ОПТИЧЕСКОЕ ОТОЖДЕСТВЛЕНИЕ РЕНТГЕНОВСКИХ ИСТОЧНИКОВ СРГ/еРОЗИТА НА ПРИМЕРЕ ГЛУБОКОГО ОБЗОРА ОБЛАСТИ ДЫРЫ ЛОКМАНА

© 2022 г. С. Д. Быков^{1,2*}, М. И. Бельведерский^{3,4}, М. Р. Гильфанов^{2,3}

¹ Казанский университет, Казань, Россия ² Институт астрофизики общества им. Макса Планка, Гархинг, Германия ³ Институт космических исследований РАН, Москва, Россия ⁴ Национальный исследовательский университет "Высшая школа экономики", Москва, Россия Поступила в редакцию 21.10.2022 г. После доработки 21.10.2022 г.; принята к публикации 15.11.2022 г.

Представлен метод оптического отождествления источников, детектируемых в широкоугольных рентгеновских обзорах неба. Для этой цели была создана и обучена нейросетевая модель для характеризации фотометрических признаков популяций оптических компаньонов рентгеновских источников и оптических объектов поля. Результат обработки фотометрической информации используется для вероятностного отождествления рентгеновских источников с источниками оптического обзора DESI LIS. Эффективность метода проиллюстрирована на примере глубокого обзора области Дыры Локмана телескопом еPO3ИTA обсерватории СРГ. Для оценки точности модели создана валидационная выборка на основе каталогов рентгеновских источников спутников Chandra и XMM-Newton. Точность отождествления в нашем методе достигает 94% по всему рентгеновскому каталогу области Дыры Локмана и 97% для источников с потоком $F_{\rm x,\ 0.5-2}>10^{-14}$ эрг см⁻² с⁻¹. Обсуждается дальнейшее развитие модели отождествления и шаги, необходимые для применения ее к данным всего обзора неба СРГ/еРОЗИТА.

Ключевые слова: рентгеновские источники, внегалактические источники, Дыра Локмана.

DOI: 10.31857/S0320010822110055

1. ВВЕДЕНИЕ

Орбитальная рентгеновская обсерватория СРГ (Сюняев и др., 2021) была запущена в июле 2019 г., и в декабре того же года начала обзор всего неба. На борту обсерватории находится два инструмента: телескоп СРГ/еРОЗИТА (Предель и др., 2021), работающий в диапазоне 0.3-9.0 кэВ, и телескоп СРГ/ART-XC им. М.Н. Павлинского (Павлинский и др., 2021), работающий в диапазоне 6-30 кэВ. Ожидается, что по итогам четырехлетнего обзора неба телескоп СРГ/еРОЗИТА откроет порядка трех миллионов активных ядер галактик, около ста тысяч скоплений и групп галактик, а также множество рентгеновски-ярких звезд и галактик (Прокопенко и Гильфанов, 2009; Мерлони и др., 2012; Пиллепич и др., 2012; Колодзиг и др., 2013a,b).

Для полноценного использования результатов рентгеновского обзора неба необходимо определить свойства объектов не только в рентгеновском, но и в других диапазонах электромагнитного спектра. Прежде всего речь идет об определении типа объекта и расстояния до него. Проблема решается через сопоставление источников рентгеновского излучения с известными объектами в оптическом (преимущественно) и других диапазонах спектра. Такая задача носит названия кроссотождествления (Сазерленд и Сандерс, 1992; Будавари и Лоредо, 2015; Пино и др., 2017). В настоящей работе мы рассматриваем задачу поиска оптических компаньонов рентгеновских источников. В дальнейшем рентгеновские источники мы будем называть "источники", а кандидаты в оптические компаньоны — "объекты". Особенностью задачи оптического отождествления является то, что ошибка положения рентгеновских источников примерно на порядок больше, чем ошибка положения оптических объектов. Кроме того, из-за высо-

^{*}Электронный адрес: sergei.d.bykov@gmail.com

кой плотности оптических объектов часто в "круге ошибок" положения рентгеновского источника находится несколько кандидатов в его оптические компаньоны. При этом часто сложно определить, какой из оптических объектов является настоящим компаньоном (или что компаньона нет вообще). Чтобы разрешить данную задачу, необходимо привлекать дополнительные данные, например, звездные величины или цвета объектов. Решению такой задачи на примере данных СРГ/еРОЗИТА посвящена эта статья.

Один из ранних подходов к учету дополнительной информации был продемонстрирован в работе Сазерленд и Сандерс (1992) с использованием метода отношений правдоподобия (MLR-метод). Метод берет в расчет положения источников в сравниваемых каталогах (включая ошибки), их число на единицу площади неба и распределение по звездной величине. Впоследствии метод был развит в работах Будавари и Сзалай (2008); Нэйлор и др. (2013); Будавари и Лоредо (2015); Пино и др. (2017); Сальвато и др. (2018) с обобщением на вариант кросс-корреляции более чем двух каталогов, а так же построением вероятностной (байесовой) модели для учета не только звездной величины, но и набора измерений любого вида.

История применения различных методов поиска компаньонов к рентгеновским обзорам (преимущественно телескопов Chandra и XMM-Newton) обширна (Бруса и др., 2007; Луо и др., 2010; Нэйлор и др., 2013; Марчеси и др., 2016; Ананна и др., 2017; Луо и др., 2017; Пино и др., 2017; Сальвато и др., 2018; Чен и др., 2018; Бельведерский и др., 2022; Сальвато и др., 2022). В широком смысле разные методы используют похожие подходы к задаче отождествления. На первом шаге вычисляются распределения звездных величин (и иногда так же цветов или других признаков) для оптических компаньонов и объектов поля, а на втором шаге эта информация комбинируется с информацией о взаимном положении источников из разных каталогов и их позиционных ошибках. Достигаемая точность отождествления зависит от конкретного метода учета фотометрической информации и от качества рентгеновских и оптических каталогов, их глубины и позиционной точности. Целью настоящей работы является решение задачи оптического отождествления рентгеновских источников для параметров обзора всего неба телескопом еРОЗИТА обсерватории СРГ (см. также Бельведерский и др., 2022). В работе излагается метод отождествления, его применение к данным глубокого обзора области Дыры Локмана, а также оценивается точность предложенного метода с помощью валидационной выборки. Основной целью работы являются представление и валидация

метода отождествления рентгеновских источников СРГ/еРОЗИТА. Характеристики рентгеновских наблюдений, каталог источников и результаты их оптического отождествления и классификации приведены в сопутствующих статьях серии публикаций, посвященных глубокому обзору Дыры Локмана телескопом СРГ/еРОЗИТА (Гильфанов и др. 2022, в подготовке; Бельведереский и др. 2022, в подготовке).

Статья построена следующим образом. В разделе 2 описываются используемые данные, в разделе 3 описываются построение модели отождествления и создание обучающей и валидационной выборок объектов. В разделе 4 мы обсуждаем результаты и делаем выводы.

2. ДАННЫЕ

2.1. Рентгеновские данные

Дыра Локмана — это зона низкого межзвездного поглощения (Локман и др., 1986), которая была изучена телескопом СРГ/еРОЗИТА в рамках верификационной фазы наблюдений телескопов обсерватории СРГ осенью 2019 г. Глубокий обзор этой области был проведен в октябре 2019 г. Отпечаток обзора на небе имеет размер $5^{\circ} \times 3.7^{\circ}$ с площадью порядка 18.5 кв. градусов. Глубина экспозиции составила 8 ксек на точку, чувствительность в диапазоне 0.3-2.3 кэВ составила $\sim 3 \times 10^{-15}$ эрг см⁻² с⁻¹. Примерные координаты центра исследованной области RA = 162°, Dec = 58°. Дальнейшие подробности наблюдений, рентгеновских данных и каталога источников приведены в работе Гильфанова и др. (2022).

Модель оптического отождествления отрабатывалась и тестировалась на каталоге рентгеновских источников еРОЗИТА Дыры Локмана средней и большой яркости (Гильфанов и др., 2022). В этот каталог включены точечные источники с достоверностью детектирования, выраженной в терминах правдоподобия DL > 10, что для Гауссова распределения примерно соответствует значимости $\approx 4\sigma$. Рентгеновский каталог включает 6885 источников.

Для построения модели отождествления и ее валидации мы использовали данные орбитальных рентгеновских обсерваторий Chandra и XMM-Newton. Данные Chandra брались из каталога Chandra Source Catalog 2.0, CSC (Эванс и др., 2010), и данные XMM-Newton из каталога 4XMM DR10 (Вебб и др., 2020). Эти каталоги были отфильтрованы по флагам качества детектирования, для того, чтобы минимизировать количество ложных рентгеновских источников¹. Таким образом, в зоне обзора Дыры Локмана СРГ/еРОЗИТА было найдено 2029 и 1316 источников Chandra и XMM соответственно. Данные этих обсерваторий покрывают около 20% площади обзора СРГ/еРОЗИТА. Расположение источников СРГ/еРОЗИТА, Chandra и XMM-Newton в поле Дыры Локмана приведено в работе Гильфанова и др. (2022).

2.2. Оптические данные

В качестве каталога оптических источников, среди которых мы будем искать компаньоны источникам СРГ/еРОЗИТА, был выбран фотометрический каталог DESI Legacy Imaging Surveys (DESI LIS, Дей и др., 2019) из-за его большой глубины и плошади². Каталог составлен по данным трех телескопов: BASS (фильтры g,r), MzLS (z) и DECaLS (g,r,z). Эти три телескопа имеют разные предельные чувствительности в разных фильтрах. DECaLS предоставляет данные на DEC $< 32^\circ$, поэтому для области Дыры Локмана (DEC $\approx 58^{\circ}$) и изучения фотометрических признаков мы использовали данные BASS и MzLS. Кроме того, каталог DESI LIS предоставляет данные об инфракрасных потоках в четырех фильтрах WISE, если они доступны. Все потоки были поправлены на поглощение (dereddened).

Данные DESI использовались во всем поле Локмана (со средней плотностью объектов ~58 000 источников на квадратный градус) для отождествления источников еРОЗИТА. Для обучения фотометрической модели использованы все объекты DESI вокруг источников CSC 2.0 на небе с DEC > 32° (см. п. 3.1.1).

Для всех объектов DESI мы вычислили отношение сигнал—шум для измерений во всех фильтрах (g, r, z, w1, w2, w3, w4). Если измерение в каком-нибудь фильтре имело отношение сигнал/шум меньше 3, то мы считали, что измерение в данном фильтре отсутствует.

3. МОДЕЛЬ

В этом разделе описаны фотометрическая модель отбора кандидатов в оптические компаньоны и ее применение к данным еРОЗИТА. В подразделе 3.1 описывается фотометрическая

модель, а именно, в п. 3.1.1 — обучающая выборка, а в п. 3.1.2 — фотометрический классификатор. В подразделе 3.2 обсуждается валидационная выборка в Дыре Локмана, а в подразделе 3.3 применение байесова подхода к задаче отождествления и его результаты для источников еРОЗИТА.

3.1. Фотометрическая модель

В этой части мы изучаем фотометрические признаки двух популяций: 1) объекты DESI LIS, которые являются настоящими компаньонами рентгеновских источников (в данном случае — источников Chandra); объекты DESI LIS, которые не являются компаньонами рентгеновских источников — так называемые полевые источники. Такое разделение оптических объектов на два класса мотивировано тем, что источники рентгеновского излучения являются достаточно редким и специфическим типом небесных объектов, так что их оптические компаньоны обладают характеристиками, отличными от объектов поля.

Для отбора оптических объектов в первый класс мы использовали однозначные оптические компаньоны рентгеновских источников из каталога Chandra, имеющего высокую точность локализации. Процедура отбора источников с однозначными оптическими компаньонами будет описана ниже.

3.1.1. Обучающая выборка. Мы использовали данные CSC 2.0 на внегалактическом небе ($|b| > 20^{\circ}$), в зоне покрытия DESI LIS, исключая о бласть Дыры Локмана. Полученная выборка из 157958 источников Chandra была кросскоррелирована с каталогом DESI с радиусом поиска 30 угл. сек. Таким образом, было выделено около 3 млн объектов DESI. Для каждого источника Chandra была определена локальная плотность оптических объектов ρ_{desi} , исходя из их количества в кольце 10–30 угл. сек. Чтобы контролировать частоту случайных совпадений между CSC и DESI, для каждого рентгеновского источника вычислен радиус r_{false} случайного совпадения по формуле

$$r_{\rm false} = \sqrt{-\frac{\ln(1 - f_{\rm lim})}{\pi \rho_{\rm desi}}},\tag{1}$$

где $f_{\rm lim}$ — вероятность найти один и более оптических объектов на расстоянии $r < r_{\rm false}$ от рентгеновского источника (Бельведерский и др., 2022). Величина $r_{\rm false}$ вычислялась в предположении однородного распределения оптических объектов с плотностью $\rho_{\rm desi}$. В качестве порогового значения вероятности мы выбрали $f_{\rm lim} = 0.03$. Для каждого источника из каталога Chandra также вычисляется ошибка положения r_{98} , соответствующая радиусу,

¹Фильтры для CSC: conf_flag, extent_flag, sat_src_flag, pileup_flag, dither_warning_flag равны false, likelihood>10. Фильтры для 4XMM DR10: sc_sum_flag=0 или 1, также sc_var_flag, sc_extent и confused равны false, sc_det_ml>10.

²https://www.legacysurvey.org

в пределах которого с вероятностью 98% находится настоящее положение рентгеновского источника.

Дальнейший отбор надежных оптических компаньонов проводился следующим образом: если в пределах r_{false} от рентгеновского источника найден единственный объект DESI LIS, при этом расстояние от него до источника Chandra меньше r_{98} , то такой объект считался *надежным оптическим компаньоном источника Chandra*. Мы также оставили только надежные источники Chandra (см. п. 2.1). Таких пар Chandra-DESI нашлось 71 993.

Для отбора *надежных источников поля* мы нашли все объекты DESI в кольце 10–30 угл. сек от источника Chandra. Количество полевых объектов составило 2 135 168.

На следующем шаге мы оставили в выборке надежных компаньонов только те объекты, у которых рентгеновский поток в диапазоне 0.5-2 кэВ источника Chandra превышал 3×10^{-15} эрг см⁻² c^{-1} , что соответствует порогу детектирования телескопа еРОЗИТА в поле Дыры Локмана. Заметим, что это важнейший шаг, поскольку обучающая и валидационная выборки должны быть эквивалентны по своим свойствам рентгеновскому каталогу, для которого будет производиться оптическое отождествление. Например, если бы мы включили в выборку надежных компаньонов все источники Chandra, то их фотометрические признаки сместились бы в сторону многочисленных тусклых источников, недоступных для СРГ/еРОЗИТА (Сальвато и др., 2018).

Наконец, мы отобрали только надежные компаньоны и объекты поля, расположенные на DEC > 32° (см. п. 2.2). Из источников поля, полученных таким образом, мы случайным образом отобрали половину, так как общее количество объектов получилось большое.

В результате этой процедуры отбора размер выборки надежных компаньонов составил 12452, а выборки надежных полевых объектов — 244008.

На рис. 1 мы иллюстрируем распределение источников этих двух выборок по звездной величине в фильтре g. Видно, что в среднем компаньоны рентгеновских источников ярче, чем полевые источники. В отсутствие дополнительной информации, этот факт можно использовать для определения более предпочтительных оптических компаньонов источникам еРОЗИТА (см. обсуждение ниже).

Предельное отношение сигнал/шум в тренировочной выборке (см. п. 2.2) изменено с 3 до 4.



Рис. 1. Плотность распределения по видимой звездной величине в фильтре g объектов DESI, которые попали в выборку надежных компаньонов (зеленая гистограмма) и надежных полевых источников (голубая гистограмма со штрихами). Вставка показывает область ярких объектов в логарифмическом масштабе по оси *y*.

3.1.2. Фотометрический классификатор. Задача классификации состоит в том, чтобы по набору фотометрической информации определить, является ли источник скорее компаньоном рентгеновского источника, или скорее является полевым источником. Такие задачи эффективно решаются с применением методов машинного обучения.

В качестве признаков для модели мы использовали звездные величины g, r, z, w1, w2 и цвета g-r, r-z, g-z, z-w1, r-w2, w1-w2. Мы не использовали фильтры w3, w4, так как в области Дыры Локмана эти измерения отсутствуют у 95 и 90% источников соответственно. Кроме того, у части рентгеновских источников отсутствуют измерения в фильтрах g, r, z, w1, w2 (18, 12, 6, 35, 60% соответственно). По этой причине мы ввели три модели: первая модель использует только признаки (звездные величины и цвета), относящиеся к фильтрам g, r, z, w1, и третья модель использует фильтры g, r, z, w1, и третья модель — g, r, z, w1, w2.

В качестве классификатора мы использовали нейронную сеть, на вход которой подаются соответствующие признаки модели, а на выходе получается число между 0 и 1, которое мы назвали пптаg. Далее опишем модель, основанную на всех признаках: g, r, z, w1 и w2.

Мы построили нейронную сеть с функцией активации tanh, в которой было 4 слоя по 8 нейронов в каждом. После каждого слоя для предотвращения переобучения добавлен исключающий (dropuot) слой с вероятностью 0.1. На выход нейронной сети добавлен слой с одним нейроном с логистической активацией. В качестве функции потерь использовалась двоичная



Рис. 2. Верхняя панель: плотность распределения по выводу нейронной сети (nnmag) объектов DESI, которые попали в тестовую выборку полевых источников (голубая гистограмма со штрихами) и компаньонов (зеленая гистограмма). Нижняя панель: кривые полноты ("recall", пунктирная линяя) и точности ("precision", сплошная линия) для компаньонов в зависимости от нижнего порога по nnmag, использованного для классификации оптического объекта в качестве компаньона рентгеновского источника.

кроссэнтропия (binary crossentropy). Количество источников класса "компаньон" — 10 480, класса "полевой" — 55 636; 30% данных использовались как тестовые, 20% использовались для кроссвалидации, а остальное — как обучающая выборка. Для обучения классификаторов использовался пакет KERAS³.

На рис. 2 (верхняя панель) представлена гистограмма вывода нейронной сети (пптад) на тестовой выборке (данные, которые не были использованы для обучения) для обоих классов объектов. Видно, что объекты класса 1 (компаньоны) имеют значения пптад ближе к единице, а объекты класса 0 (полевые источники) имеют пптад ближе к 0. Эту гистограмму можно сравнить с рис. 1, в которой разделение по величине g не такое очевидное. Также на рисунке на нижней панели показаны две метрики качества модели для разных значений порога nnmag: полнота (recall) и точность (precision). При данном значении nnmag источники назначаются компаньоном (полевым), если их значение nnmag больше (меньше) порогового⁴. В качестве общей метрики качества модели мы использовали значение точности в точке пересечения двух кривых, в случае с моделью g, r, z, w1, w2 это число равно 0.67. Отметим, что в точке пересечения двух кривых значения полноты и точности равны друг другу.

В табл. 1 описаны важные параметры для всех трех моделей: используемые признаки, размеры выборок, структуры нейронной сети и общее качество модели. Стоит отметить падение точности модели с потерей данных из инфракрасного диапазона — модель grzw1w2 имеет общую точность 0.67, grzw1 — 0.61, и grz — только 0.38.

3.2. Валидационная выборка в Дыре Локмана

Для того чтобы оценить качество кроссотождествления, нужно иметь выборку источников еРОЗИТА с известными оптическими компаньонами. Для этого мы привлекаем данные спутников Chandra и XMM-Newton в области Дыры Локмана. Мы использовали подход, аналогичный примененному в работе Бельведерского и др. (2022).

Процедура отбора следующая:

- 1. Мы провели кросс-корреляцию полных каталогов Chandra (XMM)⁵ с каталогом еРОЗИТА Дыры Локмана с радиусом поиска 30 угл. сек и выписали уникальные пары еРОЗИТА-Chandra (XMM), оставив из них только надежные детектирования Chandra/XMM. Таких получилось 461 (646).
- Для каждого источника Chandra (XMM) нашли все объекты DESI LIS в пределах 15 угл. сек и выбрали из них такие, которые (а) лежат в пределах 1.1 × r₉₈ от соответствующего рентгеновского источника и являются единственным оптическим объектом в данном радиусе, и (б) лежат в пределах r_{false}. Таким образом, было найдено 310 (383)

³https://keras.io

⁴Полнота вычисляется по формуле TP/(TP + FN), а точность как TP/(TP + FP), где true positive (TP) — количество правильно определенных компаньонов, false negative (FN) — компаньонов, определенных как полевые, false positive (FP) — полевых, определенных как компаньоны.

⁵Без фильтра по качеству, см. п. 2.1.

Модель	Признаки	Размер выборки ¹	Структура сети	Общее качество ²
nnmag_grzw1w2	g, r, z, w1, w2 g-r, r-z, g-z z-w1, r-w1, w1-w2	1: 10480 0: 55636	4 слоя по 8 нейронов Dropout rate: 0.1	0.67
nnmag_grzw1	g, r, z, w1 g-r, r-z, g-z, z-w1	1: 10814 0: 89670	4 слоя по 8 нейронов Dropout rate: 0.1	0.61
nnmag_grz	g, r, z g-r, r-z, g-z	1: 11092 0: 141845	2 слоя по 13 нейронов Dropout rate: 0.0	0.38

Таблица 1. Описание фотометрических моделей

¹ Класс 1 относится к компаньонам, класс 0 — к полевым источникам.

² Полнота в точке пересечения кривых полноты и точности, см. п. 3.1.2.

пар Chandra (XMM) — DESI (определения величин см. в п. 3.1.1). В качестве r_{false} выбран радиус 1.47", исходя из средней плотности DESI в области Дыры Локмана.

- После этого из пар Chandra (XMM)–DESI были исключены такие, в которых поток в диапазоне 0.5–2 кэВ по данным еРОЗИТА и Chandra/XMM отличается больше, чем в пять раз (Бельведерский и др., 2022) осталось 235 (374) источников.
- 4. Два каталога (eROSITA-Chandra-DESI и eROSITA-XMM-DESI) были совмещены, а дубликаты удалены. Получен каталог, состоящий из 548 троек eROSITA-Chandra (XMM)-DESI. Из него были удалены случаи, когда вокруг источника ePO3ИTA находились один надежный источник XMM/Chandra и два и более источника Chandra/XMM, не прошедшие фильтр на качество детектирования. Осталось 529 троек eROSITA-Chandra (XMM) DESI.

Это позволило точно определить соответствие между рентгеновским источником еРОЗИТА и оптическим объектом DESI. Этот каталог будет в дальнейшем использован для проверки качества отождествления.

Мы также создали каталог источников еРОЗИТА, у которых нет оптического компаньона в каталоге DESI LIS — это так называемые бездомные (hostless) источники. Для этого на втором шаге изменили критерий поиска, потребовав чтобы в пределах $2 \times r_{98}$ и $2 \times r_{false}$ не было ни одного оптического объекта DESI. Остальные шаги остались без изменений. Итоговое количество бездомных (бесхозных) источников еРОЗИТА в валидационной выборке составило 30.

3.3. NWAY и идентификация рентгеновских объектов в поле Дыры Локмана

Объединение информации о взаимном положении и позиционных ошибках, с одной стороны, и фотометрических свойствах объектов, с другой, проводилось с помощью кода NWAY⁶ (Сальвато и др., 2018), который использует байесов подход к вычислению вероятностей. На вход этой программе подаются два каталога — ePO3ИTA и DESI. Первоначально вычисляется вероятность совпадения каждой возможной пары, исходя только из астрометрических положений, ошибок на них, а также плотности источников на небе. Затем вероятности уточняются с использованием сторонней информации (далее — фотометрических приоров). При этом уточнение происходит, исходя из отношения плотностей распределения приоров для компаньонов и полевых источников. На выходе каждому объекту еРОЗИТА сопоставляется число pany, характеризующее наличие у него оптического компаньона, а каждому кандидату в компаньоны DESI — число p_i , характеризующее его вероятность быть этим компаньоном. Чем ближе числа к единице, тем надежнее соответствующее утверждение. Отметим, однако, что параметры pany и pi не следует интерпретировать как вероятности в строгом смысле.

В качестве фотометрического приора мы использовали величину пптад, полученную нейронной сетью — фотометрическим классификатором (п. 3.1.2, см. также пример на рис. 2). Выбор фотометрической модели (табл. 1) проводился в зависимости от того, какие именно признаки были доступны для того или иного объекта DESI LIS. При этом мы следили, чтобы не происходило "утечки" информации: для каждого оптического объекта использовался только один фотометрический приор, полученный моделью, обладающей наибольшей точностью, доступной для данного объекта (табл. 1). Если же фотометрической информации

⁶https://github.com/JohannesBuchner/nway

об объекте было недостаточно для применения ни одной из моделей в табл. 1 (например, отсутствовало измерение в фильтре z), то в качестве приора использовалась любая доступная информация (звездные величины и цвета), откалиброванная по выборке из секции 3.1.1, например, как проиллюстрировано на рис. 1.

При работе NWAY в качестве $1\sigma^7$ ошибки положения для источников еРОЗИТА мы использовали соответствующую откалиброванную позиционную ошибку из каталога рентгеновских источников еРОЗИТА (Гильфанов и др., 2022, в подготовке), а для источников DESI LIS — фиксированную ошибку 0.1 угл. сек.

В результате работы программы NWAY для каждого источника еРОЗИТА мы вычислили значение параметра p_{any} , а для каждого оптического объекта в радиусе 30" от рентгеновского источника — значение p_i . Кандидат с наибольшим p_i назначался наиболее вероятным компаньоном. Рентгеновские источники со значением p_{any} ниже порога $p_{any,0} = 0.12$ считались бездомными. Выбор значения порога $p_{any,0}$ обсуждается в следующем разделе.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В данном разделе представлены результаты оптического отождествления рентгеновских источников еРОЗИТА в области Дыры Локмана. В подразделе 4.1 приведена общая статистика по числу отождествлений, в подразделе 4.2 описывается калибровка позиционных ошибок рентгеновского каталога на основе валидационной выборки, в подразделе 4.3 описывается оценка точности, полноты и чистоты каталога компаньонов. В подразделе 4.4 мы обсуждаем обобщение модели на все небо и необходимые для этого изменения в процедуре.

4.1. Результаты отождествления

В результате применения вышеописанной процедуры отождествления, из 6885 источников еРОЗИТА в поле Дыры Локмана 6346 (92%) источников имеют $p_{any} > 0.12$, а 4360 (63%) $p_{any} > 0.8$. Для 5866 рентгеновских источников наиболее вероятным оптическим компаньоном оказался ближайший объект DESI LIS, а для 1019 — не ближайший. Для 6458 рентгеновских источников наиболее вероятный оптический компаньон лежит на угловых расстояниях sep < r_{98} , а для 427 — sep > r_{98} . Для 443 (6%) источников еРОЗИТА было найдено два и более кандидатов с



Рис. 3. Распределения по смещениям рентгеновских источников от их оптических компаньонов в валидационной выборке. Верхняя панель: голубая гистограмма показывает плотность распределения по величине sep/ σ . Нижняя панель: кумулятивное распределение этой же величины для двух вариантов отбора рентгеновских источников по правдоподобию детектирования DL (DL > 10 — голубая сплошная линия, DL > 15 — красная штрихпунктирная линия). На обеих панелях черной штриховой линией показано ожидаемое распределение Релея. На нижней панели горизонтальными линиями показаны уровни вероятности 95 и 98%.

близкими значениями p_i (в пределах фактора 2). Для 5919 (86%) источников еРОЗИТы один из оптических объектов имеет значение $p_i > 0.8$.

4.2. Проверка позиционных ошибок рентгеновских источников еРОЗИТА

529 валидационных пар eROSITA-DESI LIS позволяют нам проверить правильность определения ошибки положений рентгеновских источников. Если ошибки откалиброваны правильно, то ожидается, что распределение величины sep/ σ (где sep расстояние между наблюдаемым положением рентгеновского источника и его настоящим положением, σ — его ошибка положения) будет подчиняться распределению Релея (Уотсон и др., 2009;

⁷Соответствует вероятности 39% для двухмерного нормального распределения.

Чен и др., 2018; Брюннер и др., 2021; Бельведерский и др., 2022).

Результаты представлены на рис. 3, где показаны распределения этой величины, измеренные по валидационному каталогу. В качестве настоящего положения рентгеновского источника мы использовали положение его оптического компаньона, которое известно с точностью лучше долей секунды, в то время как типичная точность локализации рентгеновских источников составляет ~3–20 угл. сек. На рис. 3 приведены распределения для полного каталога, в котором правдоподобие детектирования DL > 10 и для источников с правдоподобием DL > 15 (что соответствует уровню достоверности $\approx 5\sigma$ для гауссова распределения). Чем выше порог, тем меньше в каталоге ложных источников. Так же нанесено ожидаемое распределение Релея.

Как видно на рис. 3, наблюдаемые распределения хорошо описываются распределением Релея. В случае полной валидационной выборки, включающей источники до уровня достоверности DL == 10, можно увидеть два источника на больших отклонениях sep $\gtrsim 5\sigma$. Визуальная инспекция этих областей на картах неба не выявила очевидных аномалий в рентгеновских или оптических данных. Появление этих двух источников является результатом случайного наложения, не связанных между собой источников ePO3ИTA и Chandra, при этом некоторые из источников еРОЗИТА могут быть ложными (т.е. результатом статистических флуктуаций на рентгеновском изображении), на что может указывать их сравнительно невысокая статистическая значимость ($<5\sigma$) в терминах гауссова распределения). Отметим, что при пороге значимости 4σ мы ожидаем порядка $\sim 10-20$ ложных источников в поле Дыры Локмана (см. более подробное обсуждение в работе Гильфанова и др. 2022). В распределении источников высокой значимости (DL > 15) такого хвоста в распределении нет оно хорошо согласуется с распределением Релея.

Таким образом, позиционные ошибки, приведенные в каталоге источников еРОЗИТА, адекватно описывают неопределенности в положениях рентгеновских источников, что обосновывает их использование при вычислении байесовых вероятностей отождествления источников в программе NWAY.

4.3. Точность и полнота оптического отождествления

Необходимо оценить качество нашей модели отождествления, сравнить ее с алгоритмом тривиального выбора ближайшего оптического объекта в качестве компаньона, а также откалибровать порог по параметру p_{any}. Для этого мы используем валидационную выборку и сравниваем компаньоны, подобранные нашей моделью, и настоящие компаньоны, найденные с использованием данных Chandra/XMM.

Для интерпретации результата работы модели отождествления мы применяем следующий алгоритм. Используя заранее выбранное значение порога $p_{\text{any},0}$, мы классифицируем источники еРОЗИТА с $p_{\text{any}} < p_{\text{any},0}$ как бездомные. Для источников с $p_{\text{any}} > p_{\text{any},0}$ в качестве компаньона выбираем объект с максимальным значением величины p_i . При этом возможны следующие исходы:

- Правильно определено, что источник имеет компаньон, и сам компаньон выбран правильно. Обозначим это число через *A*.
- Правильно определено, что источник имеет компаньон, но сам компаньон выбран неправильно (*B*). Это ошибка идентификации.
- Бездомный источник неправильно классифицирован как источник с компаньоном (*C*).
 Это ошибка классификации бездомного источника.
- Источник с компаньоном неправильно классифицирован как бездомный (D). Это ошибка классификации источника с компаньоном.
- Правильно определено, что источник является бездомным (*E*).

В приложении на рис. 5 приведены несколько примеров источников из валидационного каталога и их оптическое окружение. Отмечены положения рентгеновских источников (еРОЗИТА, Chandra, XMM) и оптические объекты. Мы также указываем параметры отождествления (p_{any} , p_i) и настоящий компаньон. Показаны три примера правильного и один пример неправильного отождествления.

В валидационной выборке $N_c = 529$ источников с компаньонами и $N_h = 30$ бездомных источников.

Мы задаем следующие метрики качества (Бельведерский и др., 2022):

- Общая метрика $Q = \frac{A+E}{N_c + N_h}.$
- Полнота отождествления компаньонов $C_c = \frac{A}{N_c}$.
- Полнота отождествления бездомных источников $C_h = \frac{E}{N_h}$.



Рис. 4. Метрики качества модели отождествления в зависимости от значения порога $p_{any,0}$. Сплошной толстой синей линией показано поведение метрики Q. Красные линии средней толщины соответствуют отбору компаньонов: сплошная линия — полнота C_c , штриховая — точность P_c . Тонкие зеленые линии показывают метрики для отбора бездомных источников: сплошная линия — полнота C_h , штриховая — точность P_h . (а) — Полная модель отождествления с использованием фотометрической и позиционной информации; серой линией нанесена метрика Q для случая отбора ближайшего партнера. (б) — Отождествление с ближайшим оптическим объектом; серой линией показана метрика Q фотометрической модель отождествления для ярких источников ($F > 10^{-14}$ эрг см⁻² с⁻¹); серой линией показана метрика Q фотометрической модели. полного каталога.

- Точность отождествления компаньонов $P_c = \frac{A}{A+B+C}$.
- Точность отождествления бездомных источников $P_h = \frac{E}{D+E}$.

На рис. 4 показано поведение вышеприведенных метрик для трех вариантов модели отождествления: для полной модели с использованием фотометрической и позиционной информации, для позиционной модели (в качестве компаньона выбирается ближайший оптический объект) и полной модели, примененной к ярким источникам с потоком $F_X > 10^{-14}$ эрг/с/см².

Точка пересечения кривых полноты и точности отбора компаньонов расположена на $p_{\rm any}=0.12$, при этом достигаются значения точности и полноты 94—95%. Для бездомных источников пересечение происходит при $p_{\rm any}\approx 0.15$ с точностью и полно-

774

Модель	Компаньон верный (А)	Компаньон неверный (В)	Компаньон назначен для бездомного (С)	Источник с компаньо- ном назначен бездомным (D)	Верно бездомный (Е)	Качество модели
NWAY + nnmag $p_{\text{any},0} = 0.12$	503	20	6	6	24	Q = 0.94
Ближайший партнер $p_{\mathrm{any},0}=0.24$	460	61	9	8	21	Q = 0.86
NWAY + пптад (яркие источники) $p_{\text{апу},0} = 0.11$	246	4	0	0	3	Q = 0.98

Таблица 2. Количество правильно или неправильно определенных компаньонов источников еРОЗИТА для трех вариантов модели отождествления: полная модель, отождествление с ближайшим оптическим источником; полная модель, примененная к ярким источникам валидационного каталога

той 80%. При этом на $p_{any} = 0.12$ общая точность модели отождествления достигает Q = 94%. Порог по p_{any} должен выбираться в зависимости от конкретной научной задачи, для которой производится оптическое отождествление. Для общей задачи классификации рентгеновских источников порог $p_{any,0} = 0.12$ является хорошим выбором.

В случае наивного отождествления рентгеновского источника с ближайшим оптическим объектом из каталога DESI LIS пересечение кривых отбора компаньонов происходит при $p_{any} \approx 0.24$ с точностью и полнотой 86—87%, и для отбора бездомных p_{any} такое же с точностью и полнотой 70%. Общая точность модели Q = 86%.

При отождествлении ярких источников (поток в диапазоне 0.5–2 кэВ $F_x > 10^{-14}$ эрг см⁻² с⁻¹) достигаются точность и полнота порядка 98% для $p_{\text{апу, 0}} \approx 0.12$. Отметим, что при таком фильтре по рентгеновскому потоку в валидационной выборке оказывается всего три бездомных источника, поэтому калибровка метрик для бездомных источников не производилась.

В табл. 2 показаны результаты оценки точности в соответствующих случаях.

Наблюдается сильное повышение качества отождествления при добавлении фотометрической информации — общее качество отождествления Q возрастает на 8 процентных пунктов. Однако интересно отметить, что отождествление с ближайшим оптическим источником предоставляет неплохую точность само по себе. Это связано с относительно хорошей точностью локализации рентгеновских источников телескопом еРОЗИТА и плотностью оптических объектов на уровне чувствительности обзора DESI LIS в области Дыры Локмана. Плотность объектов DESI в области Дыры Локмана соответствует одному объекту в кружке радиусом 10 угл. сек, что примерно является точностью

локализации источников СРГ/еРОЗИТА. Благодаря этому отождествление источников еРОЗИТА с ближайшим оптическим объектом во многих случаях приводит к правильному результату.

4.4. Дальнейшие шаги для обзора всего неба

Область Дыры Локмана наблюдалась телескопом еРОЗИТА в ходе верификационной фазы наблюдений обсерватории СРГ. Рентгеновская чувствительность достаточно равномерно распределена по площади обзора. Кроме того, площадь обзора не слишком велика, что позволяет игнорировать вариации межзвездного поглощения, а также глубины обзора DESI LIS. При распространении фотометрической модели на бо́льшие участки неба будет необходимо учесть эти эффекты. Обзор DESI LIS покрывает не все небо, и, вероятно, качестве фотометрического каталога можно использовать обзор Pan-STARRS⁸ или подобные.

В подразделе 3.1 описана достаточно гибкая процедура обучения фотометрической модели. Вопервых, в обучающих данных был сделан акцент на глубине рентгеновского обзора 3×10^{-15} эрг см⁻² с⁻¹, и обновление процедуры для применения в обзоре всего неба (где глубина обзора варьируется) достаточно просто и заключается в разбиении моделей по интервалам по рентгеновскому потоку. Во-вторых, упор в данной работе был сделан на внегалактических полях ($|b| > 20^\circ$), и использование новой обучающей выборки из области галактического диска позволит подойти к задаче отождествления рентгеновских источников в густонаселенных полях в Хребте нашей Галактики ($|b| < 20^\circ$).

⁸https://outerspace.stsci.edu/display/PANSTARRS/

Создание валидационного каталога по данным всего неба позволит точнее откалибровать ошибки положения рентгеновских источников (см. подраздел 4.2) за счет заметного увеличения размера выборки. Также, вероятно, можно будет учесть более тонкие эффекты, связанные с зависимостью калибровки позиционных ошибок от рентгеновского потока, достоверности детектирования (Бельведерский и др., 2022), положения на небе и т.д. Валидационный каталог большого размера поможет точнее оценить кривые полноты и чистоты алгоритма отождествления и его зависимость от параметров детектирования, в том числе от положения на небе.

5. ЗАКЛЮЧЕНИЕ

В статье построена модель оптического отождествления источников, открываемых телескопом еРОЗИТА, используя в качестве примера каталог рентгеновских источников еРОЗИТА в области Дыры Локмана, и оценили ее точность и полноту. В качестве источника оптических данных мы использовали данные обзора DESI LIS, дополненные инфракрасной фотометрией спутника WISE. Обучение и валидация моделей производились с использованием данных рентгеновских обсерваторий Chandra и XMM.

Представлена фотометрическая модель, основанная на использовании нейронных сетей для разделения фотометрических признаков оптических компаньонов рентгеновских источников от полевых источников. Модель преобразует всю совокупность доступных фотометрических данных в одно число, которое мы назвали пптад и которое варьируется от 0 до 1 и позволяет разделять эти два класса с точностью/полнотой 40—70% в зависимости от доступных данных.

Мы описываем построение выборки источников еРОЗИТА с надежно определенными оптическими компаньонами. Подобная выборка позволяет охарактеризовать точность и полноту любой модели отождествления, а так же при необходимости легко конструируется на бо́льших участках неба. В области Дыры Локмана это позволило точно найти оптический компаньон для 559 источников еРОЗИТА из 6885. Валидационная выборка позволила нам проверить точность определения ошибок положений рентгеновских источников, как описано в подразделе 4.2.

Мы применили связку из нейронной сети и кода NWAY для поиска оптических компаньонов источников СРГ/еРОЗИТА в области Дыры Локмана. Качество построенной модели отождествления обсуждается в подразделе 4.3. Построенная модель достигает полноты и чистоты (точности) отбора компаньонов 95% и отбора бесхозных источников 80%, что превосходит наивное отождествление по ближайшему партнеру на 8 процентных пунктов. Для ярких рентгеновских источников ($F_{0.5-2} > 10^{-14}$ эрг см⁻² с⁻¹) модель достигает точности 98%. Мы приводим примеры отождествлений, а также подробные кривые калибровок. В заключение мы обсудили необходимые шаги для применения модели в обзоре всего неба СРГ/еРОЗИТА.

Это исследование использовало результаты наблюдений телескопа еРОЗИТА на борту обсерватории СРГ. Обсерватория СРГ изготовлена Роскосмосом в интересах Российской академии наук в лице Института космических исследований (ИКИ) в рамках Российской федеральной научной программы с участием Германского центра авиации и космонавтики (DLR). Рентгеновский телескоп СРГ/еРОЗИТА изготовлен консорциумом германских институтов во главе с Институтом внеземной физики Общества им. Макса Планка (MPE) при поддержке DLR. Космический аппарат СРГ спроектирован, изготовлен, запущен и управляется НПО им. Лавочкина и его субподрядчиками. Прием научных данных осуществляется комплексом антенн дальней космической связи в Медвежьих озерах, Уссурийске и Байконуре и финансируется Роскосмосом. Использованные в настоящей работе данные телескопа еРОЗИТА обработаны с помощью программного обеспечения eSASS, разработанного германским консорциумом еРОЗИТА и программного обеспечения для обработки и анализа данных, разработанного российским консорциумом телескопа еРОЗИТА.

С. Д. Быков благодарит за поддержку международную исследовательскую школу по астрофизике общества Макса Планка (IMPRS) в Мюнхенском университете имени Людвига и Максимилиана. М.И. Бельведерский и М.Р. Гильфанов благодарят поддержку гранта РНФ 21-12-00343.

Программное обеспечение: NumPy (Харрис и др., 2020), Matplotlib (Хантер и др., 2007), SciPy (Виртанен и др., 2020), Pandas (Команда pandas, 2020), AstroPy (Коллаборация Astropy и др., 2018), HEALPix (Горски и др., 2005), HEALPy (Зонка и др., 2019).

ПРИЛОЖЕНИЕ



Рис. 5. Примеры изображений неба (по данным DESI LIS) в окрестности источников еРОЗИТА. Крестом и черным кругом отмечены положение источника еРОЗИТА и его ошибка положения r_{98} . Название источника и значение p_{any} указаны в заголовке. Синими маленькими кружками показаны положения объектов DESI LIS, снизу каждого кружочка подписан его идентификатор, сверху — значение p_i в процентах. В таблицах на врезках указаны кандидаты в компаньоны, у которых $p_i > 0.01$, их идентификаторы и расстояние до источника еРОЗИТА в угловых секундах. Положения источников Chandra(XMM) из валидационного каталога показаны красными (зелеными) кружками. Пунктирная линия показывает радиус $r_{false} = 1.47$ " вокруг Chandra/XMM. Настоящий компаньон указан комментарием (True). В случаях (a), (b), (c) отождествление компаньона верное, в случаях (d) — неверное, несмотря на условно большое значение $p_{any} > 0.12$.

СПИСОК ЛИТЕРАТУРЫ

- Ананна и др. (Т.Т. Ananna, M. Salvato, S. LaMassa, C.M. Urry, N. Cappelluti, C. Cardamone, et al.), Astrophys. J. 850, 66 (2017).
- Бельведерский и др. (M.I. Belvedersky, A.V. Meshcheryakov, M.R. Gilfanov, and P.S. Medvedev), arXiv e-prints, p. arXiv:2203.11531 (2022).
- 3. Бруса и др. (M. Brusa, G. Zamorani, A. Comastri, G. Hasinger, N. Cappelluti, F. Civano, et al.), Astrophys. J. Suppl. Ser. **172**, 353 (2007).
- 4. Брюннер и др. (Н. Brunner, T. Liu, G. Lamer, A. Georgakakis, A. Merloni, M. Brusa, et al.), arXiv e-prints, p. arXiv:2106.14517 (2021).
- 5. Будавари и Лоредо (Т. Budavári and T.J. Loredo), Annual Review of Statistics and Its Application **2**, 113 (2015).
- 6. Будавари и Сзалай (Т. Budavári and A.S. Szalay), Astrophys. J. **679**, 301 (2008).
- 7. Вебб и др. (N.A. Webb, M. Coriat, I. Traulsen, J. Ballet, C. Motch, F.J. Carrera, et al.), Astron. Astrophys. **641**, A136 (2020).
- 8. Виртанен и др. (P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al.), Nature Methods 17, 261 (2020).
- Горски и др. (К.М. Górski, Е. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke, et al.), Astrophys. J. 622, 759 (2005).
- 10. Дей и др. (A. Dey, D.J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan, et al.), Astron. J. **157**, 168 (2019).
- 11. Зонка и др. (A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, et al.), Journal of Open Source Software 4, 1298 (2019).
- Коллаборация Astropy и др. (Astropy Collaboration, A.M. Price-Whelan, B.M. Sip.ocz, H.M. Günther, P.L. Lim, S.M. Crawford, et al.), Astron. J. 156, 123 (2018).
- Колодзиг и др. (A. Kolodzig, M. Gilfanov, G. Hütsi, and R. Sunyaev), Astron. Astrophys. 558, A90 (2013a).
- 14. Колодзиг и др. (A. Kolodzig, M. Gilfanov, R. Sunyaev, S. Sazonov, and M. Brusa), Astron. Astrophys. **558**, A89 (2013b).
- 15. T. pandas development team, *pandas-dev/pandas*: *Pandas* (2020).
- 16. Локман и др. (F.J. Lockman, K. Jahoda, and D. McCammon), Astrophys. J. **302**, 432 (1986).
- 17. Луо и др. (B. Luo, W.N. Brandt, Y.Q. Xue, M. Brusa, D.M. Alexander, F. E. Bauer, et al.), Astrophys. J. Suppl. Ser. **187**, 560 (2010).

- 18. Луоидр. (B. Luo, W.N. Brandt, Y.Q. Xue, B. Lehmer, D.M. Alexander, F. E. Bauer, et al.), Astrophys. J. Suppl. Ser. **228**, 2 (2017).
- 19. Марчеси и др. (S. Marchesi, F. Civano, M. Elvis, M. Salvato, M. Brusa, A. Comastri, et al.), Astrophys. J. **817**, 34 (2016).
- 20. Мерлони и др. (A. Merloni, P. Predehl, W. Becker, H. Böhringer, T. Boller, H. Brunner, et al.), arXiv eprints, p. arXiv:1209.3114 (2012).
- 21. Нэйлор и др. (Т. Naylor, P.S. Broos, and E.D. Feigelson), Astrophys. J. Suppl. Ser. **209**, 30 (2013).
- 22. Павлинский и др. (M. Pavlinsky, A. Tkachenko, V. Levin, N. Alexandrovich, V. Arefiev, V. Babyshkin, et al.), Astron. Astrophys. **650**, A42 (2021).
- 23. Пиллепич и др. (A. Pillepich, C. Porciani, and T.H. Reiprich), MNRAS **422**, 44 (2012).
- 24. Пино и др. (F.X. Pineau, S. Derriere, C. Motch, F.J. Carrera, F. Genova, L. Michel, et al.), Astron. Astrophys., 597, A89 (2017).
- 25. Предель и др. (P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, et al.), Astron. Astrophys. **647**, A1 (2021).
- 26. Прокопенко и Гильфанов (I.G. Prokopenko and M.R. Gilfanov), Astronomy Letters **35**, 294 (2009).
- 27. Сазерленд и Сандерс (W. Sutherland and W. Saunders), MNRAS **259**, 413 (1992).
- 28. Сальвато и др. (M. Salvato, J. Buchner, T. Budavári, T. Dwelly, A. Merloni, M. Brusa, et al.), MNRAS **473**, 4937 (2018).
- 29. Сальвато и др. (M. Salvato, J. Wolf, T. Dwelly, A. Georgakakis, M. Brusa, A. Merloni, et al.), Astron. Astrophys. **661**, A3 (2022).
- 30. Сюняев и др. (R. Sunyaev, V. Arefiev, V. Babyshkin, A. Bogomolov, K. Borisov, M. Buntov, et al.), Astron. Astrophys. **656**, A132 (2021).
- 31. Уотсон и др. (M.G. Watson, A.C. Schröder, D. Fyfe, C.G. Page, G. Lamer, S. Mateos, et al.), Astron. Astrophys. **493**, 339 (2009).
- 32. Хантер и др. (J.D. Hunter), Computing in Science & Engineering 9, 90 (2007).
- 33. Харрис и др. (C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al.), Nature **585**, 357 (2020).
- 34. Чен и др. (С.Т.J. Chen, W.N. Brandt, B. Luo, P. Ranalli, G. Yang, D. M. Alexander, et al.), MNRAS 478, 2132 (2018).
- 35. Эванс и др. (I.N. Evans, F.A. Primini, K.J. Glotfelty, C.S. Anderson, N.R. Bonaventura, J.C. Chen, et al.), Astrophys. J. Suppl. Ser. **189**, 37 (2010).