

<https://doi.org/10.25559/SITITO.022.202601.171-182>
УДК 004.032.26:303.722.3

Оригинальная статья

Инкрементальное нейросетевое обобщённое неметрическое многомерное шкалирование

А. М. Колосов

ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова», г. Москва, Рос-
сийская Федерация

Адрес: 119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1
aleksei.kolosov@math.msu.ru

Аннотация

В статье рассматривается инкрементальная версия нейросетевого обобщённого неметриче-
ского многомерного шкалирования, предназначенная для достраивания пространства пред-
ставлений после появления новых объектов. В отличие от классического GNMDs, где весь набор
объектов оптимизируется совместно по порядковым ограничениям, инкрементальный сцена-
рий разделяет обучение на старую и новую порции данных. На датасетах MEN, SimLex-999 и
WordSim-353 исследуются режимы *Только старые объекты*, *Полное переобучение пространства*,
Фиксация старых координат и *Расширение старых координат*. Показано, что *Фиксация старых*
координат ухудшает показатель качества на парах old-new, тогда как *Расширение* пространства
с L2-регуляризацией дополнительного блока *старых координат* существенно улучшает ранго-
вый показатель качества и одновременно ограничивает деформацию старых объектов.

Ключевые слова: обобщённое неметрическое многомерное шкалирование, GNMDs, инкре-
ментальное обучение, L2-регуляризация, семантические векторные пространства

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Для цитирования: Колосов А. М. Инкрементальное нейросетевое обобщённое неметриче-
ское многомерное шкалирование // Современные информационные технологии и ИТ-образо-
вание. 2026. Т. 22, № 1. С. 171-182. <https://doi.org/10.25559/SITITO.022.202601.171-182>

© Колосов А. М., 2026



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Incremental Neural Generalized Non-Metric Multidimensional Scaling

A. M. Kolosov

Lomonosov Moscow State University, Moscow, Russian Federation

Address: 1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation

aleksei.kolosov@math.msu.ru

Abstract

The paper considers an incremental version of neural generalized non-metric multidimensional scaling intended for extending a representation space after the appearance of new objects. Unlike classical GNMDS, where the entire set of objects is jointly optimized according to ordinal constraints, the incremental scenario separates training into old and new portions of data. The modes *Old Objects Only*, *Full Space Retraining*, *Fixed Old Coordinates*, and *Expanded Old Coordinates* are studied on the MEN, SimLex-999, and WordSim-353 datasets. It is shown that *Fixed Old Coordinates* worsens the quality metric on old-new pairs, whereas *Space Expansion* with L2 regularization of the additional block of *old coordinates* substantially improves the rank-based quality metric and at the same time limits the deformation of old objects.

Keywords: generalized non-metric multidimensional scaling, GNMDS, incremental learning, L2 regularization, semantic vector spaces

Conflict of interests: The author declares no conflict of interests.

For citation: Kolosov A.M. Incremental Neural Generalized Non-Metric Multidimensional Scaling. *Modern Information Technologies and IT-Education*. 2026;22(1):171-182. <https://doi.org/10.25559/SITITO.022.202601.171-182>



1 Введение

Многомерное шкалирование используется для построения геометрических представлений объектов по данным о близости, различии или предпочтениях. В классической неметрической постановке требуется сохранить не абсолютные значения расстояний, а их порядок, что делает метод естественным инструментом для анализа экспертных сравнений и семантических оценок [1-3]. Обобщённое неметрическое многомерное шкалирование, или GNMDS, формулирует задачу через ограничения вида «одна пара объектов должна быть ближе другой пары» и поэтому допускает неполные и даже противоречивые множества сравнений [4]. В нейросетевой реализации такие ограничения удобно оптимизировать через функцию потерь на четвёрках объектов, близкую к margin-ranking и quadruplet loss [5], [6]; сама идея обучаемого представления также связана с более широкой линией нейросетевого и спектрального снижения размерности [23-25].

Современные embedding-модели, начиная с моделей Word2Vec и GloVe, строят точки в непрерывном пространстве, где расстояния или косинусные близости должны отражать семантические отношения [7-10]. Однако уже построенное пространство часто используется внешними системами: поиском, рекомендательными механизмами, индексами ближайших соседей или аналитическими инструментами. В такой ситуации появление новых объектов не всегда позволяет полностью перестроить все координаты. Требуется инкрементальная процедура, которая добавляет новые представления и по возможности сохраняет старую геометрию.

Семантические датасеты WordSim-353, SimLex-999 и MEN задают разные типы человеческих оценок: WordSim-353 связан с семантическим сходством и контекстной близостью [11], SimLex-999 специально отделяет similarity от association/relatedness [12], а MEN содержит 3000 пар слов с оценками семантической relatedness, полученными через сравнительные суждения [13]. Поэтому семантические датасеты удобны для проверки того, насколько ранговая геометрия сохраняется при инкрементальном обновлении.

2 Задачи многомерного шкалирования

Многомерное шкалирование стремится представить набор объектов $a_1 \dots a_n$ точками $g(a_i) \in \mathbb{R}^d$ так, чтобы геометрические отношения между точками согласовывались с исходными экспертными или статистическими данными. Если эксперт считает, что объекты a и b ближе, чем объекты a и c , то в построенном пространстве желательно получить

$$a \approx b, a \text{ not } \approx c \Rightarrow e(g(a), g(b)) < e(g(a), g(c)),$$

где $e(\cdot, \cdot)$ обозначает выбранную функцию расстояния. В неметрическом варианте ключевым является порядок расстояний, а не их абсолютный масштаб [1-3]. Это особенно важно для экспертных шкал, где численные оценки могут иметь разную калибровку у разных аннотаторов.

2.1 Построение представлений

В задачах семантического моделирования исходная таблица близостей обычно получается из человеческих оценок, из словарных ресурсов или из статистики совместной встречаемости.

сти. После этого строится векторное пространство, где пары с высокой экспертной близостью должны располагаться ближе пар с низкой экспертной близостью. Табличное представление ранговых отношений показано на рис. 1.

	(i, j)	.	.	.	(k, l)
(i, j)	#	<	>	<	>
.	#	#	>	>	<
.	#	#	#	<	>
.	#	#	#	#	<
(k, l)	#	#	#	#	#

Рис. 1. Представление ранговых отношений в виде таблицы сравнений пар

Fig. 1. Representation of rank relations as a table of pairwise comparisons

Источник: здесь и далее в статье все таблицы и рисунки составлены автором.
Source: Hereinafter in this article all tables and figures were made by the author.

3 Постановка задачи обобщённого неметрического многомерного шкалирования

Пусть имеется множество объектов V и набор порядковых ограничений Q . Каждый элемент Q имеет вид (i, j, k, l) и означает, что пара (i, j) должна быть ближе пары (k, l) . Тогда задача GNMDS состоит в подборе отображения $g: V \rightarrow \mathbb{R}^d$, такого что

$$d(i, j) < d(k, l) \Rightarrow e(g(i), g(j)) < e(g(k), g(l)).$$

Исходная статья Agarwal et al. подчёркивает, что GNMDS допускает неполные и противоречивые ограничения, что отличает его от классической постановки с полной матрицей различий [4]. Нейросетевой вариант заменяет прямую оптимизацию координат параметрической или полупараметрической моделью, обучаемой по наборам четвёрок. Такая схема близка к задачам метрического обучения и ordinal embedding [14], [15].

4 Подход к решению задачи

Задача GNMDS решается как задача обучения векторных представлений по порядковым сравнениям. На вход поступают утверждения о том, что одна пара объектов должна быть ближе другой пары. Модель сопоставляет каждому объекту вектор, вычисляет расстояния для двух сравниваемых пар и изменяет координаты так, чтобы порядок расстояний совпал с заданными отношениями. Такой подход не требует восстановления полной метрической шкалы: достаточно сохранить порядок близостей, который задан в исходной таблице различий.

4.1 Общая схема реализации

Для каждого ограничения (i, j, k, l) модель сравнивает две пары объектов. Пара (i, j) считается более близкой, а пара (k, l) – более далёкой. После отображения объектов в пространство

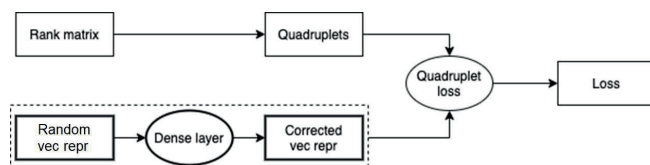


признаков вычисляются два расстояния: $e(g(i), g(j))$ и $e(g(k), g(l))$. Если более близкая пара оказывается не ближе более далёкой пары с заданным запасом, возникает штраф:

$$L(i,j,k,l) = \max(0, e(g(i), g(j)) - e(g(k), g(l)) + \gamma).$$

Здесь γ задаёт минимальный желательный зазор между расстояниями. Иными словами, модель получает ошибку тогда, когда порядок расстояний нарушен или выполнен слишком слабо. Эта ошибка используется для обновления векторов объектов.

Архитектура схемы обучения показана на рис. 2. В левой части схемы находятся пары объектов, которые образуют порядковые ограничения. Далее общий блок построения представлений переводит каждый объект в вектор. В центральной части отдельно вычисляются расстояния для «ближней» и «дальней» пар. В правой части эти расстояния сравниваются с учётом отступа γ , после чего формируется quadruplet loss. Полученный штраф передаётся назад в оптимизатор и меняет обучаемые координаты.



Р и с. 2. Архитектура схемы обучения с quadruplet loss
Fig. 2. Architecture of the training scheme with quadruplet loss

4.2 Ограничения градиентной оптимизации

Градиентная оптимизация в задаче GNMDs обычно находит приближённое решение. Функция потерь зависит от многих координат одновременно, поэтому разные запуски могут прийти к разным локальным решениям. Это означает, что результат обучения следует оценивать не как строгий глобальный минимум, а как найденное моделью согласование заданных порядковых отношений [16], [17].

Качество оценивается через коэффициент Spearman ρ . Корреляция Спирмена вычисляется между двумя списками рангов: первый список получается из исходных экспертных оценок или целевых различий, а второй – из расстояний между построенными векторами. Если пары, которые экспертно считаются более близкими, действительно получают меньшие расстояния в новом пространстве, значение ρ становится выше. Поэтому при одинаковых данных, размерностях, функции расстояния и параметрах оптимизации различия в ρ показывают, какой набор векторных представлений лучше сохраняет исходный порядок близостей.

4.3 Стохастический градиентный спуск и условия сходимости

На практике полный градиент по всем четвёркам часто заменяется стохастическим градиентным спуском (*Stochastic Gradient Descent, SGD*). SGD сходится к тому же стационарному решению, что и полный градиентный спуск, только при достаточно сильных предпосылках: мини-батчевый градиент должен быть несмещённой оценкой полного градиента, дис-

персия шума должна быть ограниченной, целевая функция должна быть гладкой, а последовательность шагов обучения должна удовлетворять условиям убывания, например $\sum_t \eta_t = \infty$ и $\sum_t \eta_t^2 < \infty$ в классической схеме Роббинса-Монро [18], [19]. При постоянном шаге SGD обычно сходится не в точку, а в окрестность стационарного решения, размер которой определяется уровнем шума и величиной шага.

Если целевая функция выпукла, эти условия дают более сильные утверждения о сходимости к глобальному минимуму или к оптимальному множеству. В невыпуклой постановке, характерной для нейросетевого GNMDs, стандартный результат слабее: ожидаемая норма градиента стремится к нулю или становится малой, но глобальная оптимальность не следует автоматически [20]. Поэтому в данной работе SGD рассматривается как масштабируемая процедура поиска приемлемого приближённого решения, а не как механизм с гарантией глобального оптимума.

5 Инкрементальные постановки: inductive-схема и transductive-альтернатива

Классическая постановка GNMDs предполагает, что весь набор объектов и все используемые порядковые ограничения зафиксированы до начала обучения. Инкрементальный сценарий отличается тем, что сначала доступна старая порция объектов, а затем появляются новые объекты и новые экспертные связи. Дальнейшие эксперименты посвящены *inductive incremental reconstruction*: на первом этапе используются только old-объекты и old-old ограничения, а на втором этапе становятся доступны ограничения, затрагивающие new-объекты.

Inductive incremental reconstruction — это постановка, в которой на первом этапе доступна только часть матрицы, относящаяся к старым объектам. После появления новых объектов становится доступна новая часть матрицы, содержащая связи новых объектов со старыми и, возможно, связи между новыми объектами; затем требуется построить представления новых объектов относительно уже имеющегося пространства.

Transductive incremental reconstruction — альтернативная постановка, в которой вся матрица экспертных близостей или весь набор известных пар доступен заранее, но координаты сначала строятся только для подмножества объектов. Такой режим измеряет реконструкцию уже известной таблицы, а не способность работать с позднее появившимися связями.

Различие между постановками существенно для интерпретации. В inductive-схеме показатель качества отражает устойчивость уже построенного пространства к добавлению новой информации. В transductive-альтернативе алгоритм не предсказывает будущие экспертные суждения: они уже известны и только временно скрываются на уровне координатной оптимизации.

Далее основной акцент делается на inductive-схеме: сначала обучается пространство старых объектов, после чего сравниваются способы добавления новых объектов без использования будущих связей на первом этапе.



6 Базовые режимы обновления

Эксперименты посвящены сравнению нескольких способов работы с новыми объектами после того, как пространство старых объектов уже построено. Сначала обучается режим Только старые объекты: он использует только old-объекты и пары внутри старого подмножества. Затем рассматриваются

два базовых способа включить новые объекты. Режим Полное переобучение пространства заново оптимизирует координаты всех old- и new-объектов по ограничениям, доступным на втором этапе. Режим Фиксация старых координат оставляет old-координаты неизменными и обучает только координаты new-объектов.

Таблица 1. Базовые режимы обновления координат в инкрементальном GNMDs

Table 1. Basic coordinate update modes in incremental GNMDs

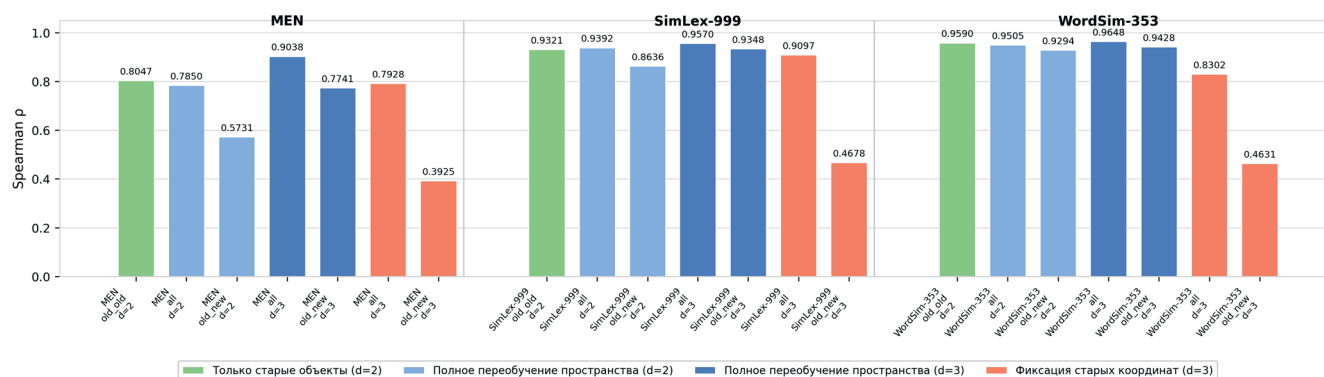
Режим	Обучаемые координаты	Используемые ограничения	Интерпретация
Только старые объекты	Только old-объекты	Пары old-old	Начальная геометрия старого пространства
Полное переобучение пространства	Все old- и new-координаты	Все известные на втором этапе пары	Контрольный ориентир качества при полной перестройке пространства
Фиксация старых координат	Только координаты new-объектов	Ограничения, затрагивающие new-объекты, при фиксированной old-геометрии	Добавление новых объектов без полной перестройки старого пространства

Таблица 2. Лучшие значения Spearman ρ для базовых режимов

Table 2. Best Spearman ρ values for the basic modes

Датасет	Расстояние	Режим	Разбиение	Размерность	Spearman ρ
MEN	cosine	Только старые объекты	old_old	2	0.8047
MEN	cosine	Полное переобучение пространства	all	2	0.7850
MEN	cosine	Полное переобучение пространства	old_new	2	0.5731
MEN	cosine	Полное переобучение пространства	all	3	0.9038
MEN	cosine	Полное переобучение пространства	old_new	3	0.7741
MEN	cosine	Фиксация старых координат	all	3	0.7928
MEN	cosine	Фиксация старых координат	old_new	3	0.3925
MEN	euclidean	Только старые объекты	old_old	2	0.9295
MEN	euclidean	Полное переобучение пространства	all	2	0.9187
MEN	euclidean	Полное переобучение пространства	old_new	2	0.7815
MEN	euclidean	Полное переобучение пространства	all	3	0.9669
MEN	euclidean	Полное переобучение пространства	old_new	3	0.8809
MEN	euclidean	Фиксация старых координат	all	3	0.9109
MEN	euclidean	Фиксация старых координат	old_new	3	0.5458
SimLex-999	cosine	Только старые объекты	old_old	2	0.9321
SimLex-999	cosine	Полное переобучение пространства	all	2	0.9392
SimLex-999	cosine	Полное переобучение пространства	old_new	2	0.8636
SimLex-999	cosine	Полное переобучение пространства	all	3	0.9570
SimLex-999	cosine	Полное переобучение пространства	old_new	3	0.9348
SimLex-999	cosine	Фиксация старых координат	all	3	0.9097
SimLex-999	cosine	Фиксация старых координат	old_new	3	0.4678
SimLex-999	euclidean	Только старые объекты	old_old	2	0.9848
SimLex-999	euclidean	Полное переобучение пространства	all	2	0.9849
SimLex-999	euclidean	Полное переобучение пространства	old_new	2	0.9574
SimLex-999	euclidean	Полное переобучение пространства	all	3	0.9867
SimLex-999	euclidean	Полное переобучение пространства	old_new	3	0.9792
SimLex-999	euclidean	Фиксация старых координат	all	3	0.9640
SimLex-999	euclidean	Фиксация старых координат	old_new	3	0.5035
WordSim-353	cosine	Только старые объекты	old_old	2	0.9590
WordSim-353	cosine	Полное переобучение пространства	all	2	0.9505
WordSim-353	cosine	Полное переобучение пространства	old_new	2	0.9294
WordSim-353	cosine	Полное переобучение пространства	all	3	0.9648
WordSim-353	cosine	Полное переобучение пространства	old_new	3	0.9428
WordSim-353	cosine	Фиксация старых координат	all	3	0.8302
WordSim-353	cosine	Фиксация старых координат	old_new	3	0.4631
WordSim-353	euclidean	Только старые объекты	old_old	2	0.9566
WordSim-353	euclidean	Полное переобучение пространства	all	2	0.9327
WordSim-353	euclidean	Полное переобучение пространства	old_new	2	0.9350
WordSim-353	euclidean	Полное переобучение пространства	all	3	0.9439
WordSim-353	euclidean	Полное переобучение пространства	old_new	3	0.9457
WordSim-353	euclidean	Фиксация старых координат	all	3	0.8973
WordSim-353	euclidean	Фиксация старых координат	old_new	3	0.7212



Р и с. 3. График к таблице 2: значения Spearman ρ для базовых режимов при cosine-расстоянииFig. 3. Chart for Table 2: Spearman ρ values for the basic modes with cosine distance

Т а б л и ц а 3. Разрыв качества между фиксацией старых координат и полным переобучением пространства

Table 3. Quality gap between fixing old coordinates and full space retraining

Датасет	Расстояние	Разбиение	Размерность	ρ полного переобучения	ρ фиксации	Фиксация – полное переобучение
MEN	cosine	all	3	0.9038	0.7928	-0.1110
MEN	euclidean	all	3	0.9669	0.9109	-0.0560
SimLex-999	cosine	all	3	0.9570	0.9097	-0.0473
SimLex-999	euclidean	all	3	0.9867	0.9640	-0.0227
WordSim-353	cosine	all	3	0.9648	0.8302	-0.1346
WordSim-353	euclidean	all	3	0.9439	0.8973	-0.0466
MEN	cosine	old_new	3	0.7741	0.3925	-0.3816
MEN	euclidean	old_new	3	0.8809	0.5458	-0.3351
SimLex-999	cosine	old_new	3	0.9348	0.4678	-0.4670
SimLex-999	euclidean	old_new	3	0.9792	0.5035	-0.4757
WordSim-353	cosine	old_new	3	0.9428	0.4631	-0.4797
WordSim-353	euclidean	old_new	3	0.9457	0.7212	-0.2245

Наиболее заметное ухудшение при фиксации координат наблюдается не на общем разбиении all, а на old-new связях. Это объяснимо: all содержит значительную долю старой структуры, уже восстановленной на базовом этапе, тогда как old_new изолирует именно размещение новых объектов относительно старых (см. рис. 3).

Таблица 3 показывает, что разрыв особенно велик на old-new разбиении: для MEN с cosine-расстоянием он достигает -0.3816 , для WordSim-353 с cosine-расстоянием -0.4797 , а для SimLex-999 с euclidean-расстоянием -0.4757 . Следовательно, фиксация обеспечивает вычислительную простоту, но может создавать жёсткий геометрический барьер.

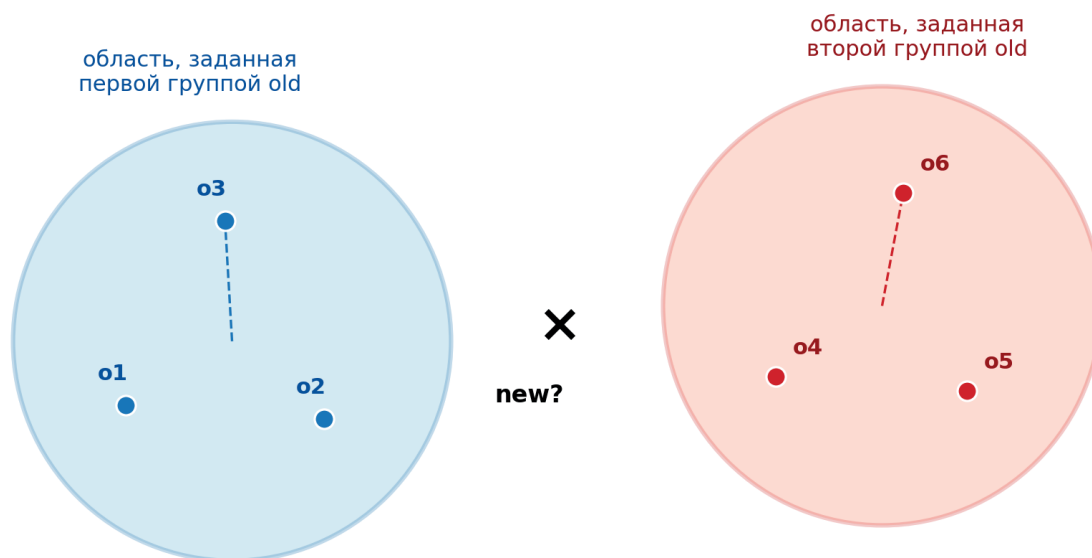
7 Почему качество для второй порции объектов ниже, чем при полной перестройке

Основное ограничение режима Фиксация старых координат является геометрическим. Когда старые координаты заморожены, новый объект должен одновременно удовлетворить всем ограничениям относительно уже расположенных old-то-

чек. Если эти ограничения несовместимы с текущей old-геометрией, то никакое положение new-точки в фиксированном пространстве не сможет восстановить все ранги.

Простейший пример можно описать через два непересекающихся круга. Пусть три old-точки задают для new-точки допустимую область расположения, которую можно аппроксимировать кругом: внутри этой области new-точка сохраняет нужные отношения расстояний к первой тройке old-точек. Другая тройка old-точек задаёт второй круг допустимых положений для той же new-точки. Если эти два круга не пересекаются, то не существует положения new-точки, которое одновременно удовлетворяет обоим наборам ограничений (см. рис. 4). При полном переобучении эта несовместимость может быть устранена за счёт небольшого сдвига old-точек или глобальной перестройки пространства. При фиксации такой возможности нет, поэтому часть old-new рангов неизбежно нарушается. Эта интерпретация согласуется с экспериментами. Разрыв между фиксацией и полным переобучением на разбиении all обычно меньше, чем на разбиении old_new, потому что all содержит значительную долю old-old структуры. Напротив, old_new является более строгой диагностикой инкрементального обновления.





Р и с. 4. Геометрическая причина ухудшения качества при замороженном old-пространстве: две группы old-точек задают для одной new-точки непересекающиеся области допустимого расположения

Fig. 4. Geometric reason for quality degradation with the frozen old space: two groups of old points define non-overlapping feasible regions for one new point

Вместе с тем полное переобучение пространства при появлении новых объектов иногда невозможно организационно или технически. Уже построенные представления могут использоваться в индексах поиска, пользовательских профилях, внешних базах признаков или внешних аналитических сервисах. В этих случаях даже небольшое изменение old-координат нарушает совместимость с ранее рассчитанными результатами. Поэтому возникает потребность в строгом расширении пространства, при котором старый базовый блок координат остаётся неизменным, а дополнительные степени свободы вводятся отдельно.

8 Расширение пространства: режим Расширение старых координат

Чтобы смягчить ограничение фиксированной старой геометрии, был рассмотрен режим Расширение старых координат. Его идея состоит в том, что пространство расширяется: исходный блок координат старых объектов сохраняется, но для old-объектов добавляется дополнительный обучаемый блок координат, а new-объекты получают координаты в полном расширенном пространстве. Тем самым алгоритм сохраняет старую геометрию как часть представления, но получает дополнительные степени свободы для согласования old-new ограничений.

Формально old-представление можно записать как

$$z_i^{\text{old}} = [x_i^{\text{old}}; h_i^{\text{old}}].$$

Здесь x_i^{old} – исходная базовая координата old-объекта, а h_i^{old} – дополнительный блок координат. Для new-объекта обучается полный вектор

$$z_j^{\text{new}} = [x_j^{\text{new}}; h_j^{\text{new}}].$$

Если h_i^{old} обучается без ограничений, модель получает возможность существенно деформировать old-объекты в дополнительных измерениях. Это улучшает согласование old-new ограничений, но создаёт риск неконтролируемого роста нормы extra-блока и ухудшает интерпретируемость инкрементального обновления: формально old-базис сохранён, но фактически old-объекты могут сильно разойтись в расширенных координатах.

Для масштабируемости инкрементального подхода принципиально важно не включать в обучение второго этапа чистые old-old пары как обычные обучающие ограничения. Если на каждом добавлении новых объектов заново использовать все old-old сравнения, число обрабатываемых четвёрок и объём пересчёта начинают расти как при полном перепостроении пространства. В таком случае инкрементальная процедура теряет своё основное преимущество: она перестаёт быть локальным обновлением и приближается по сложности к полному переобучению. Поэтому old-old структура должна сохраняться либо через фиксированный базовый блок, либо через регуляризацию, либо через ограниченный набор контрольных ограничений, но не через массовое повторное обучение на всех old-old парах.

Важно, что расширение пространства без дополнительных ограничений может ухудшать не только интерпретируемость, но и диагностическое качество старой части пространства. На втором этапе оптимизация строится по четвёрам, которые затрагивают новые объекты, тогда как среди обучающих четвёрок нет чистых сравнений old-old против old-old. Из-за этого дополнительный блок координат может изменять взаимные расстояния между старыми объектами в расширенном про-



пространстве без прямого штрафа за нарушение прежнего old-old порядка. Следовательно, old-old соотношения после второго этапа потенциально могут ухудшаться даже при приемлемом согласовании old-new ограничений.

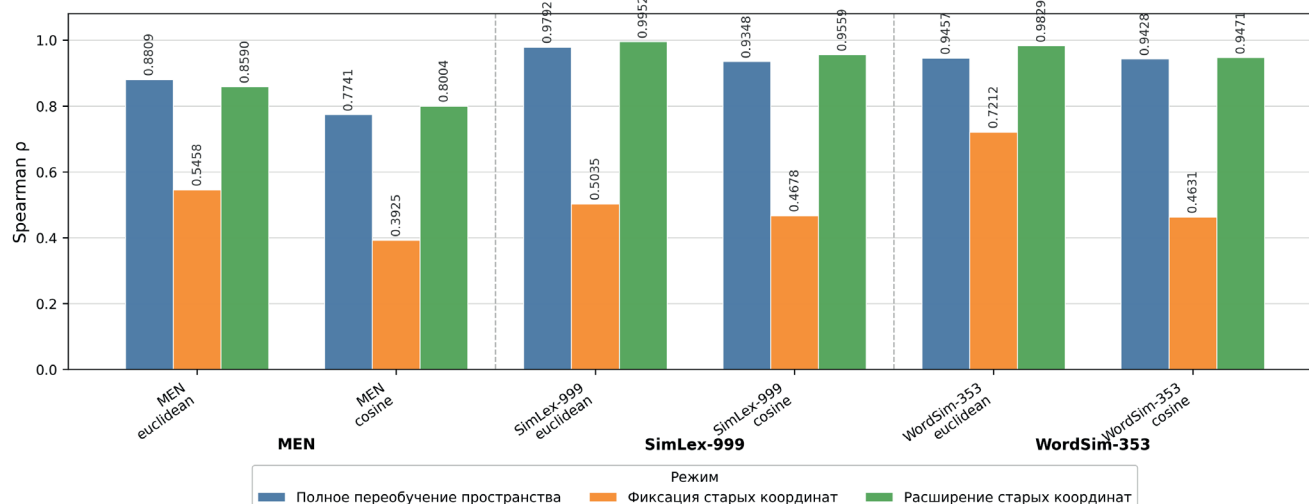
В сравнении используются четыре режима. Режим Только старые объекты строит в двухмерном пространстве базовую конфигурацию только по старым объектам и оценивается только на подмножестве old-old. Режим Полное переобучение про-

странства заново обучает все координаты для старых и новых объектов в трёхмерном пространстве и служит контрольным ориентиром. Режим Фиксация старых координат добавляет новые объекты при фиксированных старых координатах. Режим Расширение старых координат сохраняет исходные две координаты старых объектов, но добавляет им одну обучаемую дополнительную координату, новые объекты при этом обучаются в расширенном трёхмерном пространстве.

Таблица 4. Значения ρ Спирмена для сравнения режимов Расширение старых координат с Полное переобучение пространства для размерности 3
Table 4. Spearman ρ values for comparing the Expanded Old Coordinates mode with the Full Space Retraining mode for dimension 3

Датасет	Расстояние	Режим	old-old	old-new	new-new	all
MEN	euclidean	Полное переобучение пространства	0.9714	0.8809	0.7197	0.9669
MEN	euclidean	Фиксация старых координат	0.9295	0.5458	0.1255	0.9109
MEN	euclidean	Расширение старых координат	0.9301	0.8590	0.1172	0.9268
MEN	cosine	Полное переобучение пространства	0.9095	0.7741	0.6946	0.9038
MEN	cosine	Фиксация старых координат	0.8047	0.3925	0.1339	0.7928
MEN	cosine	Расширение старых координат	0.8429	0.8004	0.5523	0.8434
SimLex-999	euclidean	Полное переобучение пространства	0.9863	0.9792	—	0.9867
SimLex-999	euclidean	Фиксация старых координат	0.9848	0.5035	—	0.9640
SimLex-999	euclidean	Расширение старых координат	0.9720	0.9952	—	0.9728
SimLex-999	cosine	Полное переобучение пространства	0.9562	0.9348	—	0.9570
SimLex-999	cosine	Фиксация старых координат	0.9321	0.4678	—	0.9097
SimLex-999	cosine	Расширение старых координат	0.9107	0.9559	—	0.9154
WordSim-353	euclidean	Полное переобучение пространства	0.9484	0.9457	—	0.9439
WordSim-353	euclidean	Фиксация старых координат	0.9566	0.7212	—	0.8973
WordSim-353	euclidean	Расширение старых координат	0.9258	0.9829	—	0.9371
WordSim-353	cosine	Полное переобучение пространства	0.9702	0.9428	—	0.9648
WordSim-353	cosine	Фиксация старых координат	0.9590	0.4631	—	0.8302
WordSim-353	cosine	Расширение старых координат	0.9339	0.9471	—	0.9263

Значения ρ Спирмена на разбиении old-new для сравнения режимов (d=3)



Р и с. 5. Сравнение ρ Спирмена на связях old-new для режимов Полное переобучение пространства, Фиксация старых координат и Расширение старых координат для размерности 3

Fig. 5. Comparison of Spearman ρ on old-new links for the Full Space Retraining, Fixed Old Coordinates, and Expanded Old Coordinates modes for dimension 3



Результаты таблицы 4 показывают два разных эффекта. По общему разбиению all контрольный режим Полное переобучение пространства для размерности 3 является наиболее сильным: он лучше Полное переобучение пространства для размерности 2 во всех шести конфигурациях, а в прямом сравнении превосходит Расширение старых координат также во всех шести конфигурациях. Однако на связях old-new инкрементальный режим Расширение старых координат остаётся конкурентоспособным и превосходит режим Полное переобучение пространства в пяти из шести конфигураций. Следовательно, полная трёхмерная перестройка даёт более высокую глобальную согласованность, но расширение старых координат особенно эффективно именно для согласования новых объектов со старой частью пространства.

9 Регуляризация нормы дополнительного блока старых координат

Для контроля дополнительной деформации был проведён эксперимент с L2-регуляризацией дополнительного блока старых координат. К основной функции потерь на четвёрках добавлялся штраф:

$$L_{\text{total}} = L_{\text{quad}} + \lambda \cdot 1/|O| \sum_{i \in O} \|h_i^{\text{old}}\|_2^2.$$

Здесь O – множество old-объектов, h_i^{old} – дополнительный блок old-объекта, а $\lambda = 10^{-3}$. Такая форма соответствует общему принципу L2-регуляризации: добавление штрафа к целевой функции ограничивает норму параметров и уменьшает свободу модели [21], [22].

Результаты эксперимента показывают, что Расширение старых координат с L2-регуляризацией в большинстве конфигураций почти не меняет, а в MEN cosine old_new заметно улучшает ранговое соотношение относительно варианта без регуляризации, что видно из таблицы 5.

Главный эффект регуляризации проявляется не только в сохранении рангового показателя качества, но прежде всего в стабилизации геометрии расширенного пространства. Наиболее выраженный эффект наблюдается в конфигурациях, где L2-штраф существенно ограничивает величину дополнительного блока координат: на SimLex-999 при cosine-расстоянии норма уменьшается на 40%, а на WordSim-353 при euclidean-расстоянии – на 39%. Следовательно, L2-регуляризация делает расширение пространства более контролируемым: модель сохраняет или лишь умеренно изменяет качество рангового согласования, но достигает результата при заметно меньшей деформации старых объектов в дополнительных координатах.

Таблица 5. Влияние L2-регуляризации для режима Расширение старых координат на Spearman ρ
Table 5. Effect of L2 regularization for the Expanded Old Coordinates mode on Spearman ρ

Датасет	Расстояние	Разбиение	Расширение старых координат без регуляризации	Расширение старых координат L2, $\lambda=0.001$	Δ L2
MEN	euclidean	all	0.9268	0.9318	0.0050
MEN	euclidean	old_new	0.8590	0.8628	0.0038
MEN	euclidean	old_old	0.9301	0.9340	0.0039
MEN	cosine	all	0.8434	0.8492	0.0058
MEN	cosine	old_new	0.8004	0.8576	0.0572
MEN	cosine	old_old	0.8429	0.8471	0.0042
SimLex-999	euclidean	all	0.9728	0.9736	0.0008
SimLex-999	euclidean	old_new	0.9952	0.9958	0.0006
SimLex-999	euclidean	old_old	0.9720	0.9728	0.0008
SimLex-999	cosine	all	0.9154	0.9160	0.0006
SimLex-999	cosine	old_new	0.9559	0.9672	0.0113
SimLex-999	cosine	old_old	0.9107	0.9116	0.0009
WordSim-353	euclidean	all	0.9371	0.9375	0.0004
WordSim-353	euclidean	old_new	0.9829	0.9830	0.0001
WordSim-353	euclidean	old_old	0.9258	0.9264	0.0006
WordSim-353	cosine	all	0.9263	0.9233	-0.0030
WordSim-353	cosine	old_new	0.9471	0.9428	-0.0043
WordSim-353	cosine	old_old	0.9339	0.9304	-0.0035



Таблица 6. Снижение величины нормы дополнительного блока старых координат при L2-регуляризации
Table 6. Reduction in the magnitude of the additional block of old coordinates under L2 regularization

Датасет	Расстояние	Норма дополнительного блока старых координат без регуляризации	Норма дополнительного блока старых координат с L2	Абсолютное снижение	Относительное снижение
MEN	euclidean	77.0765	67.7518	9.3246	12.0979%
MEN	cosine	25.8606	20.8432	5.0174	19.4019%
SimLex-999	euclidean	77.5268	65.1849	12.3419	15.9195%
SimLex-999	cosine	63.9653	38.3629	25.6024	40.0255%
WordSim-353	euclidean	36.7884	22.3470	14.4414	39.2554%
WordSim-353	cosine	16.9975	16.3547	0.6428	3.7816%

10 Обсуждение

Полученные результаты позволяют разделить две задачи, которые часто смешиваются при обсуждении инкрементального GNMDs. Первая задача – оперативно включить новые объекты без полной перестройки пространства, сохранив уже построенные координаты старых объектов. Её решает режим Фиксация старых координат, но показатель качества на old-new ограничен геометрической совместимостью новых требований с фиксированным old-пространством. Вторая задача – сохранить инкрементальность, но разрешить ограниченную адаптацию старого пространства. Её решает режим Расширение старых координат: старый базовый блок остаётся неизменным, но появляется дополнительный обучаемый блок, который может компенсировать несовместимость ограничений.

Сравнение Полного переобучения пространства с Расширением старых координат в общей размерности 3 показывает различие между глобальным и инкрементальным критериями качества. По all полная перестройка выигрывает во всех шести конфигурациях, поскольку оптимизирует все координаты всех объектов заново. На old-new связях, напротив, Расширение старых координат выигрывает в пяти из шести конфигураций (см. таблицу 7). Поэтому результаты следует интерпретировать не как однозначную замену инкрементального режима полной перестройкой, а как разделение ролей: Полное переобучение пространства задаёт контрольный ориентир качества по общей реконструкции, а Расширение старых координат остаётся содержательным вариантом для включения новых объектов при сохранении исходной базы.

Таблица 7. Прямое сравнение режимов Полное переобучение пространства и Расширение старых координат в общей размерности 3
Table 7. Direct comparison of the Full Space Retraining and Expanded Old Coordinates modes in total dimension 3

Датасет	Расстояние	Полное переобучение пространства d=3, all	Расширение старых координат d=3, all	Δ полное переобучение - расширение, all	Полное переобучение пространства d=3, old_new	Расширение старых координат d=3, old_new	Δ полное переобучение - расширение, old_new
MEN	euclidean	0.967	0.927	0.040	0.881	0.859	0.022
MEN	cosine	0.904	0.843	0.060	0.774	0.800	-0.026
SimLex-999	euclidean	0.987	0.973	0.014	0.979	0.995	-0.016
SimLex-999	cosine	0.957	0.915	0.042	0.935	0.956	-0.021
WordSim-353	euclidean	0.944	0.937	0.007	0.946	0.983	-0.037
WordSim-353	cosine	0.965	0.926	0.038	0.943	0.947	-0.004

Без регуляризации Расширение старых координат может быть слишком свободным: дополнительный блок старых координат способен расти и фактически переносить значительную часть решения в новые измерения. С точки зрения Spearman ρ это иногда допустимо, но с точки зрения инкрементальной интерпретации нежелательно, поскольку old-old соотношения могут нарушаться без явного контроля. Регуляризация дополнительного блока старых координат вводит управляемый компромисс между качеством ранговой реконструкции и степенью сохранения old-old соотношений. В проведённом эксперименте регуляризация оказалась достаточно мягкой, чтобы не разрушить выигрыш по

old_new, но достаточно заметной, чтобы уменьшить норму дополнительного блока.

С методологической точки зрения результаты также показывают, что инкрементальное GNMDs следует оценивать несколькими показателями качества, а не только общим Spearman ρ на всех парах. Разбиение all может скрывать деградацию old-new или old-old структуры, поскольку объединяет разные типы пар. Поэтому для практических систем важны отдельные оценки old-old, old-new и new-new, а также контроль нормы дополнительных блоков.



11 Заключение

В работе рассмотрена инкрементальная версия нейросетевого GNMDs, в которой пространство векторных представлений обновляется после появления новых объектов. В inductive-схеме на первом этапе используются только old-old ограничения, а на втором этапе появляются связи новых объектов. Эксперименты с режимами Только старые объекты, Полное переобучение пространства и Фиксация старых координат показывают, что полное переобучение пространства остаётся контрольным ориентиром показателя качества, а фиксация старых координат приводит к заметному ухудшению на old-new парах. Геометрическая причина этого ухудшения состоит в том, что фиксированное old-пространство может задавать несовместимые ограничения на положение new-точек. Расширение старых координат добавляет необходимые степени свободы и существенно улучшает old-new качество. Однако без регуляризации такой режим может нарушать old-old соотношения и неконтролируемо деформировать старые объекты в дополни-

тельных измерениях. L2-регуляризация ограничивает величину нормы дополнительного блока координат, помогает сохранять old-old соотношения и уменьшает деформацию дополнительного блока на 40% для SimLex-999 при cosine-расстоянии и на 39% для WordSim-353 при euclidean-расстоянии, почти не меняя ранговый показатель качества.

Благодарность

Автор выражает благодарность А.И. Майсурадзе за идею общей схемы реализации неинкрементального подхода к решению задачи GNMDs.

Acknowledgement

The author expresses gratitude to A.I. Maisuradze for the idea of a general scheme for implementing a non-incremental approach to solving the GNMDs problem.

References

- [1] Shepard R.N. The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. I. *Psychometrika*. 1962;27(2):125-140. <https://doi.org/10.1007/BF02289630>
- [2] Kruskal J.B. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*. 1964;29(1):1-27. <https://doi.org/10.1007/BF02289565>
- [3] Kruskal J.B. Nonmetric Multidimensional Scaling: A Numerical Method. *Psychometrika*. 1964;29(2):115-129. <https://doi.org/10.1007/BF02289694>
- [4] Agarwal S., Wills J., Cayton L., Lanckriet G., Kriegman D., Belongie S. Generalized Non-metric Multidimensional Scaling. In: Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, PMLR. 2007;2:11-18. Available at: <https://proceedings.mlr.press/v2/agarwal07a.html> (accessed 20.02.2026).
- [5] Schroff F., Kalenichenko D., Philbin J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In: Proceedings of CVPR, 2015. p. 815-823. Available at: <https://doi.org/10.1109/CVPR.2015.7298682> (accessed 20.02.2026).
- [6] Weinberger K.Q., Saul L.K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Journal of Machine Learning Research*. 2009;10(9):207-244. Available at: <https://jmlr.org/papers/v10/weinberger09a.html> (accessed 20.02.2026).
- [7] Harris Z.S. Distributional Structure. *WORD*. 1954;10(2-3):146-162. <https://doi.org/10.1080/00437956.1954.11659520>
- [8] Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781*, 2013. <https://doi.org/10.48550/arXiv.1301.3781>
- [9] Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionality. In: Advances in Neural Information Processing Systems (NIPS 2013). Vol. 26. Curran Associates, Inc.; 2013. 9 p. Available at: <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality> (accessed 20.02.2026).
- [10] Pennington J., Socher R., Manning C. GloVe: Global Vectors for Word Representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014. p. 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [11] Finkelstein L., Gabrilovich E., Matias Y., Rivlin E., Solan Z., Wolfman G., Ruppin E. Placing Search in Context: The Concept Revisited. *ACM Transactions on Information Systems*. 2002;20(1):116-131. <https://doi.org/10.1145/503104.503110>
- [12] Hill F., Reichart R., Korhonen A. SimLex-999: Evaluating Semantic Models With Genuine Similarity Estimation. *Computational Linguistics*. 2015;41(4):665-695. https://doi.org/10.1162/COLI_a_00237
- [13] Bruni E., Tran N.K., Baroni M. Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research*. 2014;49:1-47. <https://doi.org/10.1613/jair.4135>
- [14] Terada Y., von Luxburg U. Local Ordinal Embedding. In: Proceedings of the 31st International Conference on Machine Learning, PMLR. 2014;32(2):847-855. Available at: <https://proceedings.mlr.press/v32/terada14.html> (accessed 20.02.2026).
- [15] Jamieson K.G., Nowak R.D. Low-dimensional embedding using adaptively selected ordinal data. In: 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton). Monticello, IL, USA: IEEE Press; 2011. p. 1077-1084. <https://doi.org/10.1109/Allerton.2011.6120287>
- [16] Nocedal J., Wright S.J. Numerical Optimization. *Springer Series in Operations Research and Financial Engineering*. Springer New York, NY; 2006. 664 p. <https://doi.org/10.1007/978-0-387-40065-5>



- [17] Dauphin Y.N., Pascanu R., Gulcehre C., Cho K., Ganguli S., Bengio Y. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In: Advances in Neural Information Processing Systems (NIPS 2014). Vol. 27. Curran Associates, Inc.; 2014. 9 p. Available at: <https://papers.nips.cc/paper/5486-identifying-and-attacking-the-saddle-point-problem-in-high-dimensional-non-convex-optimization> (accessed 20.02.2026).
- [18] Robbins H., Monro S. A Stochastic Approximation Method. *Annals of Mathematical Statistics*. 1951;22(3):400-407. <https://doi.org/10.1214/aoms/1177729586>
- [19] Bottou L. Large-Scale Machine Learning with Stochastic Gradient Descent. In: Lechevallier Y., Saporta G. (eds) Proceedings of COMPSTAT'2010. Physica-Verlag HD; 2010. p. 177-186. https://doi.org/10.1007/978-3-7908-2604-3_16
- [20] Ghadimi S., Lan G. Stochastic First- and Zeroth-Order Methods for Nonconvex Stochastic Programming. *SIAM Journal on Optimization*. 2013;23(4):2341-2368. <https://doi.org/10.1137/120880811>
- [21] Tikhonov A.N., Arsenin V.Y. Solutions of Ill-posed Problems. V.H. Winston and Son, Inc.; 1977. 258 p.
- [22] Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press; 2016. Available at: <http://www.deeplearningbook.org> (accessed 20.02.2026).
- [23] Hinton G.E., Salakhutdinov R.R. Reducing the Dimensionality of Data with Neural Networks. *Science*. 2006;313(5786):504-507. <https://doi.org/10.1126/science.1127647>
- [24] Belkin M., Niyogi P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems (NIPS 2001). Vol. 14. MIT Press; 2001. 7 p. Available at: <https://proceedings.neurips.cc/paper/2001/hash/f106b7f99d2cb30c3db1c3cc0fde9ccb-Abstract.html> (accessed 20.02.2026).
- [25] GNMDs Implementation Repository: obj2vec/gnmds. Available at: <https://github.com/obj2vec/gnmds> (accessed 20.02.2026).

Поступила 20.02.2026; одобрена после рецензирования 24.03.2026; принята к публикации 10.04.2026.

Submitted 20.02.2026; approved after reviewing 24.03.2026; accepted for publication 10.04.2026.

Об авторе:

Колосов Алексей Михайлович, научный сотрудник кафедры математической теории интеллектуальных систем механико-математического факультета, Московский государственный университет имени М.В. Ломоносова (119991, Российская Федерация, г. Москва, ГСП-1, Ленинские горы, д. 1), **ORCID:** <https://orcid.org/0000-0002-9474-9666>, aleksei.kolosov@math.msu.ru

Автор прочитал и одобрил окончательный вариант рукописи.

About the author:

Alexey M. Kolosov, Researcher of the Chair of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University (1 Leninskie gory, Moscow 119991, GSP-1, Russian Federation), **ORCID:** <https://orcid.org/0000-0002-9474-9666>, aleksei.kolosov@math.msu.ru

The author has read and approved the final manuscript.

