

ИНФОРМАЦИОННЫЕ  
ТЕХНОЛОГИИ  
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION  
TECHNOLOGIES  
AND TELECOMMUNICATIONS

УДК 004.8

DOI: 10.17212/2782-2001-2026-2-7-30

## Исследование алгоритмов глубокой кластеризации с контролируемой энтропией мягких присваиваний\*

А.А. АЛЕТДИНОВА<sup>1,a</sup>, Е.Н. АНТОНЯНЦ<sup>1,b</sup>, Е.И. СБОРЩИКОВ<sup>2,c</sup>

<sup>1</sup> РФ, 630073, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

<sup>2</sup> РФ, 630091, г. Новосибирск, Красный пр., 72, Новосибирский авиационный технический колледж им. Б.С. Галуцака

<sup>a</sup> [aletdinova@corp.nstu.ru](mailto:aletdinova@corp.nstu.ru)    <sup>b</sup> [bax201438@gmail.com](mailto:bax201438@gmail.com)

<sup>c</sup> [egorsborsikov250575@gmail.com](mailto:egorsborsikov250575@gmail.com)

В статье исследуется проблема преждевременного коллапса кластерных присваиваний в современных алгоритмах глубокой кластеризации, таких как глубокое встраиваемое кластерирование (DEC) и вариационное глубокое вложение (VaDE). Данное явление приводит к детерминированным, а не вероятностным распределениям принадлежности объектов, что противоречит природе реальных данных, где объекты обладают смешанной принадлежностью к нескольким кластерам (например, пользователи цифровых сервисов, биологические образцы с нечеткой классификацией, образцы в материаловедении). Для решения этой проблемы предложена модифицированная архитектура автоэнкодера на основе VaDE, в целевую функцию которой введен механизм энтропийной регуляризации, контролирующей баланс между качеством кластеризации и степенью мягкости (неопределенности) присваиваний. Используется прямое адаптивное сохранение высокой энтропии распределений мягких присваиваний на протяжении всего процесса обучения модели, что обеспечивает вероятностную интерпретацию результатов. Эксперименты на модернизированном наборе данных от Национального института стандартов и технологий (MNIST) демонстрируют оптимальное значение параметра регуляризации, равного 0,05, что позволяет повысить нормализованную взаимную информацию (NMI) на 61,6 % по сравнению с базовой моделью без регуляризации. При этом сохраняется способность модели нейронной сети к интерпретации через генерацию семантически значимых прототипов кластеров. Также модель имеет практическую значимость в таких прикладных областях, как маркетинговая сегментация и анализ поведенческих данных. Такой подход устраняет разрыв между алгоритмической эффективностью и потребностями прикладных специалистов в объяснимых и достоверных результатах кластеризации.

---

\* Статья получена 28 января 2026 г.

**Ключевые слова:** глубокая кластеризация, вариационные автоэнкодеры, мягкие присваивания, энтропийная регуляризация, интерпретируемость, обучение без учителя, неопределенность, генеративные модели, оптимизация компромиссов

## ВВЕДЕНИЕ

Кластеризация как одна из задач машинного обучения без учителя ставит своей целью выявление скрытой структуры и группировку объектов в неразмеченных данных на основе меры сходства. В современных условиях возникает рост сложности данных – от многомерных поведенческих профилей в цифровых экосистемах до высокоразмерных геномных и протеомных последовательностей. Традиционные методы кластеризации демонстрируют свою ограниченность на таких наборах данных.

Под алгоритмом К-средних (K-means) понимается классический метод кластеризации, минимизирующий сумму квадратов расстояний от объектов до центроидов кластеров. Алгоритм максимизации ожидания (Expectation-Maximization algorithm, EM) является итеративным методом оценки параметров вероятностных моделей с неполными данными, применяемыми в гауссовых смесях.

Такие алгоритмы опираются на ряд упрощающих предположений (линейная разделимость, сферичность кластеров, гомогенность дисперсий), которые редко выполняются в реальных приложениях, а их неспособность автоматически извлекать релевантные нелинейные признаки делает их неприменимыми к сырым, высокоразмерным данным [1].

Прорывное развитие глубокого обучения открыло новую эпоху в кластеризации и дало начало направлению глубокой кластеризации. Эти методы объединяют нейросетевые представления, способные улавливать сложные иерархические зависимости с формализмом кластерного анализа. Наиболее значимыми работами в этой области стали DEC (Deep Embedded Clustering) [2] и его генеративное расширение – VaDE (Variational Deep Embedding) [3]. DEC предлагает двухэтапный подход: предобучение автоэнкодера для получения представлений и последующая итеративная «заточка» кластерных распределений через минимизацию дивергенции Кульбака – Лейблера (Kullback – Leibler divergence, KL). VaDE интегрировал эту идею в байесовское распределение, используя вариационный автоэнкодер с априорными данными в виде смеси гауссиан, что формально обеспечило бы вероятностную трактовку и генеративные возможности.

Однако, как отмечается в обзорах [1, 4], эти методы имеют систематический недостаток – неконтролируемое снижение энтропии распределений мягких присваиваний в процессе обучения (явление, известное как коллапс присваиваний).

Под мягкими присваиваниями понимается вероятностное распределение принадлежности объекта к каждому из кластеров, в отличие от жестких (детерминированных) меток. Энтропия мягких присваиваний, в свою очередь, количественно измеряет степень неопределенности: чем выше энтропия, тем «мягче» присваивание [4].

Механизм итеративного уточнения в DEC и оптимизация вариационной нижней границы в VaDE неявно максимизирует уверенность модели, приводя к быстрому коллапсу апостериорных распределений в почти детерминированные векторы. Это не только противоречит пониманию о смешанной принадлежности объектов, но и снижает устойчивость модели, ее способность к обоб-

щению результатов. В прикладных контекстах, таких как сегментация клиентов или анализ транскриптомных данных, подобная излишняя уверенность может вводить исследователей в заблуждение, скрывая присущую данным неопределенность.

Таким образом, возникает ключевая проблема в необходимости разработки механизма, который бы обеспечивал устойчивый контроль над энтропией кластерных присваиваний в рамках генеративных моделей глубокой кластеризации, не жертвуя при этом дискриминативной силой и интерпретируемостью.

Целью настоящей работы является решение проблемы коллапса кластерных присваиваний в генеративных моделях глубокой кластеризации. В работе предлагается модифицированная архитектура на основе VaDE, которая за счет введения явной энтропийной регуляризации в целевую функцию позволяет гибко контролировать баланс между четкостью выделяемых кластеров и сохранением неопределенности присваиваний. Это необходимо для получения более адекватных и интерпретируемых результатов в задачах, где объекты могут принадлежать сразу к нескольким группам.

Для достижения данной цели в работе решаются следующие задачи.

1. Проведение критического анализа существующих методов глубокой кластеризации с выявлением их ограничений, связанных с контролем энтропии.
2. Предложение модификации целевой функции VaDE, включающей параметризованный член энтропийной регуляризации.
3. Проектирование, разработка и обучение соответствующей архитектуры нейронной сети.
4. Экспериментальное исследование влияния параметра регуляризации на ключевые метрики качества и на энтропию распределений присваиваний.
5. Анализ интерпретируемости полученных кластеров и предложение практических рекомендаций по выбору параметра регуляризации для различных прикладных сценариев.

Объектами исследования являются алгоритмы глубокой кластеризации, а предметом исследования – механизм контроля энтропии мягких присваиваний в алгоритмах. Эксперименты проводятся на модернизированном наборе данных от Национального института стандартов и технологий (Modified National Institute of Standards and Technology, MNIST), который в этом случае выступает в роли инструмента для верификации предложенного метода.

Научная новизна работы заключается в предложении явного и параметризованного механизма энтропийной регуляризации, напрямую воздействующего на распределение мягких присваиваний, теоретическое обоснование и экспериментальное подтверждение эффекта повышения качества кластеризации при умеренных значениях регуляризации. Это свидетельствует о предотвращении переобучения и выходе из локальных минимумов, а также о введении и исследовании параметра регуляризации как инструмента управления фундаментальным компромиссом между дискриминативностью и неопределенностью в задачах кластеризации. Практическая значимость заключается в применимости модели в предметных областях, где критична интерпретация неопределенности (сегментация клиентов с несколькими интересами, анализ биомедицинских данных с нечеткими диагнозами, рекомендательные

системы). Для обоснования актуальности предложенного подхода в следующем разделе проводится анализ существующих методов кластеризации, выявляющий их ограничения в контроле энтропии и интерпретируемости.

## 1. АНАЛИЗ СУЩЕСТВУЮЩИХ АЛГОРИТМОВ КЛАСТЕРИЗАЦИИ ОБЪЕКТОВ

### 1.1. КЛАССИЧЕСКИЕ АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ

Исторически первыми подходами к кластеризации стали алгоритмы, основанные на метриках расстояния в признаковом пространстве. Наиболее известным из них является алгоритм K-means, который минимизирует внутрикластерную сумму квадратов отклонений. Несмотря на вычислительную эффективность, этот метод опирается на ряд предположений: кластеры должны быть компактными, сферически симметричными и иметь сопоставимую дисперсию. Более того, алгоритм K-means реализует жесткую кластеризацию, где каждый объект принадлежит только одному кластеру, что игнорирует явление перекрытия групп, типичное для реальных данных (например, пользователь может одновременно интересоваться спортом и искусством).

Вероятностным расширением этого подхода является метод гауссовых смесей (Gaussian Mixture Method, GMM), использующий EM-алгоритм для оценки параметров компонентов смеси и предоставляющий мягкие присваивания в виде апостериорных вероятностей, вычисляемых по формуле Байеса. Однако и этот метод имеет существенные ограничения: данные в каждом кластере распределены согласно многомерному нормальному закону, что редко выполняется на практике. Кроме того, как отмечают специалисты, мягкие присваивания в GMM лишены прямой семантической интерпретируемости: параметры модели (средние и ковариации) не связаны с исходными признаками и не позволяют объяснить, почему объект отнесен к тому или иному кластеру [1].

Кроме того, в прикладных исследованиях, таких как анализ экономических условий функционирования агропромышленного комплекса [5], применяются иерархические методы кластеризации (например, метод Варда) и методы, основанные на плотности (например, основанная на плотности пространственная кластеризация для приложений с шумами – Density-based Spatial Clustering of Applications with Noise, DBSCAN), что подчеркивает многообразие подходов в зависимости от предметной области. Однако эти методы также предполагают определенные формы кластеров и не масштабируются в высокоразмерные данные.

Перечисленные классические методы работают напрямую в пространстве наблюдений, что делает их неспособными выявлять сложные нелинейные зависимости и требует ручной предварительной обработки данных (нормализация, отбор признаков, понижение размерности), поэтому их нельзя применять к современным высокоразмерным и гетерогенным данным, таким как изображения, тексты или пользовательские поведенческие спецификации.

## 1.2. АЛГОРИТМЫ ГЛУБОКОЙ КЛАСТЕРИЗАЦИИ

С появлением глубоких нейронных сетей открылась новая эра в кластеризации, что позволило автоматически извлекать релевантные низкоразмерные представления из сырых данных. Ключевой работой в этой области стала модель DEC [2]. В ней используется автоэнкодер для нелинейного преобразования данных, после чего в скрытом пространстве применяется механизм итеративного уточнения целевого распределения  $t$ , минимизирующий KL-дивергенцию:

$$KL(t \parallel \gamma) = \sum_{i=1}^N \sum_{j=1}^K t_{ik} \ln \left( \frac{t_{ik}}{\gamma_{ik}} \right), \quad (1)$$

где  $N$  – количество объектов в выборке;  $K$  – число кластеров;  $\gamma_{ik}$  – вероятность принадлежности  $i$ -го объекта к  $K$ -му кластеру;  $t_{ik}$  – целевое распределение, получаемое экспоненцированием и последующей нормализацией исходных оценок принадлежности для повышения уверенности модели.

Несмотря на высокую эмпирическую эффективность, DEC имеет фундаментальный недостаток: оптимизация через KL-дивергенцию неявно максимизирует уверенность модели, что приводит к быстрому коллапсу энтропии распределений  $\gamma$  уже на ранних этапах обучения [4, 6]. Это делает присваивания практически детерминированными, что противоречит природе данных со смешанной принадлежностью.

Попыткой преодолеть эти ограничения стала модель VaDE [3], интегрировавшая DEC в рамки вариационного вывода. VaDE использует вариационный автоэнкодер (Variational Autoencoder, VAE) с априорным распределением в виде гауссовой смеси, что формально обеспечивает вероятностную интерпретацию и генеративные возможности. Однако, как показано в обзоре [1], VaDE сохраняет ту же целевую функцию, что и в DEC, и, следовательно, также склонен к подавлению энтропии. Более того, использование сложных априорных распределений в VAE порождает дополнительную проблему – постериорный коллапс, когда апостериорное распределение вырождается в априорное и энкодер перестает использовать входные данные [7]. Эта проблема возникает, поскольку баланс между реконструкцией и регуляризацией в VAE является деликатным и требует тщательной настройки.

## 1.3. СОВРЕМЕННЫЕ ТЕНДЕНЦИИ И ПРОБЛЕМЫ В ОБЛАСТИ ГЛУБОКОЙ КЛАСТЕРИЗАЦИИ

Современные исследователи стремятся решить эти проблемы несколькими путями. Во-первых, предлагается явная энтропийная регуляризация для сохранения неопределенности в присваиваниях, как, например, в работе [4], однако этот подход не был интегрирован в генеративные архитектуры. Во-вторых, для борьбы с коллапсом исследуются сложные априорные распределения: сферические [8], запутанные смеси [9] и гамма-смеси [10]. Хотя результаты этих исследований повышают дискриминативность скрытого пространства, они не решают проблему жесткости присваиваний напрямую.

Немаловажным направлением является интерпретируемость данных. Работы [11, 12] связывают скрытые переменные с входными признаками или предметными концепциями, но их фокус ограничен табличными данными или специфическими архитектурами. Методы на основе диффузионных VAE [13] открывают новые горизонты, но остаются вычислительно дорогими и недостаточно исследованными в контексте управления мягкостью присваиваний.

Таким образом, несмотря на значительный прогресс, в области глубокой кластеризации сохраняются три ключевые проблемы:

- 1) недостаточный контроль над энтропией мягких присваиваний;
- 2) компромисс между качеством кластеризации и интерпретируемостью;
- 3) отсутствие универсального механизма регуляризации мягкости.

Для преодоления этих ограничений в следующем разделе предлагается модифицированная архитектура на основе VaDE с явной энтропийной регуляризацией, обеспечивающая управляемый контроль над мягкостью кластерных присваиваний.

## 2. РЕАЛИЗАЦИЯ И ОБУЧЕНИЕ МОДЕЛИ НЕЙРОСЕТЕВОЙ КЛАСТЕРИЗАЦИИ С МЯГКИМИ ПРИСВАИВАНИЯМИ

Предлагаемая архитектура модели основана на модификации VaDE, включающая энкодер (компонент, преобразующий поток входных данных в компактное векторное представление), декодер (компонент, восстанавливающий данные) и априорное распределение. В этом разделе приняты следующие обозначения:

- 1)  $x \in \mathbb{R}^{\mathcal{D}}$  – входной вектор признаков размерности  $\mathcal{D}$ ;
- 2)  $z \in \mathbb{R}^d$  – скрытое представление объекта, где  $d \ll \mathcal{D}$  – размерность скрытого пространства;
- 3)  $\gamma \in \Delta^{(K-1)} \subset \mathbb{R}^K$  – вектор мягких присваиваний к кластерам, где  $K$  – заданное число кластеров, а векторный симплекс  $\Delta^{K-1}$  имеет вид

$$\Delta^{K-1} = \left\{ \gamma \in \mathbb{R}^K \mid \gamma_k \geq 0, \sum_{k=1}^K \gamma_k = 1 \right\}, \quad (2)$$

где условие  $\gamma_k \geq 0$  для всех  $k$  и нормировка  $\sum_{k=1}^K \gamma_k = 1$  гарантируют, что вектор  $\gamma = (\gamma_1, \dots, \gamma_K)$  представляет собой корректное дискретное вероятностное распределение. Каждая компонента  $\gamma_k$  интерпретируется как вероятность принадлежности объекта к кластеру  $K$ , а сумма этих вероятностей по всем кластерам равна единице, что соответствует полной группе событий. Таким образом, пространство  $\Delta^{K-1}$  задает множество всех возможных векторов мягких присваиваний для одного объекта;

- 4)  $\phi$  – параметры энкодера;
- 5)  $\theta$  – параметры декодера;

- 6)  $\eta \geq 0$  – коэффициент энтропийной регуляризации, управляющий степенью мягкости распределений  $\gamma$ ;  
 7)  $\beta \geq 0$  – коэффициент взвешивания регуляризационного члена на основе KL-дивергенции.

### 2.1. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ПРЕДЛАГАЕМОЙ АРХИТЕКТУРЫ НЕЙРОННОЙ СЕТИ

Энкодер  $q_\phi(z, \gamma | x)$  совместно моделирует апостериорное распределение скрытой переменной  $z$  и вектора присваиваний  $\gamma$ . В отличие от оригинального VaDE, где  $\gamma$  вычисляется через  $z$ , в предлагаемой архитектуре вводится параллельная ветвь для прямого предсказания  $\gamma$ . Предполагается условная независимость этих величин, что позволяет факторизовать апостериорное совместное распределение, выдаваемое энкодером:

$$q_\phi(z, \gamma | x) = q_\phi(z | x) \cdot q_\phi(\gamma | x). \quad (3)$$

Распределение скрытого представления задается как многомерное нормальное распределение с диагональной ковариационной матрицей по формуле

$$q_\phi(z | x) = \mathcal{N}\left(z | \mu_\phi(x), \text{diag}\left(\sigma_\phi^2(x)\right)\right), \quad (4)$$

где  $\mathcal{N}$  – нормальное (гауссово) распределение  $\mu_\phi(x) \in \mathbb{R}^d$  и  $\sigma_\phi^2(x) \in \mathbb{R}^d > 0$  – параметры, генерируемые нейронной сетью с весами  $\phi$ .

Распределение мягких присваиваний моделируется категориальным распределением по формуле

$$q_\phi(\gamma | x) = \text{Cat}(\gamma | \pi_\phi(x)), \quad (5)$$

где  $\text{Cat}$  – категориальное распределение; вероятности принадлежности к  $K$  кластерам вычисляются через нормализованное экспоненциальное преобразование, или softmax-преобразование, с температурой по формуле

$$\pi_\phi(x) = \text{softmax}\left(\frac{f_\phi(x)}{T}\right), \quad (6)$$

где  $f_\phi(x) \in \mathbb{R}^K$  – логиты (выходы полносвязного слоя), а  $T = 0,3$  – температурный параметр, контролирующий «остроту» распределения: меньшие значения  $T$  увеличивают уверенность модели, приближая распределение к детерминированному.

Декодер модели  $p_\theta(x | z)$  задает условное распределение восстановления данных при заданном скрытом представлении

$$p_\theta(x | z) = \mathcal{N}\left(x | \mu_\theta(z), \sigma_{data}^2 \times I\right), \quad (7)$$

где  $\mu_\theta(z) \in \mathbb{R}^D$  – детерминированный выход нейронной сети декодера с параметрами  $\theta$ , а  $\sigma_{data}^2$  – фиксированная дисперсия наблюдений. В случае бинарных

данных полагается  $\sigma_{data}^2 = 1$ , что соответствует минимизации среднеквадратичной ошибки реконструкции, поскольку это важно для воспроизводимости. Для непрерывных данных с неизвестной шумовой структурой параметр  $\sigma_{data}^2$  может быть дополнительно введен в обучаемый скаляр или вектор.

Априорное распределение  $p(z)$  над скрытым пространством задается в виде смеси  $K$  многомерных нормальных распределений

$$p(z) = \sum_{k=1}^K \pi_k \mathcal{N}\left(z | \mu_k, \text{diag}(\sigma_k^2)\right), \quad (8)$$

где  $\pi_k \geq 0$ ,  $\sum_{k=1}^K \pi_k = 1$  – априорные веса компонентов смеси,  $\mu_k \in \mathbb{R}^d$  и  $\sigma_k^2 \in \mathbb{R}^d > 0$  – векторы центров и дисперсий  $k$ -го кластера соответственно.

Все параметры смеси  $\{\pi_k, \mu_k, \sigma_k^2\}_{k=1}^K$  являются обучаемыми. Их инициализация осуществляется следующим образом: после предобучения автоэнкодера на задаче реконструкции входные объекты проецируются в скрытое пространство с помощью энкодера, и к полученным представлениям применяется алгоритм K-means. Результаты кластеризации используются для задания начальных значений  $\mu_k$  (центроиды) и  $\sigma_k^2$  (усредненные внутрикластерные дисперсии), а веса инициализируются равномерно:  $\pi_k = \frac{1}{K}$ . Такой подход обеспечивает согласованность между структурой априорного распределения и предполагаемой кластерной организацией данных.

Целевая функция  $\mathcal{L}(\phi, \theta)$  модели формируется как задача минимизации скалярного функционала и строится на основе вариационной нижней границы логарифма правдоподобия (Evidence Lower Bound, ELBO), расширенной за счет явного члена энтропийной регуляризации. Такая функция  $\mathcal{L}(\phi, \theta)$  состоит из трех слагаемых: 1) реконструкции для обеспечения точности восстановления входных данных, 2) регуляризации пространства  $z$ , которая согласует апостериорное распределение со сложным априорным в виде гауссовой смеси, и 3) регуляризации энтропии, которая прямо управляет степенью неопределенности в мягких присваиваниях

$$\mathcal{L}_{recon} = -\mathbb{E}_{q_\phi(z|x)} [\ln p_\theta(x|z)] \approx -\frac{1}{L} \sum_{l=1}^L \ln p_\theta(x|z^{(l)}), \quad (9)$$

где  $\mathbb{E}$  – оператор математического ожидания,  $z^l \sim q_\phi(z|x)$ . Этот член заставляет энкодер сохранять релевантную информацию о входе в скрытом представлении

$$\mathcal{L}_{KL} = KL(q_\phi(z|x) \| p(z)) = \int q_\phi(z|x) \ln \frac{q_\phi(z|x)}{p(z)} dz, \quad (10)$$

где  $q_\phi(z|x)$  – апостериорное распределение скрытой переменной  $z$  (энкодер),  $p(z)$  – априорное распределение скрытой переменной (обычно  $\mathcal{N}(0, I)$ ),  $KL$ -дивергенция. Натуральный логарифм используется для линеаризации выражения.

Благодаря структуре  $p(z)$ , вычисляемой по формуле (8), этот член неявно привязывает скрытые векторы к кластерной структуре и способствует формированию разделимых групп в скрытом пространстве

$$\mathcal{L}_{entropy} = -H(q_\phi(\gamma|x)) = \sum_{k=1}^K \gamma_k \ln \gamma_k, \quad (11)$$

где  $H(\cdot)$  – энтропия Шеннона,  $\gamma = (\gamma_1, \dots, \gamma_K) \in \Delta^{K-1}$  – вектор мягких присваиваний. Этот член поощряет высокую энтропию, предотвращая преждевременное «затвердевание» распределений.

Ключевым моментом является то, что при минимизации  $\mathcal{L}$  введение отрицательной энтропии приводит к максимизации энтропии  $H(\gamma)$ , тем самым сохраняя высокую степень неопределенности (мягкость) в кластерных присваиваниях и предотвращая преждевременное «затвердевание» распределений.

Параметр  $\eta \geq 0$  позволяет контролировать степень мягкости кластеризации:

- 1)  $\eta = 0$  – модель вырождается в стандартный VaDE;
- 2)  $\eta > 0$  – модель соответствует высокой энтропии присваиваний;
- 3) большие значения  $\eta$  соответствуют более равномерным распределениям  $\gamma$ , что может быть полезно в задачах с выраженной смешанной принадлежностью.

Таким образом, формула целевой функции принимает следующий вид:

$$\mathcal{L}(\phi, \theta) = \mathcal{L}_{recon} + \beta \mathcal{L}_{KL} + \eta \mathcal{L}_{entropy}. \quad (12)$$

Значение коэффициента  $\beta$  подбирается для обеспечения достаточного «давления» априорного распределения на формирование четкой кластерной структуры в скрытом пространстве.

Предложенная архитектура реализует единый вычислительный граф, в котором входной вектор  $x \in \mathbb{R}^D$  преобразуется энкодером в скрытое представление  $z \in \mathbb{R}^d$  и в вектор мягких присваиваний  $\gamma \in \Delta^{(K-1)}$ . Декодер восстанавливает вход по  $z$ , формируя реконструкцию  $\hat{x}$ . Целевая функция объединяет реконструкцию, регуляризацию и энтропийную компоненту. Финальными выходами модели для инференса (использование знаний для предсказания или оценки результатов на новых данных) являются  $\gamma$  и  $z$ , используемые для кластеризации и анализа. Структура предложенной модели нейронной сети представлена на рис. 1.

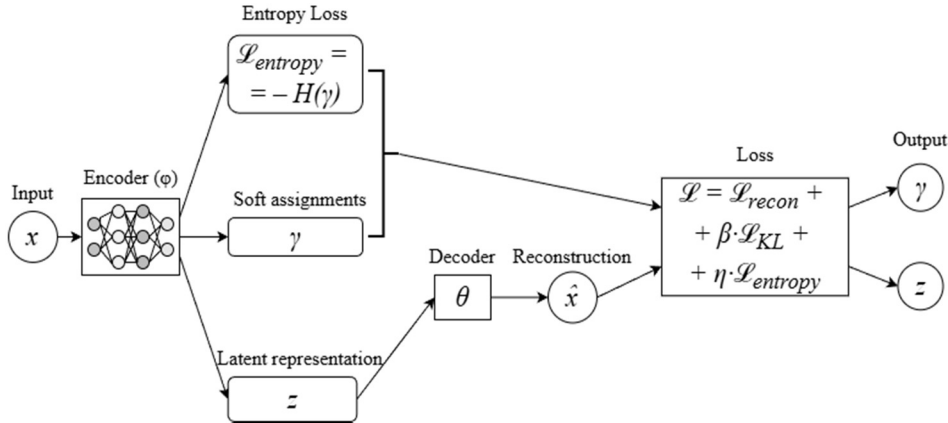


Рис. 1. Архитектура предложенной нейронной сети глубокой кластеризации

Fig. 1. Architecture of the proposed deep clustering neural network

Математическая модель нейронной сети глубокой кластеризации адаптивна под кодирование в инструментальных средах разработки, что требуется для ее обучения мягкой кластеризации объектов.

## 2.2. ПРОЦЕДУРА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Обучение предложенной модели осуществляется в два этапа: предобучение автоэнкодера для инициализации качественных скрытых представлений и основной этап совместной оптимизации всех компонент.

На этапе предобучения автоэнкодера происходит инициализация параметров энкодера и декодера для обеспечения реконструкции без учета кластерной структуры. Последовательность этапа выглядит следующим образом.

1. Инициализация: параметры  $\phi$  и  $\theta$  инициализируются согласно схеме Xavier (сохранение одинаковой дисперсии на входе и выходе каждого слоя для функции активации).

2. Упрощенная функция потерь: используется только реконструктивный член (без компонентов (9) и (10)).

3. Оптимизация: применяется оптимизатор Adam со скоростью обучения  $10^{-3}$  в течение трех-пяти эпох. Такого короткого периода достаточно для инициализации весов перед основным этапом обучения, что подтверждается стабильностью последующей сходимости.

4. Результат: получены начальные приближения для  $\mu_\phi(x)$ ,  $\sigma_\phi^2(x)$  и декодера, способные сохранять основную информацию о данных.

Далее происходит инициализация параметров GMM. Параметры априорной смеси гауссиан  $\{\pi_k, \mu_k, \sigma_k^2\}_{k=1}^K$  не могут быть инициализированы случайно, так как это приводит к нестабильному обучению. По предобученному энкодеру для всего обучающего множества вычисляются скрытые представления  $\mu_\phi(x_i)$ . На полученных векторах  $\mu_\phi(x_i)$  применяется алгоритм K-means с  $K$  кластерами для получения начальных меток.

Оценка параметров происходит следующим образом.

1. Центры  $\mu_k$  инициализируются как центроиды кластеров от K-means.
2. Дисперсии  $\sigma_k^2$  вычисляются как выборочные дисперсии точек внутри кластера с добавлением небольшого шума для устойчивости.
3. Веса  $\pi_k$  устанавливаются пропорционально размерам кластеров со сглаживанием Лапласа:

$$\pi_k = \frac{N_k + \alpha}{N + K\alpha}, \quad (13)$$

где  $N_k$  – число объектов, отнесенных к  $k$ -му кластеру на этапе инициализации,  $N = \sum_{k=1}^K N_k$  – общее число объектов,  $k$  – число кластеров,  $\alpha > 0$  – параметр сглаживания (в экспериментах используется  $\alpha = 1$ , что соответствует классическому сглаживанию Лапласа).

Алгоритм обучения реализует следующую логику:

1. Инициализация: загрузка параметров, полученных на этапах предобучения автоэнкодера и инициализации параметров GMM.
2. Итеративное обучение: для каждой эпохи и каждого батча данных выполняются шаги прямого и обратного распространения.
3. Вычисление компонент функции потерь: на каждом шаге независимо вычисляются три компоненты целевой функции.
4. Агрегация и оптимизация: компоненты суммируются с соответствующими весами, после чего вычисляются градиенты и обновляются параметры.

На основном этапе обучения оптимизируется полная целевая функция  $\mathcal{L}(\phi, \theta)$ .

Для обеспечения численной устойчивости и эффективности обучения в алгоритме применяются такие технические приемы, как метод репараметризации, вычисление  $KL$ -дивергенции для смеси гауссиан, регуляризация градиентов, температурный параметр в softmax, небольшие батчи и перемешивание данных, оптимизатор и настройка скорости обучения.

Метод репараметризации – стандартная техника в вариационном выводе, обеспечивающая дифференцируемость по параметрам распределения  $q_\phi(z|x)$ , вычисляется по формуле

$$z = \mu_\phi(x) + \epsilon \odot \exp\left(0,5 \ln \sigma_\phi^2(x)\right), \quad (14)$$

где  $\epsilon \sim \mathcal{N}(0, I)$ , а операция  $\odot$  обозначает поэлементное умножение.

Вычисление  $KL$ -дивергенции для смеси гауссиан используется как аппроксимация, осуществляемая через стабилизированный логарифм суммы экспонент. Этот прием используется, поскольку аналитическое вычисление  $\mathcal{L}_{KL} = KL(q_\phi(z|x) || p(z))$  для смеси гауссиан затруднительно. Такая аппроксимация вычисляется по формуле

$$\mathcal{L}_{KL} \approx -\ln \sum_{k=1}^K \pi_k \exp\left(-\frac{1}{2} \left[ \ln \frac{|\Sigma_k|}{|\Sigma_z|} - d + \text{tr}(\Sigma_k^{-1} \Sigma_z) + (\mu_z - \mu_k)^T \Sigma_k^{-1} (\mu_z - \mu_k) \right]\right), \quad (15)$$

где  $\Sigma_z = \text{diag}(\sigma_\phi^2(x))$  – диагональная ковариационная матрица апостериорного распределения скрытой переменной,  $\Sigma_k = \text{diag}(\sigma_k^2)$  – диагональная ковариационная матрица  $k$ -го компонента априорной смеси, параметр  $d$  является размерностью скрытого пространства,  $\text{tr}(\cdot)$  обозначает след матрицы, а  $\det|\cdot|$  – ее определитель. Выражение в квадратных скобках представляет собой аналитическую форму для  $KL$ -дивергенции между двумя многомерными нормальными распределениями: апостериорным  $q_\phi(z|x)$  и  $k$ -й компонентой априорной смеси  $p(z)$ . Логарифм суммы экспонент с весами  $\pi_k$  обеспечивает корректное усреднение по всем компонентам смеси.

Регуляризация градиентов используется для предотвращения взрыва градиентов, особенно на начальных этапах обучения. Для этого применяется их обрезка с максимальной нормой 5,0. Условие применения обрезки описано в системе

$$\nabla_\theta \mathcal{L} \leftarrow \begin{cases} \nabla_\theta \mathcal{L}, & \|\nabla_\theta \mathcal{L}\|_2 \leq 5,0, \\ 5,0 \frac{\nabla_\theta \mathcal{L}}{\|\nabla_\theta \mathcal{L}\|_2}, & \|\nabla_\theta \mathcal{L}\|_2 > 5,0, \end{cases} \quad (16)$$

где  $\nabla_\theta \mathcal{L}$  – градиент целевой функции  $\mathcal{L}$  по параметру  $\theta$ ,  $\|\cdot\|_2$  – евклидова норма вектора, значение 5,0 выбрано в качестве порога обрезки для предотвращения взрыва градиентов.

Важно отметить, что предложенная энтропийная регуляризация направлена на решение проблемы коллапса кластерных присваиваний, тогда как метод обратных липшицевых ограничений [14] разработан для предотвращения постериорного коллапса в вариационных автоэнкодерах – явления, при котором апостериорное распределение скрытой переменной перестает зависеть от входных данных и вырождается в априорное распределение, что приводит к потере информативности скрытого пространства. Это ортогональные механизмы для решения разных задач. В настоящей работе не применяется такой механизм, поскольку, во-первых, это потребовало бы существенного усложнения архитектуры, а во-вторых, как показано в разделе экспериментальных исследований, предложенный подход обеспечивает устойчивое обучение и высокое качество кластеризации без дополнительных ограничений на декодер.

Для управления «остротой» распределения вероятностей используется модификация функции softmax с добавлением температурного параметра  $T = 0,3$ . В этом случае для расчета  $\gamma$  применяется формула

$$\gamma_k = \frac{\exp(f_k(x)/T)}{\sum_{j=1}^K \exp(f_j(x)/T)}, \quad (17)$$

где  $f_k(x)$  – выход полносвязного слоя  $k$ -го кластера.

Небольшие батчи используются в обучении и имеют размер 256 объектов. Перед каждой эпохой данные случайным образом перемешиваются для улучшения сходимости.

Оптимизатор Adam используется с параметрами  $\beta_1$  и  $\beta_2$ ,  $\epsilon = 10^{-8}$ . Начальная скорость обучения устанавливается равной  $10^{-4}$ . Это величина, определяющая величину шага, на который корректируются веса модели в направлении антиградиента. Выбор целевой функции также играет ключевую роль в качестве кластеризации. Альтернативные подходы, такие как комбинированные VAE с функцией потерь на основе сходства (similarity-based loss) [15], которая дополнительно минимизирует расстояние между представлениями схожих объектов в скрытом пространстве, демонстрируют перспективность гибридных методов. Однако в настоящей работе мы фокусируемся на энтропийной регуляризации как основном механизме контроля мягкости присваиваний.

Обучение продолжается до достижения сходимости или максимального числа эпох (в экспериментах 15 эпох). В табл. 1 показаны критерии для мониторинга сходимости и предотвращения переобучения.

Таблица 1

Table 1

**Критерии мониторинга сходимости и предотвращения переобучения модели нейронной сети**

**Criteria for monitoring convergence and preventing neural network model overfitting**

Критерии		Назначение
Метрики качества кластеризации на валидационной выборке	NMI	Измеряет согласованность предсказаний кластеров с истинными метками
	Средняя энтропия $H(\gamma)$	Контролирует степень мягкости присваиваний
	Общее значение функции потерь $\mathcal{L}$	Служит основным индикатором сходимости обучения
Ранняя остановка (Early stopping)		Обучение прекращается, если NMI на валидационной выборке не улучшается в течение пяти последовательных эпох
Контроль переобучения		Разделение данных на обучающую и валидационную выборки позволяет отслеживать разрыв между метриками на обучающих и тестовых данных
Визуальный мониторинг		Дополнительно отслеживаются динамика изменения центров кластеров $\mu_k$ , распределение значений энтропии $H(\gamma)$ по объектам и качество реконструкции прототипов кластеров

После завершения обучения модель предоставляет следующие возможности.

1. Инференс для новых объектов: для произвольного входного вектора  $x$  модель предсказывает вектор мягких присваиваний  $\gamma$ .

2. Генерация прототипов: возможность визуализации «типичных» объектов каждого кластера через декодирование центров смеси гауссиан по формуле

$$x_{proto}^{(k)} = \mu_{\theta}(\mu_k), \quad (18)$$

где  $\mu_k \in \mathbb{R}^d$  – центр каждого кластера в смеси  $p(z)$ , а  $\mu_{\theta}(\cdot)$  – функция декодера.

3. Анализ неопределенности: вычисление энтропии распределений  $\gamma$  для каждого объекта позволяет идентифицировать пограничные случаи и объекты со смешанной принадлежностью.

4. Визуализация скрытого пространства: проекция скрытых представлений  $z$  в двумерное пространство с методом стохастического вложения соседей с  $t$ -распределениями ( $t$ -distribution Stochastic Neighbor Embedding,  $t$ -SNE), а также приближения и проекции равномерного многообразия (Uniform Manifold Approximation and Projection, UMAP) с раскраской по  $\hat{k} = \arg \max_k \gamma_k$ .

Предложенная процедура обучения, сочетающая теоретически обоснованную целевую функцию с практическими методами стабилизации, обеспечивает воспроизводимость результатов и устойчивую сходимость, что подтверждается результатами экспериментов.

### 3. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ВЛИЯНИЯ КОЭФФИЦИЕНТА РЕГУЛЯРИЗАЦИИ ЭНТРОПИИ НА КЛЮЧЕВЫЕ ХАРАКТЕРИСТИКИ

#### 3.1. ПОСТАНОВКА ЭКСПЕРИМЕНТА

Для проверки эффективности предложенного алгоритма и подтверждения гипотезы о контроле мягкости присваиваний через параметр энтропийной регуляризации проведена серия экспериментов на тестовом наборе данных. В качестве тестового набора данных используется MNIST, состоящий из 70 000 изображений рукописных цифр (0–9) размером  $28 \times 28$  пикселей. Набор данных разбит на обучающую (60 000) и тестовую (10 000) выборки. Этот набор данных выбран потому, что является эталоном для оценки кластеризации в компьютерном зрении, обладает четкой кластерной структурой (10 классов) и допускает визуальную интерпретацию прототипов кластеров.

В эксперименте исследовалось влияние коэффициента  $\eta$  на три ключевые характеристики модели:

- 1) качество кластеризации метрики NMI, ARI (Adjusted Rand Index – скорректированный индекс Ранда);
- 2) энтропия распределений присваиваний  $H(\gamma)$ ;
- 3) уверенность модели (среднее значение максимального элемента вектора  $\gamma$ ).

Значения  $\eta$  варьировались в диапазоне от 0,0 до 1,0 с шагом 0,05; 0,1; 0,2; 0,3; 0,5; 1,0. Каждый эксперимент проводился при фиксированных гиперпараметрах:

- 1) размер скрытого пространства  $z = 10$ ;
- 2) число кластеров  $K = 10$ ;
- 3) температурный коэффициент softmax  $\tau = 0,3$ ;
- 4) вес  $KL$ -дивергенции  $\beta = 20$ .

Для повышения надежности оценки влияния коэффициента энтропийной регуляризации  $\eta$  и исключения переобучения на конкретное разбиение данных использовалась проверка стабильности, при которой:

- 1) обучающая выборка MNIST ( $N = 60\,000$ ) фиксировалась;
- 2) для каждого значения  $\eta \in \{0; 0,05; 0,1; 0,2; 0,3; 0,5; 1\}$  обучение проводилось трижды с разными случайными начальными весами;

3) метрики качества кластеризации (NMI, ARI) и энтропия  $H(\gamma)$  оценивались на независимой тестовой выборке MNIST ( $N = 10\,000$ ), для которой доступны только те метки, которые используются только для оценки, но не для обучения. Для корректного расчета NMI и ARI при отсутствии прямого соответствия между полученными кластерами и истинными классами предсказания модели сопоставлялись с эталонными метками с помощью венгерского алгоритма, решающего задачу оптимального назначения.

Обучение в экспериментах состояло из двух этапов: 1) предобучение автоэнкодера (3 эпохи) и 2) основное обучение с энтропийной регуляризацией (15 эпох). Для каждого значения  $\eta$  проводилась одна итерация эксперимента.

Результаты, представленные в работе, соответствуют средним значениям по трем запускам. Стандартное отклонение метрик NMI не превышало 0,008, что подтверждает устойчивость и воспроизводимость полученных выводов. Такой подход позволяет отделить эффекты от гиперпараметра  $\eta$  и от случайных флуктуаций, связанных с инициализацией модели.

### 3.2. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

На рис. 2 показано, как изменяются метрики NMI и ARI в зависимости от  $\eta$ . При  $\eta = 0,0$  (базовая VaDE) модель достигает  $NMI = 0,335$  и  $ARI = 0,157$ . С увеличением  $\eta$  наблюдается нестабильное, но в целом растущее качество кластеризации. При  $\eta = 0,05$  NMI достигает значения 0,542, что на 61,6 % выше, чем у базовой модели. Дальнейшее увеличение  $\eta$  приводит к снижению NMI, но сохраняет его на уровне выше базового (например, при  $\eta = 1,0$   $NMI = 0,511$ ). Это свидетельствует о том, что умеренная энтропийная регуляризация не только не ухудшает, но и улучшает качество кластеризации, вероятно за счет выхода из локальных минимумов, и предотвращает переобучение.

На рис. 3 видно монотонное увеличение энтропии  $H(\gamma)$  с ростом  $\eta$ : от 0,337 до 0,643 при  $\eta = 1,0$ . Это прямое подтверждение эффективности предлагаемого механизма: параметр  $\eta$  действительно контролирует мягкость распределений присваиваний.

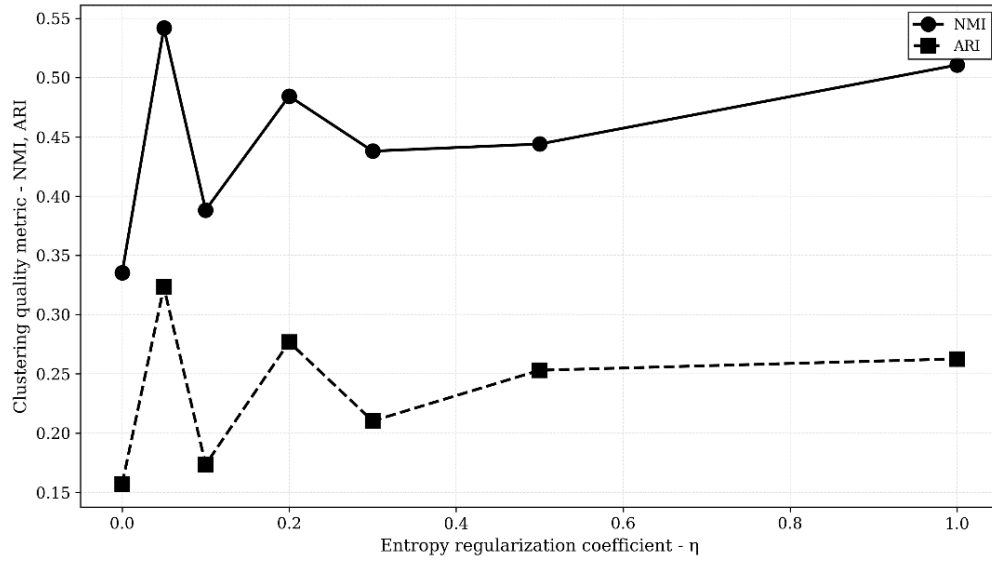


Рис. 2. Влияние  $\eta$  на качество кластеризации

Fig. 2. The influence of  $\eta$  on clustering quality

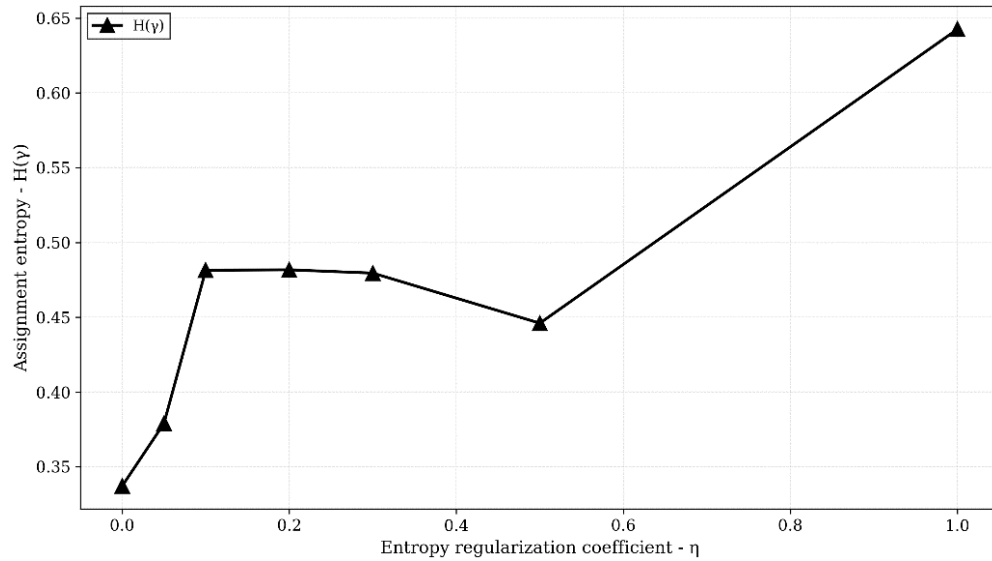


Рис. 3. Влияние  $\eta$  на энтропию присваиваний

Fig. 3. The influence of  $\eta$  on entropy assignment

На рис. 4 видно, что средняя уверенность модели (максимальное значение в  $\gamma$ ) монотонно снижается с ростом  $\eta$ : от 0,911 при  $\eta=0,0$  до 0,837 при  $\eta=1,0$ . Это подтверждает, что регуляризация снижает искусственную уверенность, характерную для коллапса присваиваний.

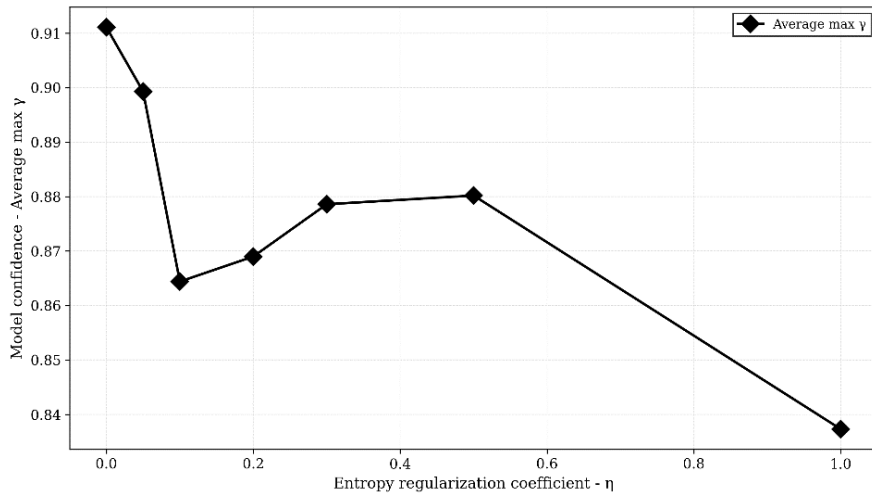


Рис. 4. Влияние  $\eta$  на уверенность модели

Fig. 4. The influence of  $\eta$  on model confidence

На рис. 5 представлена диаграмма, на которой по оси  $X$  отложена энтропия  $H(\gamma)$ , а по оси  $Y$  – NMI. Каждая точка соответствует определенному значению  $\eta$ , яркость точки кодирует величину  $\eta$ . Видно, что существует явный компромисс: высокое значение NMI достигается при низкой энтропии (жесткая кластеризация), а высокая энтропия (мягкая кластеризация) требует некоторой потери качества. Однако при  $\eta = 0,05$  модель достигает оптимального баланса:  $NMI = 0,54$  при  $H(\gamma) = 0,379$ , что значительно выше, чем у базовой модели.

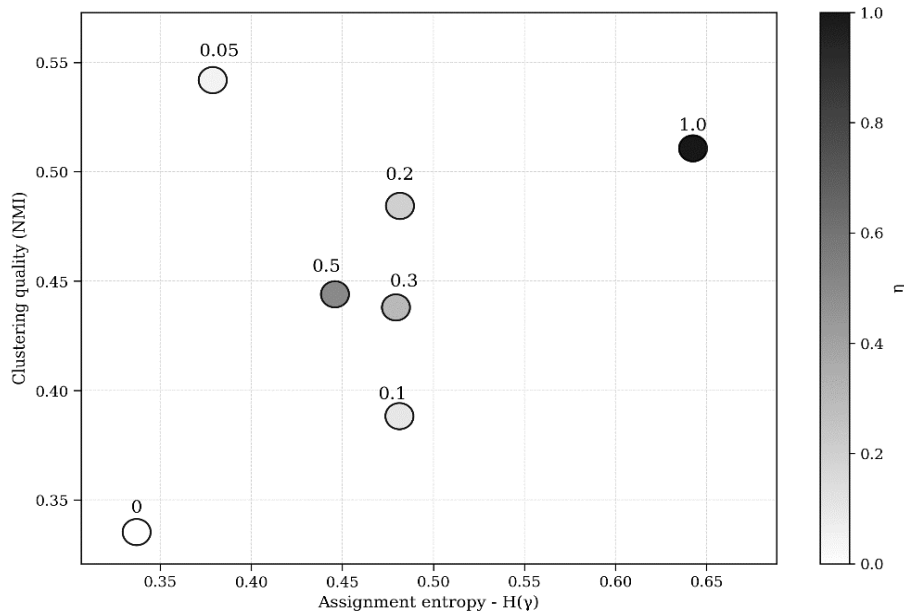


Рис. 5. Компромисс между качеством кластеризации и мягкостью присваиваний

Fig. 5. A compromise between clustering quality and assignments softness

На рис. 6 показано относительное изменение NMI и энтропии по сравнению с базовой моделью ( $\eta = 0,0$ ). При  $\eta = 0,05$  наблюдается максимальный прирост качества (61,6 %) при умеренном увеличении энтропии (12,5 %).

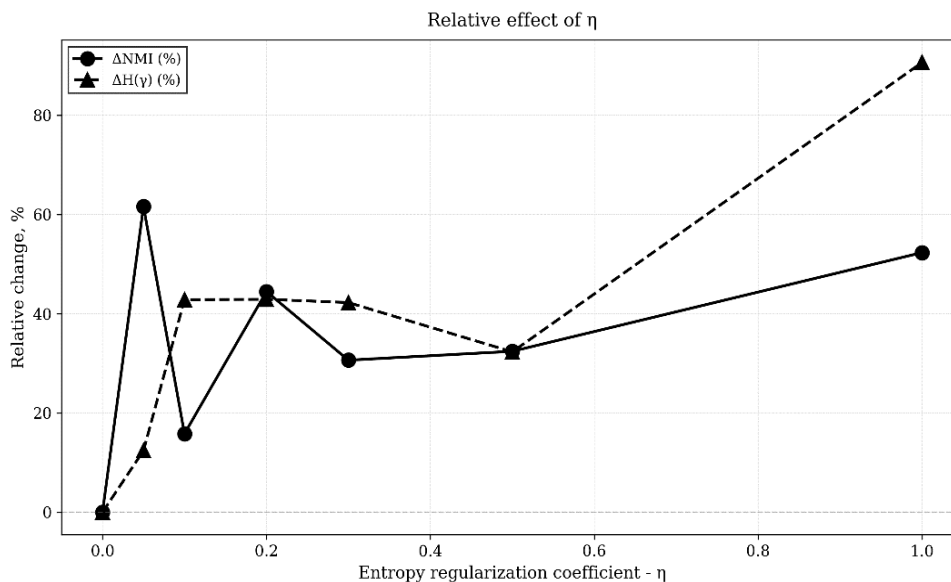


Рис. 6. Относительное влияние  $\eta$

Fig. 6. Relative influence of  $\eta$

На рис. 7 представлено соотношение  $NMI / H(\gamma)$ , которое показывает, какое значение NMI выходит на единицу энтропии.

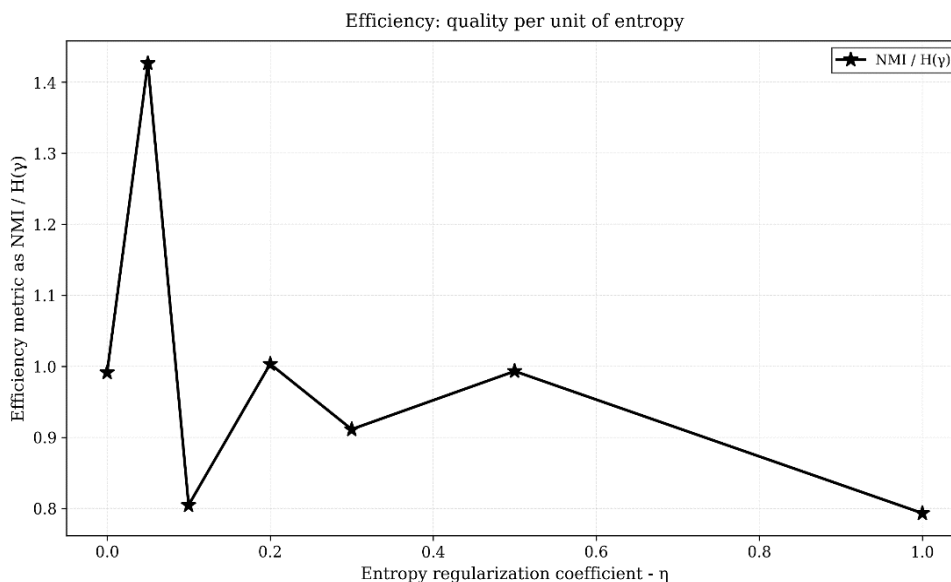


Рис. 7. Эффективность: качество кластеризации на единицу энтропии

Fig. 7. Efficiency: clustering quality per unit of entropy

Наиболее высокое значение этого соотношения достигается при  $\eta = 0,05$ , что указывает на оптимальный баланс между качеством кластеризации и неопределенностью присваивания.

### 3.3. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Полученные результаты, зафиксированные в табл. 2, позволяют сделать следующие выводы.

1. Предложенная энтропийная регуляризация эффективно контролирует мягкость присваиваний. Параметр  $\eta$  является прямым и управляемым инструментом для настройки степени неопределенности в кластеризации.

2. Умеренная регуляризация ( $\eta = 0,05$ ) не только сохраняет, но и повышает качество кластеризации. Это противоречит распространенному мнению о том, что мягкость снижает дискриминационную способность.

3. Существует фундаментальный компромисс между качеством кластеризации и мягкостью присваиваний. Этот компромисс может быть управляемым: выбор  $\eta$  позволяет адаптировать модель под конкретные прикладные задачи. Так, для задач, где важна уверенность,  $\eta = 0,0$ ; для задач, где важна интерпретируемость или учет смешанной принадлежности, –  $\eta = 0,05$  или  $\eta = 0,1$ .

Таблица 2

Table 2

**Сводные результаты экспериментов по настройке гиперпараметров модели нейронной сети и их влиянию на качество и тип кластеризации**

**Summary results of experiments on hyperparameter tuning of the neural network model and their influence on the quality and type of clustering**

$\eta$	NMI	ARI	$H(\gamma)$	Уверенность	Тип кластеризации
0,0	0,335	0,157	0,337	0,911	Очень жесткая
0,05	0,542	0,323	0,379	0,899	Оптимальная
0,1	0,388	0,173	0,481	0,864	Умеренно мягкая
0,2	0,484	0,277	0,482	0,869	Мягкая
0,3	0,438	0,210	0,480	0,879	Мягкая
0,5	0,444	0,253	0,446	0,880	Умеренно мягкая
1,0	0,511	0,263	0,643	0,837	Очень мягкая

Эксперименты подтвердили гипотезу о возможности управления степенью мягкости в кластерных присваиваниях через параметр энтропийной регуляризации  $\eta$ . Исходя из представленной выше таблицы можно утверждать, что оптимальное значение  $\eta = 0,05$  обеспечивает:

- 1) максимальное качество кластеризации ( $NMI = 0,542$ );
- 2) умеренную энтропию ( $H(\gamma) = 0,379$ );
- 3) снижение искусственной уверенности до 0,899.

Полученные результаты имеют непосредственное практическое значение. Рассмотрим в качестве примера задачу анализа региональных систем, подобную исследованию [5].

Если цель исследования – построение четкой типологии регионов для классификационных мер государственной поддержки, то может быть выбрано значение параметра  $\eta = 0,0$ , обеспечивающее жесткую кластеризацию. В этом случае каждый регион будет отнесен ровно к одному типу экономических условий, что удобно для принятия однозначных административных решений.

Однако если исследователь ставит задачу выявления регионов со смешанными или переходными экономическими условиями (например, территорий, сочетающих черты сырьевой и инновационной модели развития), то оптимальным выбором станет значение  $\eta = 0,05$  или  $\eta = 0,1$ . Полученные для таких регионов вероятностные присваивания  $\gamma$  укажут на многопрофильность: например, регион может с вероятностью 0,6 относиться к сырьевому типу и с вероятностью 0,4 – к инновационному. Это позволит избежать упрощений при анализе и разработать более точные, дифференцированные стратегии регионального развития, учитывающие неоднородность экономической структуры.

Таким образом, предложенная модель не только решает проблему коллапса присваиваний, но и позволяет настраивать баланс между дискриминативностью и неопределенностью, что делает ее пригодной для широкого спектра прикладных задач, где важны четкость классификации и учет переходных состояний.

## ЗАКЛЮЧЕНИЕ

В работе проведено исследование фундаментальной проблемы современных алгоритмов глубокой кластеризации (преждевременный коллапс присваиваний), приводящее к детерминированным распределениям и потере способности модели отражать смешанную принадлежность объектов. Для решения этой проблемы предложена модифицированная архитектура автоэнкодера на основе модели VaDE, включающая явную энтропийную регуляризацию целевой функции. Ключевым элементом предложенного подхода является введение параметризованного коэффициента  $\eta$ , который обеспечивает прямой и управляемый контроль над степенью мягкости кластерных присваиваний на всех этапах обучения.

Проведенные эксперименты на наборе данных MNIST подтвердили гипотезу исследования: значение параметра  $\eta = 0,05$  позволяет достичь оптимального баланса между качеством кластеризации и сохранением неопределенности. При этом значение NMI увеличивается на 61,6 % по сравнению с базовой моделью без регуляризации (0,542 против 0,335) при умеренном росте энтропии распределений присваиваний (0,379 против 0,337) и снижении избыточной уверенности модели. Более высокие значения параметра  $\eta$  (например, 1,0) обеспечивают более высокую энтропию (0,643), что делает модель пригодной для задач, где критически важна интерпретация неопределенности, например, в биомедицинских или маркетинговых приложениях.

Полученные результаты показывают, что мягкость и качество кластеризации не находятся в противоречии, а образуют управляемый компромисс, который может быть настроен в соответствии с требованиями конкретной предметной области. Предложенный механизм энтропийной регуляризации не только устраняет коллапс присваиваний, но и открывает путь к построению надежных, интерпретируемых и прикладных систем кластеризации, соответствующих ожиданиям конечных пользователей.

Таким образом, исследования вносят вклад в развитие теории и практики глубокой кластеризации, предлагая простое, но эффективное решение, что обеспечивает вероятностную достоверность и интерпретируемость, а также гибкость в задачах анализа неразмеченных данных.

Особый интерес представляет применение предложенного метода в задачах сегментации изображений, например, в медицине (выделение опухолевых тканей с нечеткими границами на МРТ) или при анализе данных дистанционного зондирования Земли (классификация участков со смешанными типами подстилающей поверхности). Способность модели сохранять неопределенность в таких задачах критически важна для получения более достоверных и интерпретируемых результатов, позволяя врачу или аналитику видеть не только «жесткий» вердикт, но и степень уверенности модели в каждом пикселе или участке.

В качестве направлений дальнейших исследований предполагается:

- 1) расширение подхода на гетерогенных и мультимодальных данных;
- 2) интеграция механизма регуляризации в архитектуры с механизмами внимания (например, рекуррентные нейронные сети) для усиления дискриминационных признаков;
- 3) применение предложенного алгоритма в реальных промышленных сценариях с последующей оценкой бизнес-метрик.

## СПИСОК ЛИТЕРАТУРЫ

1. Deep clustering: A comprehensive survey / Y. Ren, J. Pu, Z. Yang, J. Xu, G. Li, X. Pu // IEEE Transactions on Neural Networks and Learning Systems. – 2024. – Vol. 36 (4). – P. 5858–5878. – DOI: 10.1109/TNNLS.2024.3403155.
2. Deep embedded clustering with data augmentation / C. Song, Y. Huang, W. Ouyang, L. Wang // Proceedings of the AAAI Conference on Artificial Intelligence. – 2023. – Vol. 37 (2). – P. 2345–2353. – DOI: 10.1609/aaai.v37i2.25345.
3. DVAEGMM: Dual variational autoencoder with Gaussian mixture model for anomaly detection on attributed networks / W. Khan, M. Haroon, A.N. Khan, M.N. Hasan, A. Khan, U. Mokhtar // IEEE Access. – 2022. – Vol. 10. – P. 91160–91176. – DOI: 10.1109/ACCESS.2022.3201332.
4. Entropy regularization for unsupervised clustering with adaptive neighbors / J. Wang, X. Wang, Z. Tian, C. Liu, Y. Liu // Pattern Recognition. – 2022. – Vol. 125 (2). – P. 108517. – DOI: 10.1016/j.patcog.2021.108517.
5. Алетдинова А.А. Кластерный анализ для исследования экономических условий эффективного функционирования АПК регионов // Формирование организационно-экономических условий эффективного функционирования АПК: сборник научных статей XVI Международной научно-практической конференции, 23–24 мая 2024. – Минск, 2024. – С. 37–40. – ISBN 978-985-25-0099-9.
6. Niu C., Shan H., Wang G. SPICE: Semantic pseudo-labeling for image clustering // IEEE Transactions on Image Processing. – 2022. – Vol. 31. – P. 7264–7278. – DOI: 10.1109/TIP.2022.3221290.

7. Wang Y., Blei D., Cunningham J.P. Posterior collapse and latent variable non-identifiability // Proceedings of the 35th International Conference on Neural Information Processing Systems (NeurIPS 2021). – 2021. – Vol. 34. – P. 5443–5455. – ISBN 978-1-7138-4539-3.
8. Yang L., Fan W., Bouguila N. Deep clustering analysis via dual variational autoencoder with spherical latent embeddings // IEEE Transactions on Neural Networks and Learning Systems. – 2023. – Vol. 34 (9). – P. 6303–6312. – DOI: 10.1109/TNNLS.2021.3135460.
9. Caciularu A., Goldberger J. An entangled mixture of variational autoencoders approach to deep clustering // Neurocomputing. – 2023. – Vol. 529. – P. 182–189. – DOI: 10.1016/j.neucom.2023.01.069.
10. Deep clustering analysis via variational autoencoder with gamma mixture latent embeddings / J. Guo, M. Yand, Y. Zhang, Z. Liu // Neural Networks. – 2025. – Vol. 183. – P. 106979.
11. Cloud-VAE: Variational autoencoder with concepts embedded / Y. Liu, Z. Liu, S. Li, Z. Yu, Y. Guo, Q. Liu, G. Wang // Pattern Recognition. – 2023. – Vol. 140. – P. 109530. – DOI: 10.1016/j.patcog.2023.109530.
12. Svirsky J., Lindenbaum O. Interpretable deep clustering for tabular data // Proceedings of Machine Learning Research. – 2024. – Vol. 235: International Conference on Machine Learning (ICML 2024). – P. 47314–47330. – ISBN 979-8-3313-0223-8.
13. Deep generative clustering with multimodal diffusion variational autoencoders / E. Palumbo, L. Manduchi, S. Laguna, D. Chopard, J. Vogt // The 12th International Conference on Learning Representations (ICLR 2024). – Vienna, Austria, 2024. – P. 35404–35429. – ISBN 978-1-7138-9865-8.
14. Controlling posterior collapse by an inverse Lipschitz constraint on the decoder network / Y. Kinoshita, K. Oono, K. Fukumizu, Y. Yoshida, Sh. Maeda // Proceedings of Machine Learning Research. – 2023. – Vol. 202: International Conference on Machine Learning, 23–29 July 2023, Honolulu, Hawaii, USA. – P. 17041–17060. – ISBN 978-1-7138-8918-2.
15. Ma H. Achieving deep clustering through the use of variational autoencoders and similarity-based loss // Mathematical Biosciences and Engineering. – 2022. – Vol. 19 (10). – P. 10344–10360. – DOI: 10.3934/mbe.2022484.

*Алетдинова Анна Александровна*, доктор экономических наук, кандидат технических наук, профессор кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – анализ данных и системы искусственного интеллекта. Имеет 344 печатные работы. E-mail: aletdinova@corp.nstu.ru

*Антонянц Егор Николаевич*, аспирант, ассистент кафедры автоматизированных систем управления Новосибирского государственного технического университета. Основное направление научных исследований – машинное обучение и системы компьютерного зрения. Имеет более 50 печатных работ. E-mail: bax201438@gmail.com

*Сборщиков Егор Иванович*, студент 3-го курса отделения информационных технологий Новосибирского авиационного технического колледжа им. Б.С. Галушача. Основное направление научных исследований – машинное обучение и глубокое обучение. E-mail: egorsborsikov250575@gmail.com

*Aletdinova Anna A.*, D. Sc. (Econ.), Ph.D. (Eng), full professor at the Department of Automated Control Systems, Novosibirsk State Technical University. Her research interests are currently focused on data analysis and artificial intelligence systems. She has more than 344 publications. E-mail: aletdinova@corp.nstu.ru

*Antonyants Egor N.*, postgraduate student, assistant lecturer at the Department of Automated Control Systems, Novosibirsk State Technical University. His research interests are currently focused on machine learning and computer vision systems. He has more than 50 publications. E-mail: bax201438@gmail.com

*Sborshchikov Egor I.*, third-year student of information technologies faculty at the Galushchak Novosibirsk Aircraft Technical College. His research interests are currently focused on machine learning and deep learning. E-mail: egorsborsikov250575@gmail.com

## **Research into deep clustering algorithms with controlled entropy of soft assignments\***

A.A. ALETDINOVA<sup>1,a</sup>, E.N. ANTONYANTS<sup>1,b</sup>, E.I. SBORSHCHIKOV<sup>2,c</sup>

<sup>1</sup> Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

<sup>2</sup> Galushchak's Novosibirsk Aircraft Technical College, 72, Krasny Prospekt, Novosibirsk, 630091, Russian Federation

<sup>a</sup> aletdinova@corp.nstu.ru <sup>b</sup> bax201438@gmail.com <sup>c</sup> egorsborsikov250575@gmail.com

### **Abstract**

The article investigates the problem of premature collapse of cluster assignments in modern deep clustering algorithms, including Deep Embedded Clustering (DEC) and Variational Deep Embedding (VaDE). This phenomenon results in deterministic rather than probabilistic membership distributions, contradicting the nature of real-world data wherein objects exhibit mixed membership across multiple clusters, such as users of digital services, biological samples with ambiguous classifications, or specimens in materials science. To address this issue, we propose a modified autoencoder architecture based on VaDE, incorporating an entropy regularization mechanism into its objective function that balances clustering quality against assignment softness (uncertainty). The approach maintains direct adaptive preservation of high entropy in soft assignment distributions throughout model training, thereby ensuring probabilistic interpretation of results. Experiments on the modernized dataset from the Modified National Institute of Standards and Technology (MNIST) demonstrate an optimal regularization parameter value of 0,05, which improves the Normalized Mutual Information (NMI) by 61,6% compared to the baseline model without regularization. The neural network simultaneously preserves interpretability through the generation of semantically meaningful cluster prototypes. This model demonstrates practical utility in applied domains requiring explainability and uncertainty quantification—including marketing segmentation, behavioral analytics, biomedical technologies, and materials surface analysis. The proposed approach bridges the gap between algorithmic performance and practitioners' demand for explainable and reliable clustering outcomes.

**Keywords:** deep clustering, variational autoencoders, soft assignments, entropy regularization, interpretability, unsupervised learning, uncertainty, generative models, trade-off optimization

### **REFERENCES**

1. Ren Y., Pu J., Yang Z., Xu J., Li G., Pu X. Deep clustering: A comprehensive survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, vol. 36 (4), pp. 5858–5878. DOI: 10.1109/TNNLS.2024.3403155.
2. Song C., Huang Y., Ouyang W., Wang L. Deep embedded clustering with data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, vol. 37 (2), pp. 2345–2353. DOI: 10.1609/aaai.v37i2.25345.
3. Khan W., Haroon M., Khan A.N., Hasan M.K., Khan A., Mokhtar U.A. DVAEGMM: Dual variational autoencoder with Gaussian mixture model for anomaly detection on attributed networks. *IEEE Access*, 2022, vol. 10, pp. 91160–91176. DOI: 10.1109/ACCESS.2022.3201332.
4. Wang J., Wang X., Tian Z., Liu C., Liu Y. Entropy regularization for unsupervised clustering with adaptive neighbors. *Pattern Recognition*, 2022, vol. 125 (2), p. 108517. DOI: 10.1016/j.patcog.2021.108517.

---

\* Received 28 January 2026.

5. Aletdinova A.A. [Cluster analysis for research the economic conditions for the effective functioning of regional AIC]. *Formirovanie organizatsionno-ekonomicheskikh uslovii effektivnogo funktsionirovaniya APK* [Formation of organizational and economic conditions for the effective functioning of the AIC]. Collection of scientific articles from the XVI International Scientific and Practical Conference. Minsk, 2024, pp. 37–40. ISBN 978-985-25-0099-9. (In Russian).
6. Niu C., Shan H., Wang G. SPICE: Semantic pseudo-labeling for image clustering. *IEEE Transactions on Image Processing*, 2022, vol. 31, pp. 7264–7278. DOI: 10.1109/TIP.2022.3221290.
7. Wang Y., Blei D., Cunningham J.P. Posterior collapse and latent variable non-identifiability. *Proceedings of the 35th International Conference on Neural Information Processing Systems (NeurIPS 2021)*, 2021, vol. 34, pp. 5443–5455. ISBN 978-1-7138-4539-3.
8. Yang L., Fan W., Bouguila N. Deep clustering analysis via dual variational autoencoder with spherical latent embeddings. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, vol. 34 (9), pp. 6303–6312. DOI: 10.1109/TNNLS.2021.3135460.
9. Caciularu A., Goldberger J. An entangled mixture of variational autoencoders approach to deep clustering. *Neurocomputing*, 2023, vol. 529, pp. 182–189. DOI: 10.1016/j.neucom.2023.01.069.
10. Guo J., Yand M., Zhang Y., Liu Z. Deep clustering analysis via variational autoencoder with gamma mixture latent embeddings. *Neural Networks*, 2025, vol. 183, p. 106979.
11. Liu Y., Liu Z., Li S., Yu Z., Guo Y., Liu Q., Wang G. Cloud-VAE: Variational auroencoder with concepts embedded. *Pattern Recognition*, 2023, vol. 140, p. 109530. DOI: 10.1016/j.patcog.2023.109530.
12. Svirsky J., Lindenbaum O. Interpretable deep clustering for tabular data. *Proceedings of Machine Learning Research*, 2024, vol. 235. *International Conference on Machine Learning (ICML 2024)*, pp. 47314–47330. ISBN 979-8-3313-0223-8.
13. Palumbo E., Manduchi L., Laguna S., Chopard D., Vogt J. Deep generative clustering with multimodal diffusion variational autoencoders. *The 12th International Conference on Learning Representations (ICLR 2024)*, Vienna, Austria, 2024, pp. 35404–35429. ISBN 978-1-7138-9865-8.
14. Kinoshita Y., Oono K., Fukumizu K., Yoshida Y., Maeda Sh. Controlling posterior collapse by an inverse Lipschitz constraint on the decoder network. *Proceedings of Machine Learning Research*, 2023, vol. 202. *International Conference on Machine Learning*, 23–29 July 2023, Honolulu, Hawaii, USA, pp. 17041–17060. ISBN 978-1-7138-8918-2.
15. Ma H. Achieving deep clustering through the use of variational autoencoders and similarity-based loss. *Mathematical Biosciences and Engineering*, 2022, vol. 19 (10), pp. 10344–10360. DOI: 10.3934/mbe.2022484.

Для цитирования:

Алетдинова А.А., Антонянц Е.Н., Сборщиков Е.И. Исследование алгоритмов глубокой кластеризации с контролируемой энтропией мягких присваиваний // Системы анализа и обработки данных. – 2026. – № 2 (102). – С. 7–30. – DOI: 10.17212/2782-2001-2026-2-7-30.

For citation:

Aletdinova A.A., Antonyants E.N., Sborshchikov E.I. Issledovanie algoritmov glubokoi klasterizatsii s kontroliruemoi entropiei myagkikh prisvaivanii [Research into deep clustering algorithms with controlled entropy of soft assignments]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2026, no. 2 (102), pp. 7–30. DOI: 10.17212/2782-2001-2026-2-7-30.