

Выбор тензорных представлений для прогнозирования по мультимодальным измерениям

Алсаханова Н. Ю., Стрижов В. В.

alsahanova.nyu@phystech.edu, strijov@phystech.edu

В данной работе решается задача прогнозирования по измерениям с разнородных источников путем выбора оптимального тензорного представления исходных данных. Оптимальное представление предлагается получать согласованием мультимодальных измерений. Далее строится прогностическая модель путем согласования независимых и целевой переменных.

Ключевые слова: *декодирование сигналов, электрокортикограмма, метод частичных наименьших квадратов, тензорный метод главных компонент.*

1 Введение

Исходные данные в области декодирования сигналов [1], [2], [3] являются многомерными и сильно избыточными. В следствие чего модели, построенные на таких данных, неустойчивы. Так же значительно увеличивается время вычислений. Для решения этих проблем применяют модели отбора признаков [4], [5] и снижения размерности [6], [27].

Часто в задаче декодирования имеется несколько входных сигналов, мультимодальные измерения. Например, в задаче восстановления траектории движения руки имеется 32 входных сигнала электрокортикограмм [2]. Для решения задачи с мультимодальными измерениями применяются две различные группы подходов, описанные в [7]. Одна группа состоит из методов, переводящих мультимодальные измерения в одно низкоразмерное представление. Другая группа методов путем согласования мультимодальных измерений проецирует их в низкоразмерное пространство, сохраняя их количество. Например, метод главных компонент PCA и его тензорные модификации, а так же автоэнкодеры. Применяется канонический корреляционный анализ (CCA) и некоторые его модификации [8], [9], [10]. Так же согласование мультимодальных измерений проводится путем применения метода частичных наименьших квадратов (PLS) и его разновидностей [11], [12].

При прогнозировании временных рядов исходные данные имеют высокую размерность, так же в пространствах целевой и независимой переменных есть скрытые зависимости. Чересмерно высокая размерность пространств и наблюдаемая множественная корреляция приводит к неустойчивости прогностической модели. В данном случае регрессионная модель, основанная на методе частичных наименьших квадратов широко используется для восстановления зависимостей между наборами данных [13], [14], [15], [16]. Для построения устойчивой прогностической модели так же используют метод отбора признаков, основанный на задаче квадратичного программирования (QPFS) [3]. Данный метод

максимизирует релевантность и минимизирует избыточность признаков. В статье [5] показано, что метод QPFS превосходит многие существующие методы отбора признаков для задачи одномерной регрессии.

Все вышеуказанные методы используют матричное представление данных для прогнозирования временных рядов. Но в задаче с мультимодальными измерениями данные представляются в тензорном виде. Можно решать задачу путем перехода от тензоров к матрицам, но это снижает интерпретируемость и увеличивает избыточность данных. Поэтому для работы с тензорными данными предлагались различные модели [17], [18]. Были предложены тензорные модификации метода главных компонент (tensor KPCA и MPCA) [19], [20], канонического корреляционного анализа (tensor CCA и deep tensor CCA) [21], [9] и метода частичных наименьших квадратов (HOPLS) [22]. В статьях [23], [24] были рассмотрены нелинейные версии тензорного метода частичных наименьших квадратов. В статье [27] был применен автоэнкодер, учитывающий взаимосвязь между исходными данными и целевой переменной в задаче декодирования траектории руки по ЭКоГ данным.

В данной работе предлагается модифицировать предложенный в статье [27] автоэнкодер таким образом, чтобы данные на выходе энкодера оставались в тензорном виде. Более того предлагается новая функция потерь для такого автоэнкодера, который будет учитывать взаимную корреляцию между столбцами нового признакового пространства, полученного с помощью энкодера.

2 Постановка задачи

В данной главе ставится задача декодирования временных рядов. Приводится обзор стандартных методов анализа временных рядов. Ставится задача построения оптимальной линейной регрессионной модели декодирования. Приведен обзор методов снижения размерности пространства, их обобщений и модификаций.

2.1 Предлагаемая модель

Определение 1 *Временным рядом называется функция дискретного аргумента $\mathbf{s}(t)$, сопоставляющая отчетам времени $t_i \in \mathcal{T}$ вектор значения измеряемых переменных $\mathbf{s}(t_i) = \mathbf{s}_i \in \mathbb{R}^M$.*

Определение 2 *Пусть заданы набор временных рядов $\{\mathbf{s}_n(t)\}_{n=1}^{N_s}$ и набор целевых временных рядов $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$. Задача восстановления значений $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$ по предыдущим значениям $\{\mathbf{s}_n(t)\}_{n=1}^{N_s}$ называется задачей декодирования временных рядов $\{\mathbf{y}_n(t)\}_{n=1}^{N_y}$.*

Пусть из временных рядов $\mathbf{s}(t)$, $\mathbf{y}(t)$ составлена выборка $(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$:

$$\underline{\mathbf{X}} \in \mathbb{R}^{M \times n_1 \times \dots \times n_D}, \quad \underline{\mathbf{Y}} \in \mathbb{R}^{M \times K}, \quad (1)$$

где $\mathbf{y}_m = \mathbf{y}(t_m)$, а $\underline{\mathbf{X}}_m \in \mathbb{R}^{n_1 \times \dots \times n_D}$ - тензор (многоиндексная матрица).

Требуется восстановить значения \mathbf{y}_m по заданному тензору $\underline{\mathbf{X}}_m$, $m = 1, \dots, M$.

В данной работе предлагается применять композицию моделей h и g , такую что:

$$\varphi = h \circ g : \mathbb{R}^{n_1 \times \dots \times n_D} \rightarrow \mathbb{R}^K, \quad (2)$$

где функции h и g удовлетворяют процессу:

$$\underline{\mathbf{X}}_m \xrightarrow{h(\cdot, \theta)} \underline{\mathbf{Q}}_m \xrightarrow{g(\cdot, \eta)} \mathbf{Y}_m$$

Определение 3 Модель снижения размерности $h : \mathbb{R}^{n_1 \times \dots \times n_D} \rightarrow \mathbb{R}^{\ell_1 \times \dots \times \ell_D}$, где $\ell_i < n_i$ является преобразованием, которое переводит из пространства большей размерности $\mathbb{R}^{n_1 \times \dots \times n_D}$ в пространство меньшей размерности $\mathbb{R}^{\ell_1 \times \dots \times \ell_D}$, то есть $\underline{\mathbf{Q}}_m = h(\underline{\mathbf{X}}_m, \theta)$, где θ – параметры преобразования.

Определение 4 Модель согласования $g : \mathbb{R}^{\ell_1 \times \dots \times \ell_D} \rightarrow \mathbb{R}^K$ – преобразование, переводящая из пространства меньшей размерности $\mathbb{R}^{\ell_1 \times \dots \times \ell_D}$ в пространство целевой переменной \mathbb{R}^K , то есть $\mathbf{Y}_m = g(\underline{\mathbf{Q}}_m, \eta)$, где η – параметры преобразования.

Модель $\varphi^*(\cdot, \theta, \eta)$ является оптимальной, если она минимизирует некоторый функционал ошибки \mathcal{L} :

$$\varphi^* = \arg \min_{\{\theta, \eta\}} \mathcal{L}(\varphi(\underline{\mathbf{X}}, \theta, \eta), \mathbf{Y}) \quad (3)$$

Возможные модели снижения размерности будут рассмотрены в пункте 4. В качестве модели согласования могут быть использованы мультимодальная, тензорная и нелинейная тензорная регрессии.

2.2 Мультимодальная регрессия

Тензор признаков $\underline{\mathbf{X}}$ имеет $D+1$ размерность. Чтобы восстановить целевые временные ряды, можно матрицизовать тензор признаков $\underline{\mathbf{X}}$ по первой размерности:

$$\underline{\mathbf{X}}_{(1)} = \left[\text{vec}(\underline{\mathbf{X}}_1)^\top, \dots, \text{vec}(\underline{\mathbf{X}}_M)^\top \right]^\top \in \mathbb{R}^{M \times (n_1 \dots n_D)} \quad (4)$$

таким образом задача сводится к задаче мультимодальной регрессии, где $\mathbf{X} \in \mathbb{R}^{M \times (n_1 \dots n_D)}$ – исходная матрица, полученная матрицизацией исходного тензора, $\mathbf{Y} \in \mathbb{R}^{M \times K}$ – целевая матрица. Требуется построить прогностическую модель $\mathbf{f} : \mathbb{R}^{M \times (n_1 \dots n_D)} \rightarrow \mathbb{R}^{M \times K}$. Данная модель должна быть оптимальной \mathbf{f}^* , то есть минимизирующей некоторую функционал ошибки \mathcal{L} :

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \mathcal{L}(\mathbf{f}, \mathbf{X}, \mathbf{Y}) \quad (5)$$

Будем рассматривать только класс параметрических моделей $\mathbf{f}(\mathbf{X}, \Theta)$, где Θ – матрица параметров, для сужения пространства поиска моделей. Тогда задача (5) сводится к задаче поиска оптимальных параметров:

$$\Theta^* = \arg \min_{\Theta} \mathcal{L}(\Theta, \mathbf{X}, \mathbf{Y}) \quad (6)$$

В данной работе рассматриваемые пространства исходной и целевой переменных имеют избыточную размерность. В таком случае решение задачи (6) неустойчиво. Например, при восстановлении линейной регрессии:

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \varepsilon, \quad \mathbf{\Theta} \in \mathbb{R}^{(n_1 \dots n_D) \times K} \quad (7)$$

Оптимальные параметры $\mathbf{\Theta}^*$ определяются минимизацией функционала ошибки:

$$\mathcal{L}(\mathbf{\Theta}, \mathbf{X}, \mathbf{Y}) = \|\mathbf{Y} - \mathbf{X}\mathbf{\Theta}\|_2^2 \rightarrow \min_{\mathbf{\Theta}} \quad (8)$$

Решением задачи оптимизации (8) является:

$$\mathbf{\Theta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad (9)$$

Линейная зависимость между признаками в матрице \mathbf{X} приводит к неустойчивости решения задачи (8), так как матрица $\mathbf{X}^\top \mathbf{X}$ близка к сингулярной. Для решения данной проблемы используются методы снижения размерности пространства, описываемые в пункте 4. Они позволяют избежать сильной линейной зависимости между признаками.

2.3 Тензорная регрессия

Матрицизация тензора $\underline{\mathbf{X}}$ приводит к чрезмерно большому количеству столбцов. Так же теряется информация для структурированных данных при матрицизации. И так как тензоры это многоиндексные матрицы, то тензорная регрессия может быть определена подобным образом, как (7):

$$\mathbf{y}_m = \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle + \varepsilon \quad (10)$$

где $\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle$ - тензорная свертка по первым D размерностям D -мерного исходного тензора $\underline{\mathbf{X}}_m \in \mathbb{R}^{n_1 \times \dots \times n_D}$ и $(D+1)$ -мерного тензора параметров модели $\underline{\mathbf{W}} \in \mathbb{R}^{n_1 \times \dots \times n_D \times K}$, $\varepsilon \in \mathbb{R}^K$ - ошибка, $\mathbf{y}_m \in \mathbb{R}^K$ - целевой вектор.

k -ый элемент вектора, получившегося после тензорной свертки $\underline{\mathbf{X}}_m$ и $\underline{\mathbf{W}}$ считается следующим образом:

$$\langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle_k = \sum_{i_1=1}^{n_1} \dots \sum_{i_D=1}^{n_D} x_{i_1, \dots, i_D} w_{i_1, \dots, i_D, k} \quad (11)$$

Оптимальный тензор параметров $\underline{\mathbf{W}}^*$ находится путем минимизации квадратичной функции ошибки:

$$\underline{\mathbf{W}}^* = \arg \min_{\underline{\mathbf{W}}} \sum_{m=1}^M \|\mathbf{y}_m - \langle \underline{\mathbf{X}}_m | \underline{\mathbf{W}} \rangle\|_2^2 \quad (12)$$

На практике тензор параметров $\underline{\mathbf{W}}$ представляют в разложении Такера:

$$\underline{\mathbf{W}} \approx \underline{\mathbf{G}}_{\times_1} \mathbf{U}^{(1)} \dots \times_D \mathbf{U}^{(D)} \times_{D+1} \mathbf{U}^{(D+1)} \quad (13)$$

где $\underline{\mathbf{G}}$ - центральный тензор меньшего размера, $\mathbf{U}^{(i)}$ - унитарные матрицы.

2.4 Снижение размерности

Для снижения размерности пространства часто применяется метод главных компонент PCA и его тензорные модификации. Основным недостатком которого является то, что он не учитывает взаимосвязь между объектами \mathbf{X} и целевой переменной \mathbf{Y} .

Этого недостатка нет у метода PLS (partial least squares), так как он находит в скрытом пространстве матрицы \mathbf{T}, \mathbf{U} , которые лучше всего описывают исходные матрицы \mathbf{X}, \mathbf{Y} . Основная идея метода PLS заключается в том, что он максимизирует ковариацию между столбцами матриц \mathbf{T} и \mathbf{U} . Тем самым, учитывается взаимосвязь между \mathbf{X} и \mathbf{Y} . Более того, для работы с тензорными данными были разработаны тензорные модификации PLS, например HOPLS. Основным ограничением методов PLS и HOPLS является их линейность.

Еще одним из популярных методов снижения размерности пространства являются автоэнкодеры. Неглубокие автоэнкодеры по существу эквивалентны методу главных компонент (PCA). Но если автоэнкодер имеет нелинейный энкодер и декодер, то он становится эквивалентным нелинейному PCA [26]. Важным преимуществом автоэнкодеров так же является то, что они работают с тензорами. Но многие автоэнкодеры обучаются, минимизируя функционал потерь, никак не учитывающим \mathbf{Y} , тем самым не учитывается взаимосвязь между \mathbf{X} и \mathbf{Y} .

В данной работе предлагаются автоэнкодеры, учитывающие взаимосвязь между признаковым описанием объектов и целевой переменной, а так же корреляцию между столбцами нового признаково описания в скрытом пространстве.

В последующих главах будут подробно рассмотрены PLS, HOPLS и предлагаемые автоэнкодеры.

3 Метод частичных наименьших квадратов

Пусть задана выборка (\mathbf{X}, \mathbf{Y}) , где $\mathbf{X} \in \mathbb{R}^{m \times n}$ — матрицы объектов, $\mathbf{Y} \in \mathbb{R}^{m \times r}$ — матрица ответов. Метод частных наименьших квадратов (partial least squares, PLS) проецирует матрицу объектов \mathbf{X} и матрицу ответов \mathbf{Y} в скрытое пространство \mathbb{R}^ℓ меньшей размерности ($\ell < r < n$). Метод PLS находит в скрытом пространстве матрицы $\mathbf{T}, \mathbf{U} \in \mathbb{R}^{m \times \ell}$, которые лучше всего описывают исходные матрицы \mathbf{X} и \mathbf{Y} . При этом PLS максимизирует ковариацию между столбцами матриц \mathbf{T} и \mathbf{U} соответственно.

Исходная матрица \mathbf{X} и целевая матрица \mathbf{Y} проецируются на скрытое пространство следующим образом:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^\top + \mathbf{F} = \sum_{k=1}^{\ell} \mathbf{t}_k \cdot \mathbf{p}_k^\top + \mathbf{F}, \mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^\top + \mathbf{E} = \sum_{k=1}^{\ell} \mathbf{u}_k \cdot \mathbf{q}_k^\top + \mathbf{E}.$$

Здесь \mathbf{T} и \mathbf{U} — образы исходных матриц в скрытом пространстве, причем столбцы матрицы \mathbf{T} ортогональны; \mathbf{P} и \mathbf{Q} — матрицы перехода; \mathbf{F} и \mathbf{E} — матрицы остатков. Метод PLS восстанавливает линейную зависимость между столбцами матриц \mathbf{T} и \mathbf{U} :

$$\mathbf{U} \approx \mathbf{T}\mathbf{B}, \quad \mathbf{B} = \text{diag}(\beta_k), \quad \beta_k = \mathbf{u}_k^\top \mathbf{t}_k / \|\mathbf{t}_k\|_2^2,$$

где $\{\mathbf{t}_k\}_{k=1}^\ell, \{\mathbf{u}_k\}_{k=1}^\ell$ — столбцы матриц \mathbf{T} и \mathbf{U} соответственно. *PLS* решает следующую оптимизационную задачу:

$$\text{cov}(\mathbf{X}\mathbf{p}, \mathbf{Y}\mathbf{q})^2 = \frac{\mathbf{p}^\top \mathbf{X}^\top \mathbf{Y} \mathbf{q}}{\|\mathbf{p}\|_2 \|\mathbf{q}\|_2} \rightarrow \max_{\mathbf{p}, \mathbf{q}}, \text{ s.t. } \mathbf{p}^\top \mathbf{p} = 1, \quad \mathbf{q}^\top \mathbf{q} = 1.$$

Линейное преобразование между объектами в исходном и скрытом пространствах имеет вид

$$\mathbf{T} = \mathbf{X}\mathbf{W}^*,$$

где $\mathbf{W}^* = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1}$.

Матрица параметров модели находится следующим образом:

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{W}^*\mathbf{Q}^\top + \mathbf{E} = \mathbf{X}\mathbf{\Theta} + \mathbf{E}.$$

Таким образом, параметры модели равны

$$\mathbf{\Theta} = \mathbf{W}(\mathbf{P}^\top \mathbf{W})^{-1} \mathbf{B}\mathbf{Q}^\top.$$

3.1 Тензорный метод частичных наименьших квадратов

4 Автоэнкодер

Для задачи снижения размерности часто применяются автоэнкодеры. Так, например, для задачи снижения размерности при работе с данными электрокортикограмм был применен автоэнкодер *ReducedNet* [27]. Предсказание координат движения рук были оказались более точным после снижения размерности с помощью *ReducedNet*, чем после таких методов, как *PLS* или *KernelPCA*. Поэтому было решено применить эту модель в данной работе, но с небольшими модификациями.

Предложенная в [27] модель состояла из двух блоков: модуля понижения размерности, состоящего из энкодера h и декодера ψ ; и модуля декодирования ω . Модуль понижения размерности состоял из сверточных блоков (*conv1D*) и *LSTM* блоков. Каждый сверточный блок состоял из одномерного сверточного слоя, одного слоя *Batch Normalization*, функции активации *GELU* (*gaussian error linear unit*) и *Dropout*. *LSTM* блок — это один *LSTM* слой. Модуль декодирования представлял собой один линейный слой. Единственным настраиваемым параметром нейронной сети была размерность данных на выходе из *LSTM* слоя.

В данной работе предлагается несколько модификаций данной нейронной сети: *ModifiedReducedNet*, *TensorReducedNet* и *TensorDimReducedNet*.

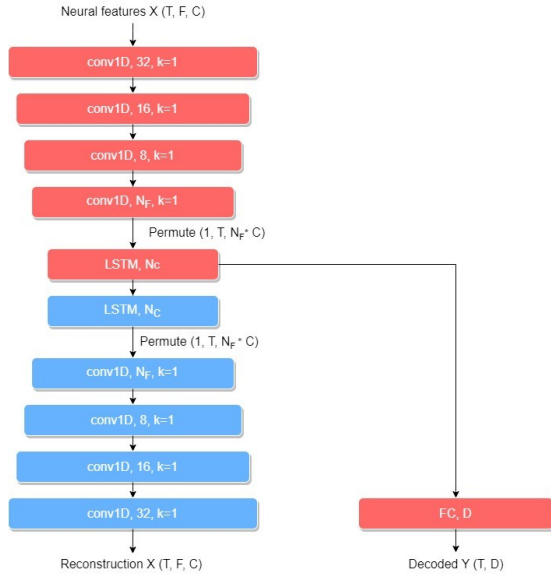


Рис. 1 ModifiedReducedNet.

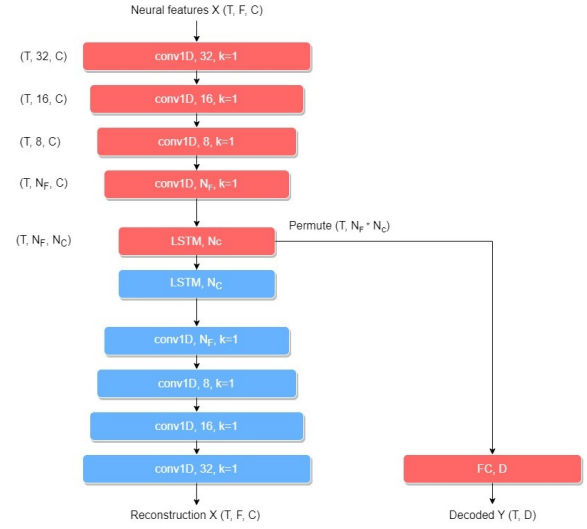


Рис. 2 TensorReducedNet.

4.1 Автоэнкодер ModifiedReducedNet

Главным отличием *ModifiedReducedNet* (Рис. 1) от *ReducedNet* является наличие второго параметра сети, такого как количество каналов, которое получается на выходе из сверточных блоков (N_F). То есть на вход подается тензор $X \in \mathbb{R}^{T \times F \times C}$, на выходе из четвертого сверточного слоя получается тензор размерности $\mathbb{R}^{T \times N_F \times C}$. Более того было увеличено количество каналов, которое получается на выходе из первого сверточного блока, чтобы усложнить немного модель для наших данных.

Однако выход после автоэнкодера получается матричным и теряет информацию, которую можно было бы сохранить не применяя матрицизацию перед LSTM блоком.

4.2 Автоэнкодер TensorReducedNet

Тензорной модификацией *ReducedNet* является *TensorReducedNet* (Рис. 2). Данная модификация позволяет не терять исходную структуру данных, создавая на выходе энкодера тензор. Это позволяет применять как тензорные, так и матричные регрессионные модели.

4.3 Автоэнкодер TensorDimReducedNet

С помощью *TensorDimReducedNet* (Рис. 3) можно создавать на выходе из энкодера тензора с дополнительными размерностями, которые можно трактовать, как дополнительные нелинейные признаки. Это позволяет учитывать нелинейные взаимосвязи между \mathbf{X} и \mathbf{Y} , используя линейные модели.

4.4 Функции потерь

Для обучения автоэнкодеров, представленных выше, использовалось две разные функции потерь. Первая, как в статье [27]:

$$\mathcal{L}_1 = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{dec}, \quad (14)$$

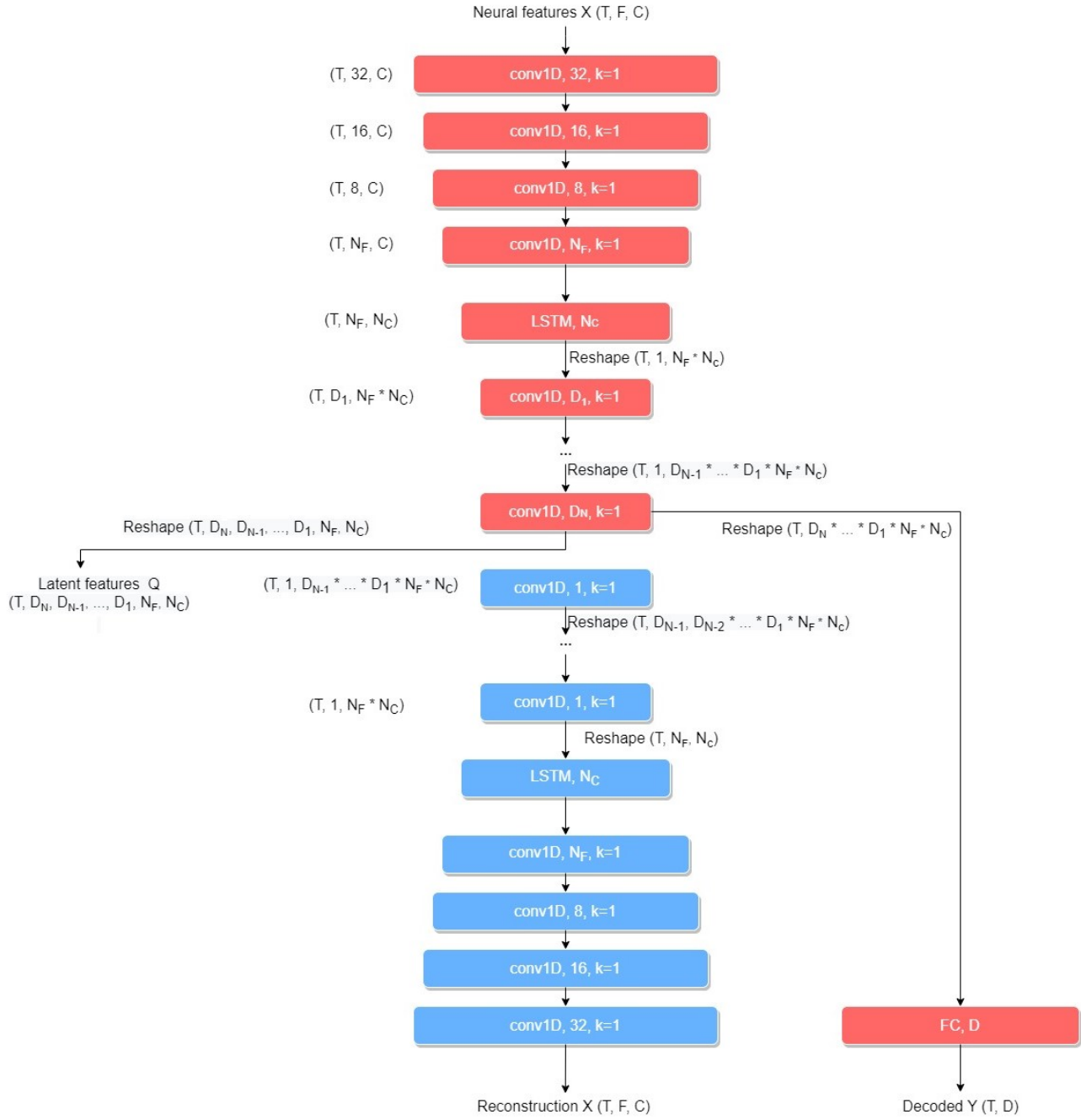


Рис. 3 TensorDimReducedNet.

где:

$$\mathcal{L}_{rec} = \frac{1}{M} \sum_{m=1}^M \|\underline{\mathbf{X}}_m - h \circ \psi(\underline{\mathbf{X}}_m)\|^2$$

$$\mathcal{L}_{dec} = \frac{1}{M} \sum_{m=1}^M \|\mathbf{Y}_m - \omega \circ h(\underline{\mathbf{X}}_m)\|^2$$

Вторая, предлагается впервые в данной работе:

$$\mathcal{L}_2 = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{dec} + \beta \cdot \mathcal{L}_{cor}, \quad (15)$$

где:

$$\mathcal{L}_{cor} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{corr}(h(\underline{\mathbf{X}}_m)_i, h(\underline{\mathbf{X}}_m)_j)$$

где n - количество столцов в тензоре $h(\underline{\mathbf{X}}_m)$.

Она позволяет получать новые признаковые описания объектов в скрытом пространстве слабо коррелирующие между собой.

5 Эксперимент

5.1 Данные

Таблица 1 Данные

Название	Источники	Предсказываемый сигнал
Sensors	$\underline{\mathbf{X}} \in \mathbb{R}^{t \times 6 \times 15}$	$\mathbf{Y} \in \mathbb{R}^{t \times 3}$
ECoG	$\underline{\mathbf{X}} \in \mathbb{R}^{t \times 32 \times 27}$	$\mathbf{Y} \in \mathbb{R}^{t \times 3}$
BerkeleyMHAD-video	$\underline{\mathbf{X}} \in \mathbb{R}^{t \times 15 \times 35 \times 4}$	$\underline{\mathbf{Y}} \in \mathbb{R}^{t \times 3 \times 6}$

Данные Sensors

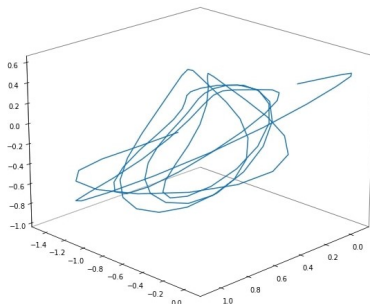


Рис. 4 Данные с гироскопа

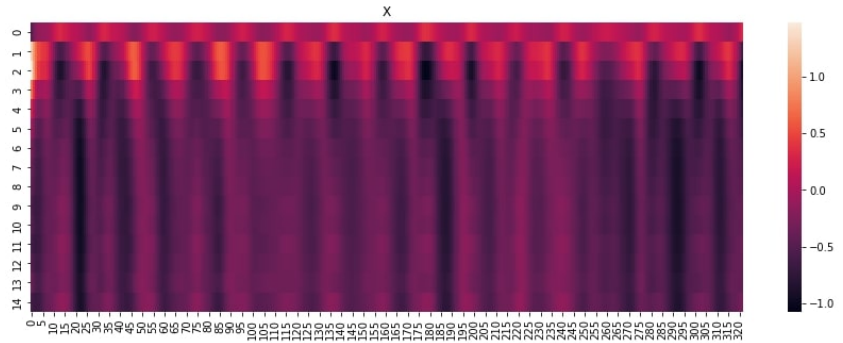


Рис. 5 Признаки в частотно-временном пространстве

Одним из наборов данных для вычислительного эксперимента являются данные собранные с помощью сенсоров на телефонах (один гироскоп и два акселерометра). Это простые для предсказания данные, так как процесс записи таких данных был простым. Один из телефонов был размещен в руке, второй в кармане. Человек с телефонами сделал около 20 шагов по кругу. Данные с двух акселерометров были переведены в частотно-временное представлении и объединены в X . Данные с гироскопа были взяты за целевую

переменную Y (Рис. 4). Таким образом, $X \in \mathbb{R}^{T \times 6 \times 15}$, где 15 – количество частот (Рис. 5), и $Y \in \mathbb{R}^{T \times 3}$.

Данные ECoG

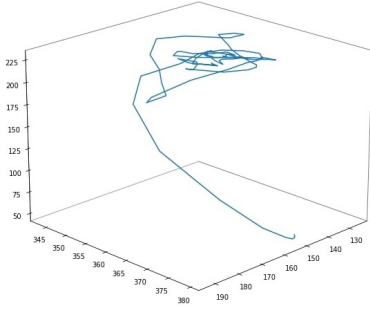


Рис. 6 Траектория движения руки

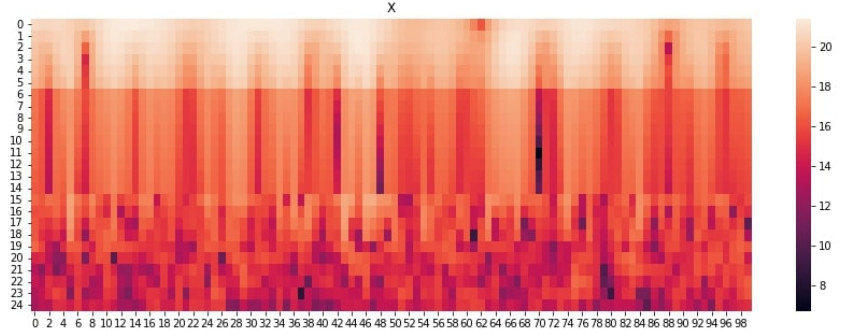


Рис. 7 Данные с одного канала в частотно-временном пространстве

Другим набором данных для вычислительного эксперимента являются данные электрокортикограмм (ECoG) из датасета Neurotycho [28]. Данные состоят из 32-канальных сигналов напряжения, снятых с головного мозга. Данные ECoG являются многомерными и измерения коррелируют как во временной, так и в пространственной областях. Целевой переменной являются координаты положения руки в пространстве (Рис. 6). Исходные сигналы напряжения преобразуются в частотно-временное представление с помощью вейвлет-преобразования с материнским вейвлетом Морле, так как такой тип преобразования часто используется в задачах с ECoG данными [29], [30]. Описание исходного сигнала в каждый момент времени имеет размерность 32 (каналы) \times 27 (частоты) = 864. Каждый сигнал представляет собой локальный отрезок времени длительностью $\Delta t = 1s$. Временной шаг между сигналами $\delta t = 0.05s$. Данные имеют размеры $\mathbf{X} \in \mathbb{R}^{18900 \times 32 \times 27}$ (Рис. 7) и $\mathbf{Y} \in \mathbb{R}^{18900 \times 3}$. Данные разбиты на тренировочную и тестовую выборки в соотношении 0.7.

Данные BerkeleyMHAD-video

5.2 Параметры сети

Обучение всех моделей происходило с оптимизатором Adam, с learning rate 0.01 при работе с первой функцией потерь (14) и с learning rate 0.005 при работе с второй функцией потерь (15). Learning rate умножался каждые 50 эпох на 0.75

Подбор параметров модели, таких как $N_F, N_{CH}, D_1, \dots, D_N$, для моделей перечисленных в пункте 4 проводилось путем подсчета качества предсказания линейной регрессии по данным, полученным с помощью энкодера, на кросс-валидации с тремя фолдами. Качество предсказания оценивалось по $nRMSE$ (normalized root mean squared error) метрике. Обучение автоэнкодера для подбора параметров проходило 200 эпох для Sensor данных и 100 эпох для ECoG данных.

Параметр α в функции потерь \mathcal{L}_1 (14). был взят, как в статье [27], то есть $\alpha = 3$. Параметры в функции потерь \mathcal{L}_2 (15) подбирались способом, аналогичным подбору параметров сети, только параметры сети были зафиксированы. Для *Sensors* данных оптимальными α , β оказались 10, 3, соответственно. Для *ECoG* данных оптимальными α , β оказались 5, 1, соответственно.

После подбора всех параметров модели обучались 600 эпох для *Sensor* данных и 200 эпох для *ECoG* данных. С помощью энкодера из обученной модели были получены данные в пространстве меньшей размерности. После этого целевая переменная была предсказана с помощью линейной регрессии (LR) по матрицизованным данным, либо тензорной регрессией (TR).

5.3 Результаты

Таблица 2 Результаты экспериментов с автоэнкодером с функцией потерь \mathcal{L}_1 (14)

Данные	Модель	Количество компонент	RMSE	Количество параметров
Sensors	PLS	5	0.757	-
	HOPLS	3×1	0.811	-
	MatrixReducedNet	6	0.720	24163
	TensorReducedNet+LR	5×2	0.682	3069
	TensorReducedNet+TR	5×2	0.675	3069
	TensorDimReducedNet+LR	$2 \times 6 \times 2$	0.710	3163
	TensorDimReducedNet+TR	$2 \times 6 \times 2$	0.691	3163
ECoG	PLS	4	0.957	-
	HOPLS	1×1	1.138	-
	MatrixReducedNet	19	0.925	57106
	TensorReducedNet+LR	6×14	0.926	12418
	TensorReducedNet+TR	6×14	0.922	12418
	TensorDimReducedNet+LR	$2 \times 5 \times 14$	0.937	12808
	TensorDimReducedNet+TR	$2 \times 5 \times 14$	0.931	12808

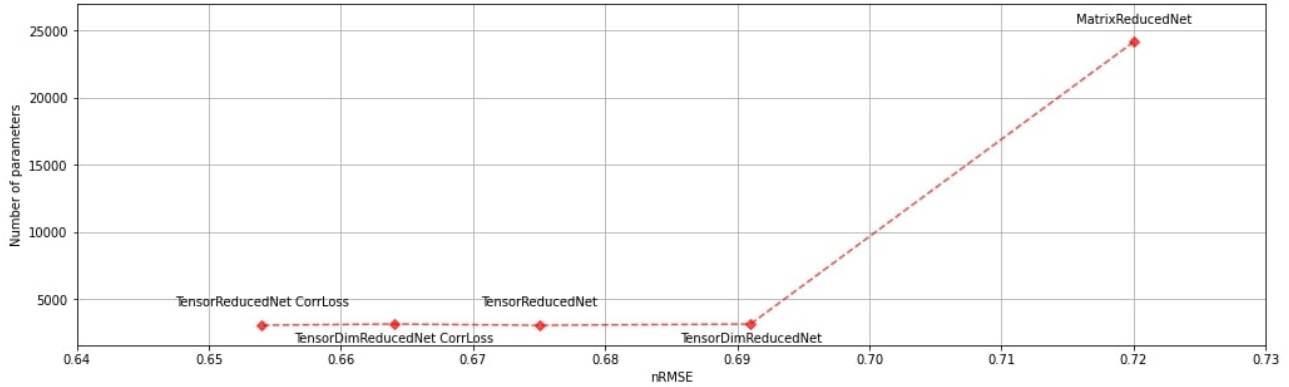
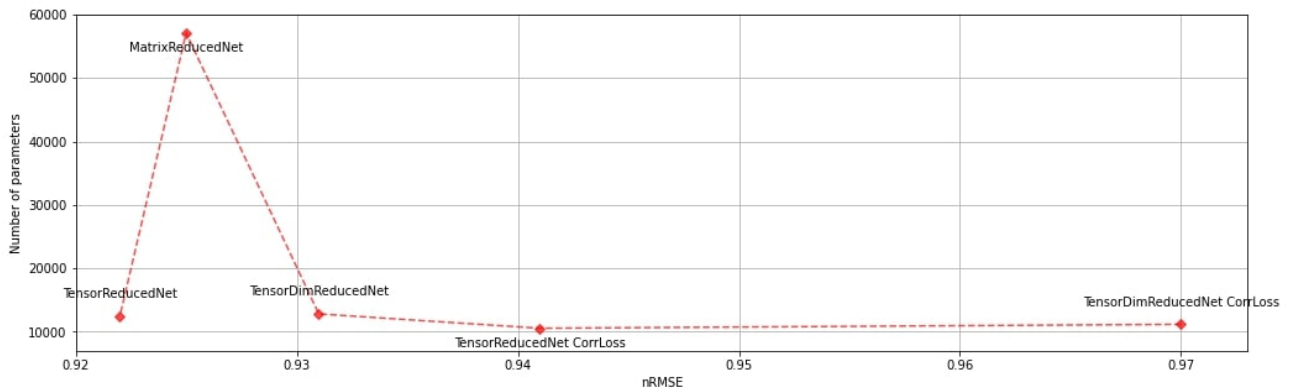
6 Заключение

Литература

- [1] Jean Faber Alexander Wyss Napoleon Torres Corinne Mestais Alim Louis Benabid Andrey Eliseyev, Cecile Moro and Tetiana Aksenova. *L1-penalized n-way pls for subset of electrodes selection in bci experiments*. Journal of neural engineering, 9(4), 2012.
- [2] Andrey Eliseyev and Tatiana Aksenova. *Stable and artifact-resistant decoding of 3d hand trajectories from ecog signals using the generalized additive model*. Journal of neural engineering, 11(6), 2014.

Таблица 3 Результаты экспериментов с автоэнкодером с функцией потерь \mathcal{L}_2 (15)

Данные	Модель	Количество компонент	RMSE	Количество параметров
Sensors	TensorReducedNet+LR	2×3	0.705	3076
	TensorReducedNet+TR	2×3	0.654	3076
	TensorDimReducedNet+LR	$6 \times 4 \times 1$	0.704	3165
	TensorDimReducedNet+TR	$6 \times 4 \times 1$	0.664	3165
ECoG	TensorReducedNet+LR	6×9	0.939	10548
	TensorReducedNet+TR	6×9	0.941	10548
	TensorDimReducedNet+LR	$2 \times 4 \times 10$	0.970	11169
	TensorDimReducedNet+TR	$2 \times 4 \times 10$	0.973	11169

**Рис. 8** Сравнение моделей на Sensors данных**Рис. 9** Сравнение моделей на ECoG данных

- [3] Anastasia Motrenko and Vadim Strijov. Multi-way feature selection for ecog-based brain-computer interface. Expert Systems with Applications, 114:402–413, 2018.
- [4] Lars Snipen Tahir Mehmood, Kristian Hovde Liland and Solve Saebo. A review of variable selection methods in partial least squares regression. Chemometrics and Intelligent Laboratory Systems, 118(12):62–69, 2012.

-
- [5] *Alexandr Katrutsa and Vadim Strijov. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. Expert Systems with Applications, 76:1–11, 2017.*
 - [6] *Alexandr Katrutsa and Vadim Strijov. Stress test procedure for feature selection algorithms. Chemometrics and Intelligent Laboratory Systems, 142:172–183, 2015.*
 - [7] *Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. IEEE transactions on knowledge and data engineering, 31(10):1863–1883, 2018.*
 - [8] *Adrian Benton, Huda Khayrallah, Biman Gujral, Drew Reisinger, Sheng Zhang, and Raman Arora. Deep generalized canonical correlation analysis. CoRR, abs/1702.02519, 2017.*
 - [9] *Hok Shing Wong, Li Wang, Raymond H. Chan, and Tieyong Zeng. Deep tensor cca for multi-view learning. CoRR, abs/2005.11914, 2020.*
 - [10] *Heather D. Couture, Roland Kwitt, J. S. Marron, Melissa A. Troester, Charles M. Perou, and Marc Niethammer. Deep multi-view learning via task-optimal cca. CoRR, abs/1907.07739, 2019.*
 - [11] *Andrey Eliseyev and Tetiana Aksenova. Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. PloS one, 11(5), 2016.*
 - [12] *Roman Rosipal. Nonlinear partial least squares an overview. Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques, pages 169–189, 2011.*
 - [13] *Roman Rosipal and Nicole Kramer. Overview and recent advances in partial least squares. C. Saunders et al. (Eds.) SLSFS 2005, LNCS 3940, page 34–51, 2006.*
 - [14] *Alexander E. Stott, Bruno Scalzo Dees, Ilia Kisil, and Danilo P. Mandic. A class of multidimensional nipals algorithms for quaternion and tensor partial least squares regression. Signal Process., 160:316–327, 2019.*
 - [15] *Solve Sæbø Tahir Mehmood and Kristian Hovde Liland. Comparison of variable selection methods in partial least squares regression. Journal of Chemometrics, 2020.*
 - [16] *Petre Manolescu Julien Lauzon-Gauthier and Carl Duchesne. The sequential multi-block pls algorithm (smb-pls): Comparison of performance and interpretability. Chemometrics and Intelligent Laboratory Systems, (180):72–83, 2018.*
 - [17] *Eric F. Lock. Tensor-on-tensor regression. Journal of Computational and Graphical Statistics, 27(3):638–647, 2018.*
 - [18] *Kamran Paynabar Mostafa Reisi Gahrooei, Hao Yan and Jianjun Shi. Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. Technometrics, 2020.*
 - [19] *Johan A. K. Suykens Lynn Houthuys. Tensor learning in multi-view kernel pca. Artificial Neural Networks and Machine Learning - ICANN 2018, 11140, 2018.*
 - [20] *Haiping Lu, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos. Multilinear principal component analysis of tensor objects for recognition. pages 776–779, 2006.*

- [21] Yong Luo, Dacheng Tao, Yonggang Wen, Kotagiri Ramamohanarao, and Chao Xu. *Tensor canonical correlation analysis for multi-view dimension reduction*. CoRR, abs/1502.02330, 2015.
- [22] Qibin Zhao, Cesar F Caiafa, Danilo P Mandic, Zenas C Chao, Yasuo Nagasaka, Naotaka Fujii, Liqing Zhang, and Andrzej Cichocki. *Higher order partial least squares (hopls): a generalized multilinear regression method*. IEEE transactions on pattern analysis and machine intelligence, 35(7):1660–1673, 2012.
- [23] Y. Xu X. Feng and Y. Meng. *Short-term load forecasting with tensor partial least squares-neural network*. Energies, 990(12), 2019.
- [24] Qibin Zhao, Liqing Zhang, and Andrzej Cichocki. *Multilinear and nonlinear generalizations of partial least squares: an overview of recent advances*. Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 4(2):104–115, 2014.
- [25] Jun Huang, Kehua Su, Jamal El-Den, Tao Hu, and Junlong Li. *An mpca/lda based dimensionality reduction algorithm for face recognition*. Mathematical Problems in Engineering, 2014:1–12, 08 2014.
- [26] Yoshua Bengio Ian Goodfellow. *Deep learning*. 2006.
- [27] Chen W. Yvert B. Zhang S. Ran, X. *A hybrid autoencoder framework of dimensionality reduction for brain-computer interface decoding*. Computers in Biology and Medicine, page 148, 2022.
- [28] Nagasaka Y. Chao Z. C. Fujii N. Shimoda, K. *Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in japanese macaques*. Journal of neural engineering, page 9(3), 2012.
- [29] Aksenova T. Eliseyev, A. *Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording*. PloS one, page 11(5), 2016.
- [30] Nagasaka Y. Fujii N. Chao, Z. C. *Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey*. Frontiers in neuroengineering, page 3, 2010.