Задача обучения с экспертом для построения интерпретируемых моделей машинного обучения

Грабовой Андрей Валериевич

Кафедра интеллектуального анализа данных
Научный руководитель: д.ф.-м.н. В. В. Стрижов

Московский физико-технический институт
8 декабря 2020 г.

Обучение с экспертной информацией о данных

Цель
Предложить метод обучения выбора моделей машинного обучения на основе *экспертной информацией* об исследуемых объектах.

Исследуемая проблема
Снижение размерности пространства параметров моделей глубокого обучения при помощи интерпретируемых моделей машинного обучения на основе экспертной информации.

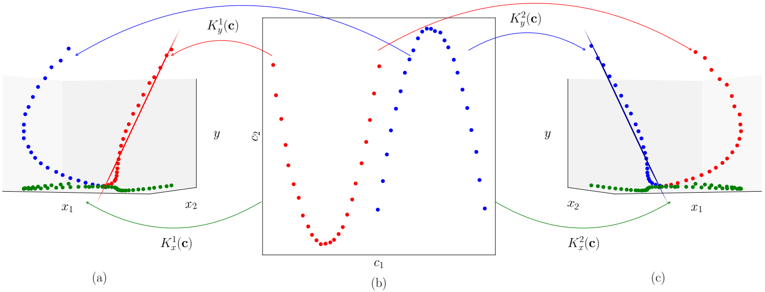
Требуется

1. Формально поставить задача обучения на основе экспертного описания данных.
2. Предложить метод решения на основе построения экспертного признакового описания объектов.
3. Использовать этот метод для решения задачи аппроксимации кривых второго порядка на бинарном изображении.

Список литературы

1. *Грабовой А.В., Стрижов В.В.* Анализ свойств вероятностных моделей обучения с экспертом // в процессе подачи.
2. *Грабовой А.В., Стрижов В.В.* Анализ выбора априорного распределения для смеси экспертов // Журнал Вычислительной математики и математической физики, 2021. Т. 61. № 5.
3. *Yuksel Seniha Esen, Wilson Joseph N., Gader Paul D* Twenty Years of Mixture of Experts // IEEE Transactions on Neural Networks and Learning Systems, 2012. Vol. 23. No 8. Pp. 1177–1193.

Аппроксимации кривых второго порядка



(b) Исходный набор точек на изображении.

(a) Представление первого эксперта: K_x^1, K_y^1 — отображение в данное представление.

(c) Представление второго эксперта: K_x^2, K_y^2 — отображение в данное представление.

Постановка задачи: кривые второго порядка

Изображение

$$\mathbf{M} \in \{0, 1\}^{m_1 \times m_2},$$

где 1 — точка изображение, а 0 — точка фона.
Точки изображения — кривая второго порядка Ω. Координаты точек изображения $\mathbf{C} \in \mathbb{R}^{N \times 2}$. Задана экспертная информация $E\bigl(\Omega\bigr)$ о фигуре Ω.
Признаковое описание экспертов:

$$K_x\bigl(E\bigl(\Omega\bigr)\bigr) : \mathbb{R}^2 \to \mathbb{R}^n, \quad K_y\bigl(E\bigl(\Omega\bigr)\bigr) : \mathbb{R}^2 \to \mathbb{R},$$

где K_x, K_y отображения в признаковое описание и пространство ответов.
Набор данных для аппроксимации кривых:

$$\mathfrak{D} = \{(\mathbf{x}, y) \mid \forall \mathbf{c} \in \mathbf{C} \quad \mathbf{x} = K_x(\mathbf{c}),\ y = K_y(\mathbf{c})\}.$$

В данной работе предполагается, что выборка $\mathfrak{D}$ аппроксимируется линейной моделью $g(\mathbf{x}, \mathbf{w}) = \mathbf{x}^{\mathsf{T}}\mathbf{w}$, где $\mathbf{w}$ вектор, параметр, который требуется найти.
Требуется решить оптимизационную задачу:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w} \in \mathbb{R}^n} \sum_{(\mathbf{x}, y) \in \mathfrak{D}} \|g(\mathbf{x}, \mathbf{w}) - y\|_2^2.$$

Постановка задачи: признаковое описание кривых

Кривая второго порядка: главная ось которой не параллельна оси ординат:

$$x^2 = B'xy + C'y^2 + D'x + E'y + F',$$

также на коэффициенты B', C' могут быть наложены ограничения.

Отображение в экспертное описание:

$$K_x(\mathbf{c}_i) = \left[x_i y_i, y_i^2, x_i, y_i, 1\right], \quad K_y(\mathbf{c}_i) = x_i^2.$$

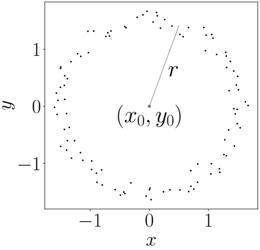
Частный случай: аппроксимации окружности

Пусть x_0, y_0 — центр окружности, которую требуется найти, а r ее радиус.

Точки $(x_i, y_i) \in \mathbf{C}$ удовлетворяют уравнению окружности:

$$(x_i - x_0)^2 + (y_i - y_0)^2 = r^2 \Rightarrow$$

$$\Rightarrow (2x_0) \cdot x_i + (2y_0) \cdot y_i + (r^2 - x_0^2 - y_0^2) \cdot 1 = x_i^2 + y_i^2.$$



Линейная регрессия для аппроксимации окружности:

$$\mathbf{Xw} \approx \mathbf{y}, \quad \mathbf{X} = [\mathbf{C}, \mathbf{1}], \quad \mathbf{y} = [x_1^2 + y_1^2, x_2^2 + y_2^2, \cdots, x_N^2 + y_N^2]^\mathsf{T},$$

где оптимальные параметры $\mathbf{w} = \left[w_1, w_2, w_3\right]^\mathsf{T}$ восстанавливают окружность:

$$x_0 = \frac{w_1}{2}, \quad y_0 = \frac{w_2}{2}, \quad r = \sqrt{w_3 + x_0^2 + y_0^2}.$$

Постановка задачи: мультимодель

Заданы K кривых второго порядка $\Omega_1, \cdots, \Omega_K$ с $E_k = E\bigl(\Omega_k\bigr)$.

Определение

Функция f называется смесью K экспертов, если:

$$f = \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) g_k(\mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n \times |\mathbf{V}|} \to [0,\, 1], \quad \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) = 1,$$

где g_k — локальная модель, $\mathbf{x}$ — признаки, π_k — шлюзовая функция, $\mathbf{w}_k$ — параметры, $\mathbf{V}$ — параметры шлюзовой функции.

Мультимодель, описывающую кривые $\Omega_1, \ldots, \Omega_K$ на изображении $\mathbf{M}$:

$$f = \sum_{\mathbf{c} \in \mathbf{C}} \sum_{k=1}^{K} \pi_k(\mathbf{c}, \mathbf{V}) g_k\bigl(K_x^k(\mathbf{c}), \mathbf{w}_k\bigr), \quad \mathbf{x} = K_x^1(\mathbf{c}) = \cdots = K_x^K(\mathbf{c}).$$

Требуется решить оптимизационную задачу:

$$\mathcal{L} = \sum_{(\mathbf{x}, y) \in \mathfrak{D}} \sum_{k=1}^{K} \pi_k(\mathbf{x}, \mathbf{V}) (y - \mathbf{w}_k^{\mathsf{T}} \mathbf{x})^2 + R\bigl(\mathbf{V}, \mathbf{W}, E(\Omega)\bigr) \to \min_{\mathbf{V}, \mathbf{W}},$$

где $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$ — параметры локальных моделей, $R\bigl(\mathbf{V}, \mathbf{W}, E(\Omega)\bigr)$ — регуляризация параметров, основанная на экспертной информации.

Вероятностная постановка задачи

Заданы:

1) правдоподобие выборки $p_k\left(y_i|\mathbf{w}_k, \mathbf{x}_i\right) = \mathcal{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)$, где β уровень шума,
2) априорное распределение параметров $p^k\left(\mathbf{w}_k\right) = \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right)$, где $\mathbf{w}_k^0$ — вектор размера $n \times 1$, $\mathbf{A}_k$ — ковариационная матрица параметров,
3) регуляризация априорного распределения $p\left(\varepsilon_{k,k'}|\boldsymbol{\alpha}\right) = \mathcal{N}\left(\varepsilon_{k,k'}|\mathbf{0}, \boldsymbol{\Xi}\right)$, где $\boldsymbol{\Xi}$ — ковариационная матрица общего вида, $\varepsilon_{k,k'} = \mathbf{w}_k^0 - \mathbf{w}_{k'}^0$.

Правдоподобие модели включает правдоподобие выборки, априорное распределение параметров, а также их регуляризацию

$$p\big(\mathbf{y}, \mathbf{W}|\mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \Xi, \beta\big) = \prod_{k,k'=1}^{K} \mathcal{N}\left(\varepsilon_{k,k'}|\mathbf{0}, \Xi\right) \cdot \\ \cdot \prod_{k=1}^{K} \mathcal{N}\left(\mathbf{w}_k|\mathbf{w}_k^0, \mathbf{A}_k\right) \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(y_i|\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i, \beta^{-1}\right)\right),$$

где $\mathbf{A} = \big\{\mathbf{A}_1, \cdots, \mathbf{A}_K\big\}$.

Оптимизационная задача

Введем скрытые переменные $\mathbf{Z} = [z_{ik}]$, где $z_{ik} = 1$ тогда и только тогда, когда $k_i = k$:

$$\log p\bigl(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\boldsymbol{\Xi},\beta\bigr) =$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left[\log \pi_k\left(\mathbf{x}_i,\mathbf{V}\right) - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}\log\frac{\beta}{2\pi}\right] +$$

$$+ \sum_{k=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k - \mathbf{w}_k^0\right)^{\mathsf{T}}\mathbf{A}_k^{-1}\left(\mathbf{w}_k - \mathbf{w}_k^0\right) + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\right] +$$

$$+ \sum_{k=1}^{K}\sum_{k'=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k^0 - \mathbf{w}_{k'}^0\right)^{\mathsf{T}}\hat{\boldsymbol{\alpha}}^{-1}\left(\mathbf{w}_k^0 - \mathbf{w}_{k'}^0\right) + \frac{1}{2}\log\det\boldsymbol{\Xi} - \frac{n}{2}\log 2\pi\right].$$

Задача оптимизации параметров локальных моделей и параметров смеси принимает следующий вид:

$$\mathbf{W},\mathbf{Z},\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta = \arg\max_{\mathbf{W},\mathbf{Z},\mathbf{V},\mathbf{W}^0,\mathbf{A},\beta}\log p\bigl(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\boldsymbol{\Xi},\beta\bigr).$$

Для оптимизации используется вариационный ЕМ–алгоритм с предположением $q\left(\mathbf{Z},\mathbf{W}\right) = q\left(\mathbf{Z}\right)q\left(\mathbf{W}\right)$.

EM–алгоритм для решения задачи смеси экспертов

Итерационные формулы EM–алгоритма:

1. E–шаг:

$$p\left(z_{ik}=1\right)=\frac{\exp\left(\log\pi_k\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\right)\right)}{\sum_{k'=1}^{K}\exp\left(\log\pi_{k'}\left(\mathbf{x}_i,\mathbf{V}\right)-\frac{\beta}{2}\left(\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\mathbf{w}_{k'}^{\mathsf{T}}\mathbf{x}_i-\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_{k'}\right)\right)},$$

$$q(\mathbf{w}_k)=\mathcal{N}(\mathbf{w}_k|\mathbf{m}_k,\mathbf{B}_k),$$

$$\mathbf{m}_k=\mathbf{B}_k\left(\mathbf{A}_k^{-1}\mathbf{w}_k^0+\beta\sum_{i=1}^{N}\mathbf{x}_iy_i\mathsf{E}z_{ik}\right),\quad \mathbf{B}_k=\left(\mathbf{A}_k^{-1}+\beta\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}z_{ik}\right)^{-1}.$$

2. M–шаг:

$$\mathbf{A}_k=\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}-\mathbf{w}_k^0\mathsf{E}\mathbf{w}_k^{\mathsf{T}}-\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{0\mathsf{T}}+\mathbf{w}_k^0\mathbf{w}_k^{0\mathsf{T}},$$

$$\frac{1}{\beta}=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\left[y_i^2-2y_i\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k+\mathbf{x}_i^{\mathsf{T}}\mathsf{E}\mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right]\mathsf{E}z_{ik},$$

$$\mathbf{w}_k^0=\left[\mathbf{A}_k^{-1}+(K-1)\,\boldsymbol{\Xi}\right]^{-1}\left(\mathbf{A}_k^{-1}\mathsf{E}\mathbf{w}_k+\boldsymbol{\Xi}\sum_{k'=1,\ k'\neq k}^{K}\mathbf{w}_{k'}^0\right),$$

$$\mathbf{V}=\arg\max_{\mathbf{V}}\mathsf{E}_{q^s}\log p\big(\mathbf{y},\mathbf{Z},\mathbf{W}|\mathbf{X},\mathbf{V},\mathbf{A},\mathbf{W}^0,\boldsymbol{\Xi},\beta\big).$$

Вычислительный эксперимент

Эксперимент с окружностями:

1. Синтетическое изображение трех непересекающих окружностей с шумом.
2. Сравнивается модели: с заданием априорного распределения и без него.

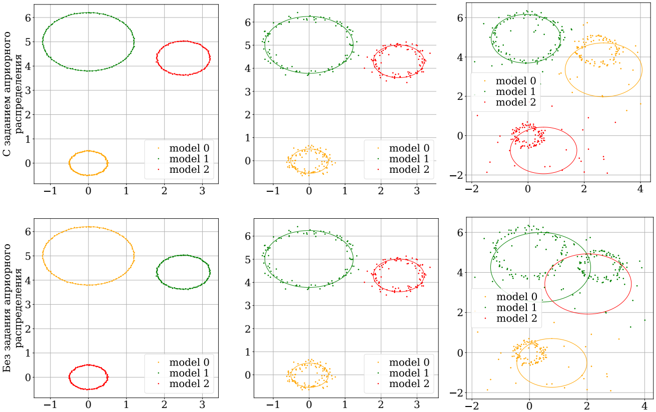
Эксперимент с разным уровнем шума в данных:

1. Синтетическое изображение двух парабол.
2. Анализ качества аппроксимации S от уровня шума β в данных и от параметра априорных распределений γ. Качество аппроксимации следующее:
$$S = ||\mathbf{w}_1^{\text{pred}} - \mathbf{w}_1^{\text{true}}||_2^2 + ||\mathbf{w}_2^{\text{pred}} - \mathbf{w}_2^{\text{true}}||_2^2.$$

Аппроксимация радужки глаза

1. Реальные изображения радужки глаза с предобработкой для их бинаризации.
2. Сравниваются различные регуляризаторы:
   - $R_0\bigl(\mathbf{V}, \mathbf{W}, E(\Omega)\bigr) = 0;$
   - $R_1\bigl(\mathbf{V}, \mathbf{W}, E(\Omega)\bigr) = -\sum_{k=1}^{K} \mathbf{w}_k^{\mathsf{T}}\mathbf{w}_k;$
   - $R_2\bigl(\mathbf{V}, \mathbf{W}, E(\Omega)\bigr) = -\sum_{k=1}^{K} \mathbf{w}_k^{\mathsf{T}}\mathbf{w}_k + \sum_{k=1}^{K}\sum_{k'=1}^{K}\sum_{j=1}^{2}\left(w_k^j - w_{k'}^j\right)^2.$

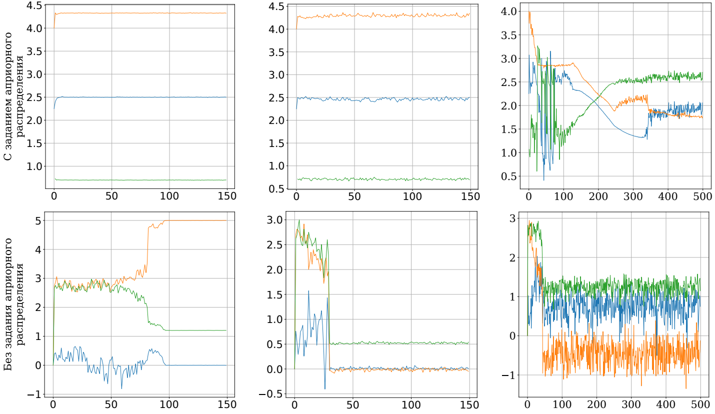
Эксперимент с окружностями



1. Сверху вниз: с заданием априорного распределения; без задания априорного распределения.
2. Слева на право: без шума; шум в радиусе; шум в радиусе окружности и произвольные точки по всему изображению.

При добавлении шума на изображение, качество аппроксимации значительно ухудшается.

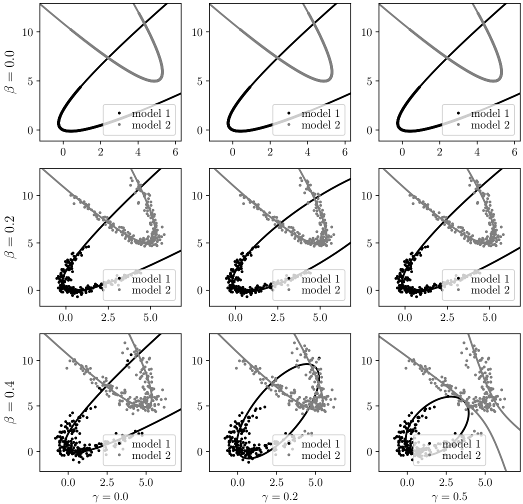
Эксперимент с разным уровнем шума



1. Сверху вниз: с заданием априорного распределения; без задания априорного распределения.
2. Слева на право: без шума; шум в радиусе; шум в радиусе окружности и произвольные точки по всему изображению.

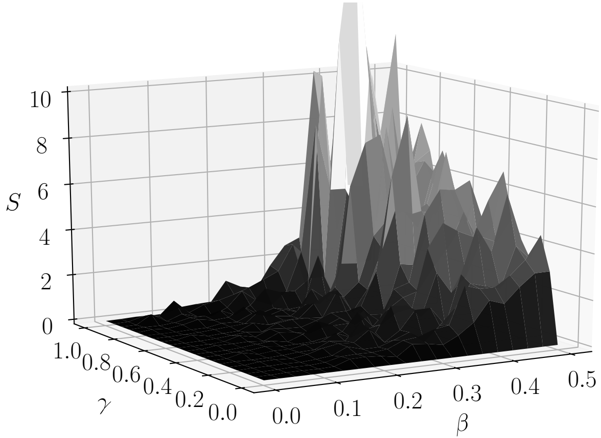
Модель с заданием априорного распределения является более устойчивой, чем аналогичная без него.

Эксперимент с разным уровнем шума



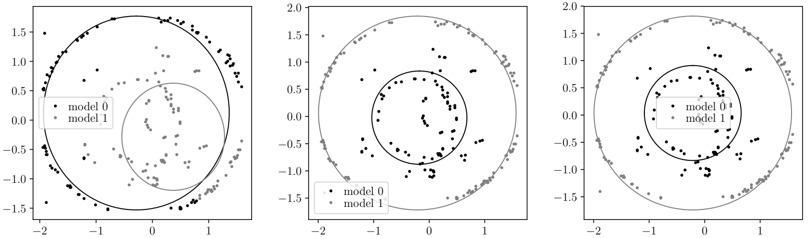
При малом шуме качество аппроксимации не зависит от параметра γ, при увеличении уровня шума, качество аппроксимации зависит от параметра γ.

Эксперимент с разным уровнем шума



При увеличении шума β в данных ошибка аппроксимации S увеличивается. Параметр γ не сильно влияет при фиксированном параметре β.

Эксперимент с аппроксимаций радужки глаза



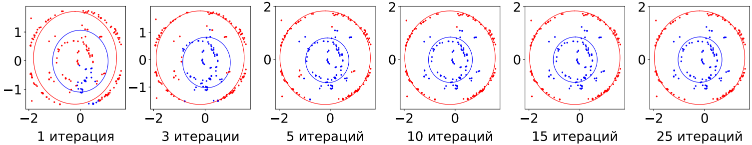
(a) (b) (c)

Аппроксимация радужки глаза: a) в случае, если задан регуляризатор R_0; b) в случае, если задан регуляризатор R_1; b) в случае, если задан регуляризатор R_2.

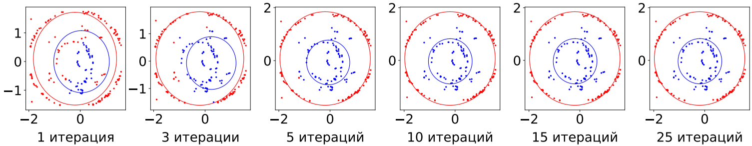
При увеличении сложности регуляризатора с R_0 до R_2 качество аппроксимации улучшается.

Визуализация сходимости

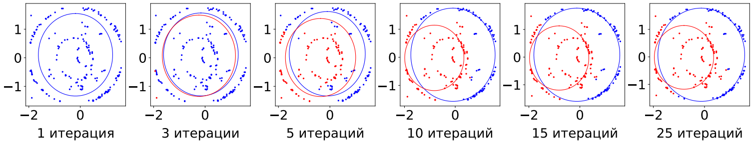
Регуляризация априорных распределений



С заданием априорного распределения на модели



Без задания априорного распределения



Заключение

1. Поставлена задача обучения с экспертной информацией.
2. Предложен метод решения задачи обучения с экспертной информацией.
3. Приведен частный случай обучения с экспертной информацией для решения задачи поиска кривых второго порядка.
4. Введено понятие регуляризации априорных распределений для улучшения качества мультимодели.
5. Проведен вычислительный эксперимент для анализа предложенной модели.

Планируется:

1. Улучшить мультимодель при помощи задания априорного распределения на шлюзовую функцию.
2. Рассмотреть в качестве локальных моделей не только модели, которые описывают данные, а также модель, которая отвечает за шум в данных.
3. Провести адаптация предложенного метода для методов глубокого обучения.

Публикации ВАК по теме

1. *Грабовой А. В., Бахтеев О. Ю., Стрижов В. В.* Определение релевантности параметров нейросети // Информатика и ее применения, 2019, 13(2).
2. *Грабовой А. В., Бахтеев О. Ю., Стрижов В. В.* Введение отношения порядка на множестве параметров аппроксимирующих моделей // Информатика и ее применения, 2020, 14(2).
3. *A. Grabovoy, V. Strijov.* Quasi-periodic time series clustering for human. Lobachevskii Journal of Mathematics, 2020, 41(3).
4. *Грабовой А. В., Стрижов В. В.* Анализ выбора априорного распределения для смеси экспертов // Журнал Вычислительной математики и математической физики, 2021. 61(5).
5. *Грабовой А. В., Стрижов В. В.* Анализ моделей привилегированного обучения и дистилляции // Автоматика и телемеханика, 2021 (текущая работа, на рецензировании).
6. *T. Gadaev, A. Grabovoy, A. Motrenko, V. Strijov* Numerical methods of minimum sufficient sample size estimation for linear models // in progress.
7. *Базарова А. И., Грабовой А. В., Стрижов В. В.* Анализ свойств вероятностных моделей в задачах обучения с экспертом // подано.