор или электронный шум счи-

тивания, который иногда может быть аппроксимирован как аддитивная случайная величина, полученная из распределения Гаусса  $G$  с фиксированным средним  $m$  и дисперсией  $\sigma$ , и шум, возникающий в результате подсчета ограниченного числа электронов или фотонов, которые извлекаются из распределения Пуассона  $P$  со средним значением и дисперсией, заданной истинным значением данных.

Таким образом, достаточно общее представление  $y$  задается формулой

$$y = P(x) + G(m, \sigma) \quad (1)$$

где  $P$  и  $G$  — случайные величины.

Алгоритмы рендеринга часто предполагают, что в данных преобладает аддитивный белый гауссовский шум, что справедливо для сильных сигналов, например при фотографировании. При том, что гауссовские случайные переменные имеют хорошо развитые математические свойства. Однако он не является хорошей моделью для многих экспериментов по микроскопии материалов на основе частиц, в которых преобладают шумы Пуассона. И поскольку дисперсия распределения Пуассона усредняется, уровень шума контролируется силой сигнала, а не коррелируется с ним. В результате простое применение кодов, разработанных применительно к аддитивному белому гауссовскому шуму для данных микроскопии, может не дать корректных результатов.

Существует два подхода к адаптации методов, разработанных для аддитивного белого гауссовского шума к шуму Пуассона. Одним из них является использование преобразования Анскомба для изменения данных с преобладанием шума Пуассона в форме с постоянной дисперсией, не зависящей от  $x$  [7]. Преобразование Анскомба заменяет  $y$  на  $A(Y) = 2\sqrt{y} + 3/8$ . Затем данные обрабатываются с использованием алгоритма аддитивного белого гауссовского шума. Обратное преобразование является прямым для  $y$  без шума, но оценка обратного преобразования при наличии шума в  $y$  более сложна и при очень низкой интенсивности изображения становится ненадежной.

Другой подход — модифицирование алгоритма, чтобы непосредственно учитывать математические свойства пуассоновского шума. Например, общей проблемой в науке о данных является оценка подобия  $S$  между двумя сигналами  $y_{1,i}$  и  $y_{2,i}$ . Наиболее распространенным подходом является вычисление квадратной разности или нормы разницы между двумя векторами данных,

$$S = \frac{1}{N} \sum_{i=1}^N (y_{1,i} - y_{2,i})^2 \quad (2)$$

Эта мера  $S$  нестабильна при наличии большого пуассоновского шума, так как она подчеркивает экстремальные значения в  $y$ . Лучшей мерой  $S$  является логарифм отношения максимального правдоподобия Пуассона:

$$S = \frac{1}{N} \sum_{i=1}^N \log[pml(y_{1,i}, y_{2,i})] \quad (3)$$

где:

$$pml(y_{1,i}, y_{2,i}) = \frac{\max_{\lambda \in (0, \infty)} [P(y_{1,i}, \lambda) P(y_{2,i}, \lambda)]}{\max_{\lambda \in (0, \infty)} P(y_{1,i}, \lambda) \max_{\lambda \in (0, \infty)} P(y_{2,i}, \lambda)} \quad (4)$$

$\lambda$  — дисперсия распределения Пуассона.

Тогда:

$$S = \frac{1}{N} \sum_{i=1}^N y_{1,i} \log y_{1,i} + y_{2,i} \log y_{2,i} - (y_{1,i} + y_{2,i}) \log \left( \frac{y_{1,i} + y_{2,i}}{2} \right)$$

Такие изменения более устойчивы к высоким уровням шума, но требуют как модификаций алгоритмов, так и их реализации.

Один из подходов к шумоподавлению опирается на концепцию низкоразмерного представления данных. Каждый набор данных  $x$  представляется как точка в  $N$ -мерном пространстве. Однако конкретный набор данных в этом  $N$ -мерном пространстве часто занимает только одно подпространство. Если найти базисные векторы для подпространства можно представить данные  $x$  как линейную комбинацию этих базисных векторов.

Существуют различные подходы к определению размерности подпространства и нахождению правильного набора базисных векторов. Наиболее ранним и наиболее широко применяемым в материалах микроскопии является метод главных компонент, который может быть адаптирован для шума Пуассона в данных.

Весовой метод главных компонент использует преобразование данных, аналогичных описанному выше преобразованию Анскомба, а в пуассоновском методе главных компонент используется модификация меры сходства, аналогичная максимальной вероятности Пуассона (формула 5). Весовой метод главных компонент для данных электронной микроскопии реализуется в различных пакетах, включая коммерческое программное обеспечение и программное обеспечение с открытым исходным кодом.

Наиболее распространенным подходом для метода главных компонент является график собственных значений, который показывает дисперсию компонентных сигналов от самых больших до самых маленьких. Количество компонент выбирается, когда дисперсия падает ниже некоторого порога или когда есть разрыв в наклоне графика в линейном масштабе либо в масштабе шкалы.

Данные экспериментальной микроскопии также часто повреждаются дрейфом и искажениями. Это может быть связано как с дрейфом самого образца, или конструктивных элементов микроскопа так и с дрейфом того и другого. Искажения могут быть согласованными при каждом измерении, или они могут варьироваться случайным образом при каждом измерении.

Дрейф и искажение особенно распространены и проблематичны для сканирующих микроскопов. Сканирование часто происходит медленнее, чем визуализация, что приводит к увеличению времени дрейфа, смещению образца и искажению изображения.

Обобщенно проблема дрейфа и искажения имеет два типа решения. Простейшая форма задачи задается набором из  $i=1...N$  изображений идентичного объекта с дрейфом и искажением, и сводится к восстановлению истинного изображения. Случай только дрейфа, скорректированного вектором сдвига  $f_i$ , применяемого ко всем пикселям в каждом изображении  $i$  может быть реализован рядом изображений одного и того же объекта. Поскольку детектор фиксирован, сетка выборки пикселей одинакова для каждого изображения. Вектор сдвига  $f_i$  может быть оценен либо по самим данным изображения, либо по образам отдельной опорной зоны между кадрами, которая, будет жестко соединена с отображаемой областью.

Регистрация эталонной области может периодически использоваться в течение получения одного изображения. Более точной может быть встроенная регистрация, поскольку она напрямую использует зарегистрированные изображения.

Наиболее распространенным средством для получения  $f_i$  является максимизация взаимной корреляции между изображениями. Для решеточных изображений с атомным разрешением фазовая кросс-корреляция сильнее, чем корреляция амплитуды, но для непериодических изображений лучше корреляция амплитуды.

Перекрестная корреляция может быть заменена средней квадратичной разницей (уравнение 2) максимальным правдоподобием Пуассона (уравнение 5) или полной вариацией.

В случае нежесткой регистрации каждый пиксель в каждом изображении допускает сдвиг  $f_i(x, y)$ , поэтому в алгоритм должна быть встроена предварительная информация о режиме захвата изображения или выборке, а также предварительные знания об образце, например, использование эталонного изображения.

Другим подходом к коррекции дрейфа и искажения является введение предварительной информации о данных полученных путем поворота сканирования относительно образца по серии изображений. Искажения в проверке привязаны к кадру отсчета сканирования, и дрейф образца привязан к образцу отсчета, поэтому два эффекта могут быть отделены друг от друга и исправлены. В отличие от нежесткой регистрации, метод вращения также может корректировать систематические искажения сканирования, присутствующие в каждом изображении.

Одной из доступных возможностей микроскопии материалов и материаловедения в более широком смысле является адаптация инструментов и подхода, разработанных в других областях науки, к материалам. Как показывают приведенные выше примеры, разные проблемы требуют разных подходов, и зачастую наилучшие результаты исходят из разумного сочетания нескольких подходов.

Существенной проблемой является рост распространенности наборов данных по микроскопии материалов, которые слишком велики для детального анализа при индивидуальных измерениях.

Связанная с этим возможность заключается в том, чтобы пропустить этап улучшения данных и перейти непосредственно к улучшению информации о материалах.

Поскольку все обсуждаемые выше подходы становятся все более существенными для материальной микроскопии становится актуальным вопрос программного обеспечения, которое их реализует, и вычислительных ресурсов, которые их поддерживают.

Для выполнения вычислительных алгоритмов на больших наборах данных целесообразно применение программного обеспечения с открытым исходным кодом, написанное для работы на платформах, таких как MATLAB, IDL или Digital Micrograph.

В последнее время альтернативой установленному программному обеспечению конечного пользователя стали среды и приложения для анализа данных в Интернете. Большинство проблем с микроскопией материалов были полезны при применении более чем одной наукоемкой технологии данных, такой как нелокальный анализ компонент, а также анализ данных микроскопии сложных сканирующих задач.

ЛИТЕРАТУРА

1. Rajan K. Informatics for Materials Science and Engineering. Data-driven Discovery for Accelerated Experimentation and Application. — 1st Edition. — Elsevier, 2013. 544 p. — ISBN: 9780123943996.
2. Фульц Б. Просвечивающая электронная микроскопия и дифрактометрия материалов / Б. Фульц, Дж. М. Хау. — М.: Техносфера, 2011. — 872 с.
3. Burger, M. Level set and PDE based reconstruction methods in imaging / M. Burger. — Springer, 2008—319 p.
4. М. Кристал, И. Ясников. Сканирующая электронная микроскопия и рентгеноспектральный микроанализ. М.: Техносфера, 2009, 208 с.
5. Нестеренко Д. В. Формирование и обработка изображений электронной микроскопии / Компьютерная оптика, том 35, № 2—2011
6. Gorban A. N., Kegl B., Wunsch D., Zinovyev A. Y. (Eds.), Principal Manifolds for Data Visualisation and Dimension Reduction, Series: Lecture Notes in Computational Science and Engineering 58, Springer, Berlin — Heidelberg — New York, 2007, XXIV, 340 p. 82 illus. ISBN978-3-540-73749-0
7. Mäkitalo, M.; Foi, A. (2013), "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise", IEEE Transactions on Image Processing, 22 (1), pp. 91–103.
8. A. B. Yankovich, et al. Non-rigid registration and non-local principle component analysis to improve electron microscopy spectrum images / Nanotechnology, 27 (2016), p. 364

© Темиргалиев Жолдасбек Елубаевич ( tzh@yandex.ru ).

Журнал «Современная наука: актуальные проблемы теории и практики»

