

УДК 004.85:004.912:519.61

АЛГОРИТМ СОГЛАСОВАНИЯ СЕМАНТИЧЕСКИХ ПРОСТРАНСТВ РАЗНОРОДНЫХ ИНТЕЛЛЕКТУАЛЬНЫХ МОДЕЛЕЙ НА ОСНОВЕ ЯКОРНОГО ОРТОГОНАЛЬНОГО ОТОБРАЖЕНИЯ

Некрасов А. Н., гр. ИИБ-231, III курс,
Кузбасский государственный технический университет имени Т.Ф. Горбачева
г. Кемерово

Современные интеллектуальные системы редко строятся на единственной модели. Поисковый модуль, классификатор намерений, генеративная языковая модель и специализированный отраслевой кодировщик обучаются независимо, в разное время и на разных корпусах, а результатом каждого из них является собственное векторное представление объектов предметной области. Такие представления образуют семантические пространства, в которых близость векторов отражает смысловую близость исходных объектов [1]. Проблема состоит в том, что пространства разных моделей несопоставимы: одинаковые по смыслу объекты получают координаты, между которыми отсутствует прямое соответствие. Прямое сравнение или объединение эмбедингов приводит к деградации качества поиска и переноса знаний.

Причина несопоставимости носит алгоритмический характер. Обучение представлений инвариантно относительно широкого класса преобразований: функция потерь трансформерной модели зависит от скалярных произведений и нормировок, а не от конкретной системы координат [2]. Поэтому близость исходных данных при обучении не гарантирует совпадения ориентации полученных пространств. В результате две модели могут формировать представления с близкой геометрией, но различным расположением координатных осей, включая поворот или отражение конфигурации точек [3]. Это позволяет рассматривать часть наблюдаемого расхождения между пространствами как следствие выбора базиса, а не различий в представляемой семантике. В рамках данной работы предполагается, что указанная составляющая может быть компенсирована посредством единственного линейного преобразования.

Работа направлена на разработку метода приведения семантических пространств различных интеллектуальных моделей к согласованному представлению. В качестве исходной информации используется ограниченное множество пар-якорей, по которому определяется преобразование между пространствами. При этом требуется минимально изменять геометрические свойства исходных представлений и обеспечивать возможность последующей независимой оценки полученного соответствия. Задачи работы: формализовать критерий согласования; получить замкнутое решение; предложить критерий отбора

4.21.2

якорей; оценить вычислительную стоимость; выполнить расчётно-алгоритмическую проверку.

Введём обозначения. Пусть ϕ и ψ – кодировщики моделей A и B , отображающие объект x в векторы размерности d . Для конечного якорного набора A из m пар объектов, смысловая эквивалентность которых установлена заранее, строятся матрицы X_A и Y_A размера $m \times d$, строки которых нормированы на единичную длину. Требуется найти оператор W , переводящий пространство модели A в пространство модели B . Критерий согласования записывается как задача о наименьшем квадратичном расхождении при ограничении ортогональности:

$$W^* = \arg \min_{W^T W = I_d} \|X_A W - Y_A\|_F^2 \quad (1)$$

Ограничение $W^T W = I_d$ принципиально: ортогональный оператор сохраняет длины векторов, углы между ними и, следовательно, все косинусные близости внутри переносимого пространства. Это означает, что согласование не может «подогнать» результат за счёт искажения внутренней структуры – оно способно лишь повернуть и отразить пространство целиком. Задача (1) известна как ортогональная задача Прокруста и имеет замкнутое решение через сингулярное разложение матрицы взаимной ковариации якорей [4]:

$$M = X_A^T Y_A = U \Sigma V^T, \quad W^* = U V^T \quad (2)$$

В выражении (2) M – матрица размера $d \times d$, U и V – ортогональные матрицы левых и правых сингулярных векторов, Σ – диагональная матрица сингулярных чисел. Полученный оператор W^* ортогонален по построению, не требует итерационной настройки и не содержит гиперпараметров, что выгодно отличает его от градиентного обучения согласующего слоя [5]. Устойчивость решения определяется обусловленностью M : при вырожденном спектре малые возмущения якорей приводят к заметному повороту соответствующих сингулярных направлений [6].

Качество согласования критически зависит от состава якорного набора. Пары, в которых объект модели A эквивалентен сразу нескольким объектам модели B , вносят в матрицу M противоречивый вклад. Для отбора вводится оценка надёжности якоря, сопоставляющая близость внутри пары с близостью к её локальному окружению:

$$s_i = \langle x_i, y_i \rangle - \frac{1}{k} \sum_{j \in N_k(i)} \langle x_i, y_j \rangle \quad (3)$$

Здесь x_i и y_i – нормированные векторы i -й пары, $N_k(i)$ – множество k ближайших к x_i объектов в пространстве модели B , исключая парный. Величина s_i положительна, когда парное соответствие выражено сильнее фонового, и близка к нулю для неоднозначных пар. В якорный набор включаются пары с $s_i \geq \tau$; порог τ задаётся квантилью распределения оценок, что делает критерий независимым от абсолютного масштаба близостей [7].

После построения оператора требуется контроль на объектах, не входивших в якорный набор. Локальная невязка согласования определяется как косинусное расхождение перенесённого и целевого векторов:

$$\delta(x) = 1 - \frac{\langle \varphi(x)W, \psi(x) \rangle}{\|\varphi(x)W\| \cdot \|\psi(x)\|} \quad (4)$$

Помимо усреднённой невязки (4) применяется ранговый контроль: для контрольной выборки вычисляется доля объектов, для которых целевой вектор попадает в первую позицию либо в первые пять позиций ранжирования по косинусной близости [8]. Ранговые показатели чувствительнее среднего косинуса, поскольку отражают не абсолютную величину близости, а её преимущество над конкурирующими объектами.

Вычислительная стоимость алгоритма складывается из формирования ковариационной матрицы, сингулярного разложения и однократного переноса всей базы из n объектов, а постоянная память определяется хранением оператора:

$$T(d, m) = 2md^2 + \alpha d^3 + 2nd^2, S_W = bd^2 \quad (5)$$

В (5) b – число байт на элемент, α – эмпирический коэффициент реализации сингулярного разложения (для двусторонних алгоритмов с накоплением обоих семейств векторов $\alpha \approx 21$). Существенно, что стоимость не зависит от объёма обучающих корпусов исходных моделей: алгоритм работает только с их выходами. Это позволяет согласовывать пространства закрытых моделей, доступных исключительно через программный интерфейс кодирования, включая кодировщики предложенческого уровня [9].

Таблица 1 – Классы согласующих отображений и условия их применения

Класс отображения	Ограничение на оператор	Условие применения
А. Ортогональное	$W^T W = I$; сохраняются длины и углы	Пространства близки по геометрии; якорей мало ($m \approx d$)
В. Аффинное	W произвольна, добавлен сдвиг b	Различаются масштабы осей; якорей достаточно ($m \gg d$)
С. Кусочно-нелинейное	Ортогональная база и лёгкая нелинейная поправка	Устойчивая нелинейная невязка после этапа А

Выбор класса отображения определяется соотношением между числом якорных пар и размерностью пространства. Ортогональное преобразование содержит $(d(d-1)/2)$ независимых параметров и при наличии достаточного числа линейно независимых якорных векторов может быть однозначно восстановлено без введения дополнительных параметров масштаба и сдвига. Аффинное отображение имеет существенно больше степеней свободы и, в отличие от

4.21.4

ортогонального, не обеспечивает сохранение норм и косинусных углов между векторами. Класс С применяется только как уточнение поверх результата класса А и лишь при статистически значимой систематической невязке.

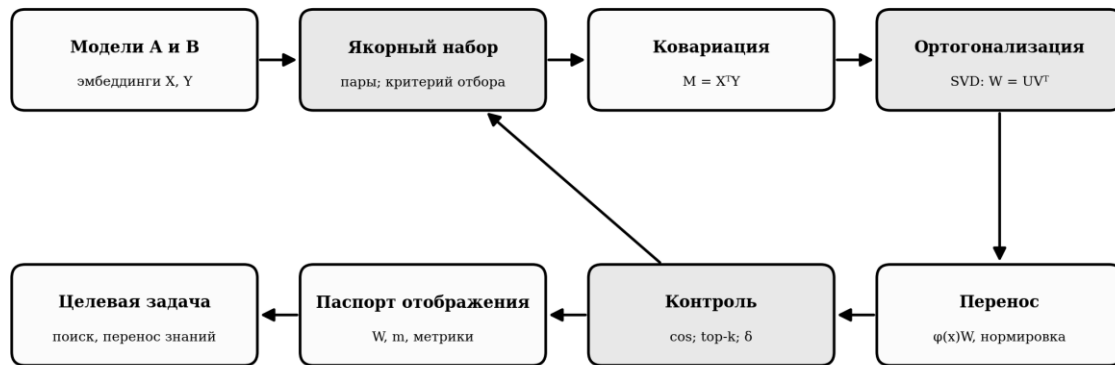


Рисунок 1 – Схема построения и контроля согласующего оператора семантических пространств

Расчётно-алгоритмическая проверка выполнена на контролируемой модели данных, что позволяет отделить свойства самого алгоритма от свойств конкретных предобученных кодировщиков. Сформировано $n = 20000$ нормированных векторов размерности $d = 256$, задан эталонный ортогональный оператор Q , а второе пространство получено как образ первого с аддитивным гауссовым возмущением уровня $0,15$ с последующей перенормировкой. Возмущение моделирует ту часть расхождения, которая не объясняется поворотом базиса. Оператор восстанавливался по случайной подвыборке якорей, контроль проводился на 2000 объектах, не участвовавших в построении. Вычисления выполнены средствами библиотеки линейной алгебры NumPy в двойной точности [10].

Таблица 2 – Зависимость качества согласования от объёма якорного набора

Число якорей m	Норма $\ W^T W - I\ $	Средний косинус	top-1, %	top-5, %
125	$4,4 \cdot 10^{-14}$	0,071	0,8	3,2
250	$5,8 \cdot 10^{-14}$	0,109	6,2	14,8
500	$5,9 \cdot 10^{-14}$	0,169	25,2	46,8
1000	$6,2 \cdot 10^{-14}$	0,236	65,8	84,8
2000	$6,0 \cdot 10^{-14}$	0,309	94,3	99,1
4000	$6,4 \cdot 10^{-14}$	0,348	99,0	99,9

Отклонение от ортогональности во всех запусках не превышает $6,4 \cdot 10^{-14}$ по норме Фробениуса, то есть находится на уровне машинной точности

4.21.5

двойной точности и не зависит от числа якорей. Это подтверждает, что ограничение задачи (1) выполняется решением (2) точно, а не приближённо. Качество же согласования зависит от m резко нелинейно: при $m = 500$ (около двух якорей на измерение) доля точных попаданий составляет 25,2 %, при $m = 1000$ – 65,8 %, а при $m = 2000$ достигает 94,3 %. Дальнейшее увеличение якорного набора вчетверо добавляет менее пяти процентных пунктов.

Практически значим переход в области $m \approx 4d \dots 8d$: именно здесь оценка ковариационной матрицы становится статистически устойчивой. Средний косинус при этом остаётся умеренным (0,309 при $m = 2000$), поскольку возмущение принципиально неустранимо ортогональным оператором, однако ранговые показатели близки к предельным. Это подтверждает тезис о разной информативности метрик: усреднённая невязка (4) характеризует остаточный шум, а ранговый контроль – практическую пригодность согласования в задачах поиска.

Таблица 3 – Ресурсные характеристики алгоритма при $m = 5000$ и $n = 10^6$

Размерность d	Память W , МиБ	Ковариация, ГФлоп	SVD, ГФлоп	Доля SVD, %
384	0,56	1,47	1,19	45
768	2,25	5,90	9,51	62
1024	4,00	10,49	22,55	68
1536	9,00	23,59	76,10	76

Расчёт по формуле (5) показывает, что постоянная память оператора остаётся малой даже для больших размерностей: 9 МиБ при $d = 1536$. Вычислительная стоимость смещается в сторону сингулярного разложения при росте d – его доля увеличивается с 45 % при $d = 384$ до 76 % при $d = 1536$, что соответствует кубическому росту против квадратичного у ковариационной части. Отсюда следует практический вывод: для высокоразмерных пространств оправдано предварительное понижение размерности с последующим согласованием в редуцированном базисе, если контроль подтверждает сохранение ранговых показателей. Для баз объёмом 10^6 и выше однократный перенос удобно совмещать с перестроением индекса приближённого поиска [11].

Границы применимости метода определяются исходной гипотезой. Алгоритм согласует пространства, различающиеся выбором базиса, и не способен восстановить соответствие там, где модели кодируют семантически различные признаки: если модель B не различает объекты, разделяемые моделью A , никакое ортогональное преобразование этого различия не создаст. Диагностическим признаком служит одновременно высокое отклонение (4) и низкие ранговые показатели при заведомо достаточном якорном наборе; для независимой оценки степени структурного подобия пространств применимы меры типа

центрированного ядрового выравнивания [12]. Второе ограничение – требование доверия к якорям: ошибочные пары смещают оценку ковариации, и критерий (3) снижает, но не устраняет этот риск.

Полученное преобразование предлагается сопровождать формализованным паспортом согласования. В нём фиксируются размерности обоих пространств, характеристики якорного набора, используемый порог τ , найденная матрица W и результаты контрольных измерений, выполненных на независимой выборке. Наличие такого описания позволяет воспроизводить процедуру и контролировать изменение качества при замене или обновлении одной из исходных моделей. Повторное вычисление контрольных показателей на неизменной выборке в этом случае позволяет обнаружить изменение положения представлений без необходимости повторного обучения процедуры согласования.

В работе рассмотрен алгоритм согласования семантических пространств разнородных интеллектуальных моделей, в котором объединены ортогональная постановка преобразования, отбор якорных пар с учётом надёжности локального соответствия и двухступенчатая процедура контроля результата. В вычислительном эксперименте полученная матрица удовлетворила заданному условию ортогональности, а доля точных соответствий на контрольной выборке составила 94,3 %. Установлено также, что для достижения указанного результата достаточно якорного набора объёмом порядка восьми пар на одну размерность пространства. Оценка вычислительных затрат показывает возможность включения процедуры согласования в регулярный цикл сопровождения многомодельной интеллектуальной системы.

Список литературы:

1. Mikolov T., Le Q. V., Sutskever I. Exploiting Similarities Among Languages for Machine Translation // arXiv. – 2013. – arXiv:1309.4168. – DOI: 10.48550/arXiv.1309.4168.
2. Vaswani A., Shazeer N., Parmar N. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.
3. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. – 2019. – P. 4171–4186. – DOI: 10.18653/v1/N19-1423.
4. Schönemann P. H. A Generalized Solution of the Orthogonal Procrustes Problem // Psychometrika. – 1966. – Vol. 31, No. 1. – P. 1–10. – DOI: 10.1007/BF02289451.
5. Xing C., Wang D., Liu C., Lin Y. Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation // Proceedings of NAACL-HLT. – 2015. – P. 1006–1011. – DOI: 10.3115/v1/N15-1104.

6. Golub G. H., Van Loan C. F. Matrix Computations. – 4th ed. – Baltimore: Johns Hopkins University Press, 2013. – 756 p.
7. Artetxe M., Labaka G., Agirre E. A Robust Self-Learning Method for Fully Unsupervised Cross-Lingual Mappings of Word Embeddings // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. – 2018. – P. 789–798. – DOI: 10.18653/v1/P18-1073.
8. Conneau A., Lample G., Ranzato M., Denoyer L., Jégou H. Word Translation Without Parallel Data // International Conference on Learning Representations. – 2018. – URL: <https://openreview.net/forum?id=H196sainb> (дата обращения: 31.08.2026).
9. Reimers N., Gurevych I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks // Proceedings of EMNLP-IJCNLP. – 2019. – P. 3982–3992. – DOI: 10.18653/v1/D19-1410.
10. Harris C. R., Millman K. J., van der Walt S. J. et al. Array Programming with NumPy // Nature. – 2020. – Vol. 585. – P. 357–362. – DOI: 10.1038/s41586-020-2649-2.
11. Johnson J., Douze M., Jégou H. Billion-Scale Similarity Search with GPUs // IEEE Transactions on Big Data. – 2021. – Vol. 7, No. 3. – P. 535–547. – DOI: 10.1109/TBDATA.2019.2921572.
12. Kornblith S., Norouzi M., Lee H., Hinton G. Similarity of Neural Network Representations Revisited // Proceedings of the 36th International Conference on Machine Learning. – 2019. – P. 3519–3529.