

Моя первая научная статья

Вадим Викторович Стрижов

Кафедра интеллектуальных систем МФТИ

2022

Научиться точно и ясно сообщать свои и чужие идеи.

Задачи:

- ▶ спланировать и выполнить научное исследование в области машинного обучения,
- ▶ выполнить вычислительный эксперимент с анализом ошибки,
- ▶ представить полученные результаты в виде научной статьи и доклада.

Ожидаемый результат: подать подготовленную работу к публикации в рецензируемом научном журнале.

План на mlp.org

Потеря информации при передаче сообщения¹

Небольшая группа программистов работает над новым проектом. Сколько времени пройдет, прежде чем

- 1) в группе выработается свой уникальный лабораторный жаргон,
- 2) новый сотрудник сможет разобраться, чем занимается группа,
- 3) руководитель группы перестанет понимать ход проекта,
- 4) каждый член группы перестает понимать, чем занимаются его коллеги?

¹ в отсутствие планирования

Критерий качества сообщения

В группе исследователей мы быстро приходим к соглашению

даже по сложным вопросам.

Исследователь-аналитик в коммерческой компании



Director



Customer



Analyst



Expert

Исследователь-аналитик в стартапе



Startup team



Investors

Исследователь-аналитик в научной группе



Research team



Fund, company

До начала планирования исследования аналитик и (эксперт) обсуждают ключевые вопросы

1. Цель проекта. (Ожидаемый результат разработки.)
Ожидаемая цель исследования.
2. Прикладная задача, решаемая в проекте. (Как результат будет использован?) **Чем результат будет проиллюстрирован?**
3. Описание исторических измеряемых данных. (Форматы баз данных.) **Алгебраические структуры и статистические гипотезы.**
4. Критерии качества. (Как измеряется качество полученного результата, что будет в отчете?) **Функция ошибки, что будем оптимизировать.**
5. Выполнимость проекта. (Как показать, что проект выполним, список возможных рисков.) **План анализа ошибки.**
6. Условия, необходимые для успешного выполнения проекта. (Организация работ.) **Требования к выборке.**
7. Методы решения. (Библиотеки процедур.) **Поставленные гипотезы, оптимальные вероятностные модели.**

НИР или ОКР? Новизна или технологичность

Эксперт:

(Как долго будет эксплуатироваться модель? Что заменит ее в дальнейшем?)

Аналитик:

**Какое влияние окажет исследование на область знаний?
Насколько она будет полезна?**

Риски научного исследования студента

1. Я хочу заниматься темой X.
2. Вложу усилия, а у меня ничего не выйдет.
3. Как повысить шансы на успех?

Риск управляется научным руководителем при планировании.

Самая интересная научная работа для меня — разобраться в новой научной статье и запрограммировать ее.

Покупая птицу, проверьте, есть ли у нее зубы

Является ли ваша деятельность **научным** исследованием?

Проект — это либо наука, либо коммерция, либо самообразование. Дипломная работа является квалификационной, она должна показывать квалификацию студента.

Положение о выпускной квалификационной работе студентов МФТИ

Содержательная часть рецензии содержит заключение об

- 1) актуальности (**и новизне**) исследования,
- 2) **достоверности** результатов исследования,
- 3) теоретической и практической **значимости** полученных результатов,
- 4) основных результатах исследования (**личный вклад**),
- 5) положительных сторонах и недостатках исследования (**единство исследования**).

Формула требований

Научность, личный вклад, единство, новизна, достоверность

Критерий выбора

Максимум значимости при минимуме усилий.

Смотри как построены такие работы.

Например, Neur ODE 2018 и работы Понтрягина 1962.

Чек-лист руководителя при планировании

Ясно видеть

- 1) результат,
- 2) реакцию научного сообщества на него.

Прочее — в Положении о ВКР.

За какую задачу браться?

1. Масштабность проблемы: решение проблемы должно касаться большого числа людей, специалистов, лиц принимающих решения.
2. Зброшенность (популярность) проблемы. Общая ошибка: решать популярные проблемы.
3. Решаемость проблемы. Выбор просто и элегантно решаемых проблем.
4. Наша готовность к решению проблемы, квалификация: похожими проектами мы уже занимались.*

Пункт 4. Иначе риском является студент, а не проект.

Синдром FoMO при выборе темы исследований

«Все, кроме меня»

1. Программируют на Питоне
2. Используют нейросети
3. Знают, что такое NeurODE
4. ...

Плюс — актуальность (даже при дальнейшей тупиковости)

Минус — нет новизны (подготовка и публикация занимает 1–3 года) **Полезно** следить за призовыми публикациями и их судьбой

Исследование как спорт

Цель — показать на новой модели качество (precision, F1, AUC, ...) выше, чем у альтернатив.

Насколько быстро и каким способом будет побит твой рекорд?

Достаточное условие признания квалификации

Публикация в **рецензируемом** журнале

Выбрать нужное в списке ArXiv, ВАК, Скопус/Конференции, WoS, Core, опубликовать, наклеить обложку, — диплом готов!

До момента выбора студентом темы

1. Понимание объекта исследования: что именно развиваем, метод или код, какой именно
2. Точное целеполагание: каких именно новых результатов ждем
3. Формальная постановка задачи вместе с критериями качества ее решения: обоснованность предполагает анализ ошибки
4. Что выносится на защиту: что студент объявит как личный значимый вклад

Это требует квалификации научного руководителя. Квалификации студента для этого недостаточно.

Торжественный комплект — универсальная инструкция по изготовлению исследований

Планируем исследование по мере возрастания риска

- 1) повторить чужую статью (анализ свойств алгоритма),
- 2) старую задачу решить старым способом (пара “задача–способ решения” не была найдена в литературе),
- 3) прочие три варианта предыдущего пункта,
- 4) ...
- 5) применить новые эвристики в оптимизации (“рационализаторское предложение”)
- 6) наконец-то выписать все предположения о порождении выборки и вывести вероятностную модель
- 7) применить методы абстрактной математики и теоретической физики к методам машинного обучения

По мотивам Ильи Ильфа и Евгения Петрова «Золотой телёнок».

NB! Все эти способы призваны повышать достоверность научных результатов.

Признаки **качественной статьи**

1. Рассматривается одна выделенная проблема, имеющая новое, ясное решение.
2. Решение имеет теоретическое обоснование. Свойства предлагаемого алгоритма исследованы.
3. Статья проиллюстрирована интересными прикладными примерами, поставлен вычислительный эксперимент, сделан анализ ошибки.
4. Студент понимает, кто является читателем этой работы, и готов, на основании анализа публикаций, назвать журнал, в который она будет подана.

Текст, скорее всего, **статьей** не является

1. Известная задача решена одним из известных способов, но подобная пара в литературе не встречалась (отсутствие новизны).
2. Задача посвящена повышению значения единственного критерия качества на выборках из UCI (невостребованность).
3. Для решения задачи данные обрабатываются последовательностью эвристических процедур (необоснованность).

О, ужас!

- ▶ доклад посвящен обзору известных решений, в конце невнятный эксперимент
- ▶ целеполагание неясно, пересказ модной работы
- ▶ работа является сборной солянкой из проектов студента
- ▶ терминология не проработана, термины не определены
- ▶ теория и формулы не связаны с кодом, так как студент не понимает значения формул, а код для него черный ящик
- ▶ личный вклад неясен, невозможно отделить то, чем пользовался студент от того, что он сделал
- ▶ цитируются схемы, значение которых студенту не ясно
- ▶ нет анализа ошибки, неясно, верить ли таблицам
- ▶ практика противоречит теоретическим предположениям по неосведомленности

Низкое качество работы более всего заметно в докладе.

Резюме

Математика — наука цеховая

Способ получения научных знаний — обсуждение

оно требует квалификации

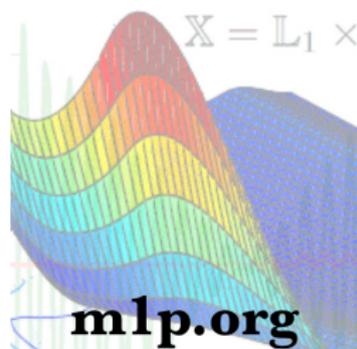
Если что-то можно не делать, то этого делать не нужно

Publish or perish

The course produces student research papers

Machine learning projects: how to

- ▶ state the problem,
- ▶ make the project feasible,
- ▶ present results of the experiment



Science requires community:

- ▶ **Student** is a project driver, who wants to plunge into scientific research activities.
- ▶ **Consultant**, a graduated student, conducts the research and helps the student.
- ▶ **Expert**, a professor, states the problem and enlightens the road to the goal.

Lectures and seminars

Lecture

- ▶ Theoretical part: Machine learning for researchers
- ▶ Practice: The routine of scientific research and homeworks

Seminar

- ▶ Collective games
- ▶ Analysis of projects
- ▶ Talks

We read aloud every paper to prepare it for publication

Four talks to convey your message to the audience

Week 3 Introductory pitch

6 The message

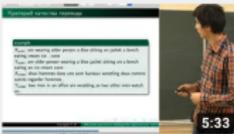
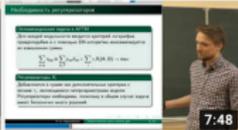
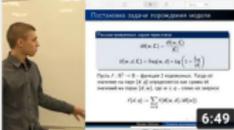
9 Computational experiment

12 Conference talk



YouTube channel: Machine Learning



 <p>7:13</p>	 <p>6:19</p>	 <p>4:15</p>	 <p>5:33</p>	 <p>3:51</p>
Евгения Веселова: Reliable detection of tandem repeat... 59 views • 1 year ago	Наталья Вареник: Сферические сверточные... 133 views • 1 year ago	Игорь Юсупов: Динамическое... 55 views • 1 year ago	Саттаров Тагир: Машинный перевод без параллельны... 42 views • 1 year ago	Алексей Гришанов: Автоматическая настройк... 74 views • 1 year ago
 <p>6:28</p>	 <p>7:48</p>	 <p>5:33</p>		 <p>6:56</p>
Алексей Григорьев: Распознавание... 29 views • 1 year ago	Николай Герасименко: Тематический поиск... 23 views • 1 year ago	Абдурахмон Садиев: Порождение признаков с... 81 views • 1 year ago		Глеб Моргачев, Владислав Смирнов: Динамическое... 62 views • 1 year ago
 <p>6:03</p>	 <p>4:57</p>	 <p>4:25</p>	 <p>6:49</p>	 <p>4:35</p>
Александр Рубинштейн: Scoring function for protein...	Михаил Макаров: Определение...	Антон Черепков: Привилегированное...	Михаил Лепехин: Ранжирующе модели...	Валентин Бучнев: Раннее прогнозирование...

Roadmap

1. Set the toolbox
2. Select your project
3. Read papers
4. Write introduction
5. State the problem
6. Set your experiment
7. Develop your theory
8. Make error analysis
9. Paper draft
10. Share your results
11. Finalize your paper
12. Present your talk

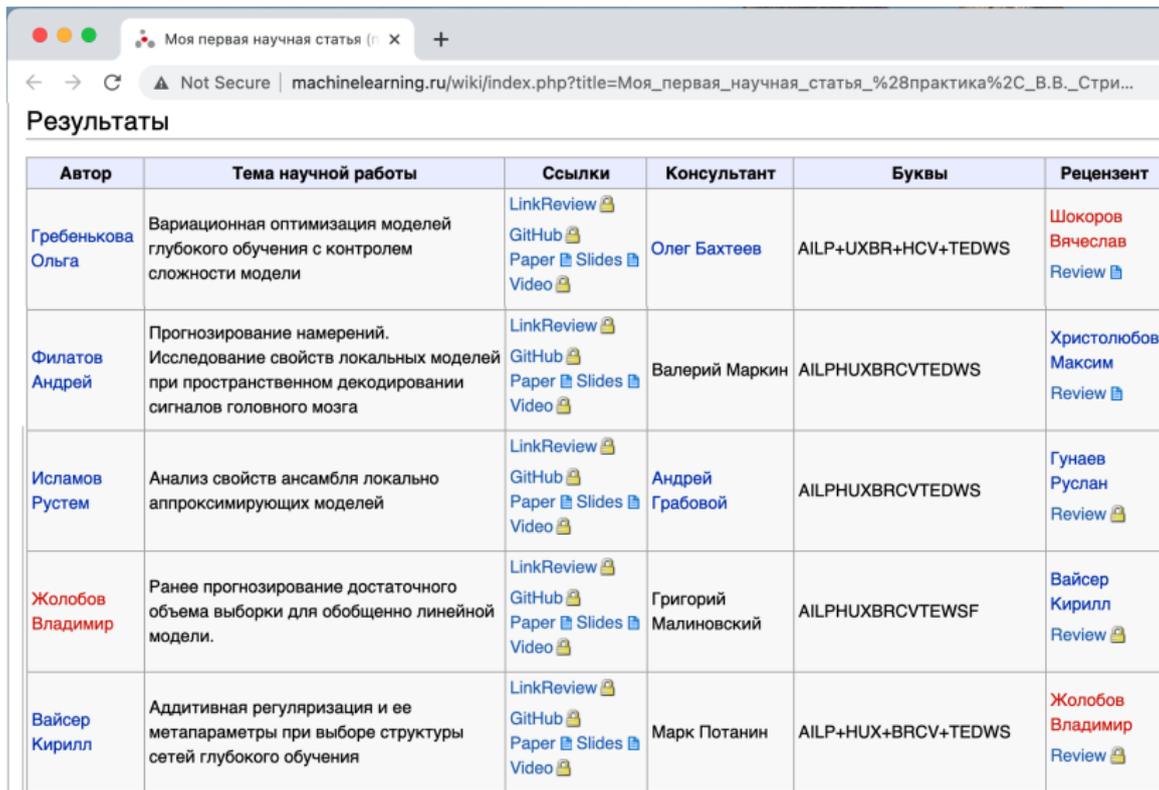


Deliveries are scheduled

- ▶ [LinkReview](#) with references and the literature review
- ▶ [GitHub](#) with the code and computational experiment
- ▶ [Paper](#) is ready for submission
- ▶ [Slides](#) for the presentation
- ▶ [Video](#) of the conference talk



Table with links to the deliveries and the marks



The image shows a browser window with the address bar containing the URL: machinelearning.ru/wiki/index.php?title=Моя_первая_научная_статья_%28практика%2C_В.В._Стри... The page title is "Моя первая научная статья". Below the browser is a table with the following columns: Автор, Тема научной работы, Ссылки, Консультант, Буквы, and Рецензент. The table contains five rows of data, each representing a different author and their work.

Автор	Тема научной работы	Ссылки	Консультант	Буквы	Рецензент
Гребенькова Ольга	Вариационная оптимизация моделей глубокого обучения с контролем сложности модели	LinkReview GitHub Paper Slides Video	Олег Бахтеев	AILP+UXBR+HCV+TEDWS	Шокоров Вячеслав Review
Филатов Андрей	Прогнозирование намерений. Исследование свойств локальных моделей при пространственном декодировании сигналов головного мозга	LinkReview GitHub Paper Slides Video	Валерий Маркин	AILPHUXBRCVTEWDS	Христолюбов Максим Review
Исламов Рустем	Анализ свойств ансамбля локально аппроксимирующих моделей	LinkReview GitHub Paper Slides Video	Андрей Грабовой	AILPHUXBRCVTEWDS	Гунаев Руслан Review
Жолобов Владимир	Ранее прогнозирование достаточного объема выборки для обобщенно линейной модели.	LinkReview GitHub Paper Slides Video	Григорий Малиновский	AILPHUXBRCVTEWSF	Вайсер Кирилл Review
Вайсер Кирилл	Аддитивная регуляризация и ее метапараметры при выборе структуры сетей глубокого обучения	LinkReview GitHub Paper Slides Video	Марк Потанин	AILP+HUX+BRCV+TEDWS	Жолобов Владимир Review

Мы диктуем формат. Творчество ваше.



Информация о курсе

Telegram



t.me/m1p_org

Info



m1p.org

YouTube



Machine Learning

Вадим Викторович Стрижов, МФТИ

Автоматизация научных исследований

mlalgorithms@gmail.com

<http://www.machinelearning.ru/wiki/index.php?title=m1>

ПОИСК

Перейти

Найти

The project topics

Any topic is welcome on the open access basis!

- ▶ Analysis of wearable devices' data
- ▶ Brain computer interface modelling
- ▶ Image and symbol recognition
- ▶ Synchrotron image analysis
- ▶ CERN LHCb data analysis
- ▶ Receptor-ligand docking
- ▶ Chemical reaction forecasting
- ▶ Multiscale time series forecasting
- ▶ Expert learning and knowledge distillation
- ▶ Neural architecture search
- ▶ Hyper-networks for model selection
- ▶ Model generation, processes
- ▶ Stochastic model selection
- ▶ Metaparameter optimisation
- ▶ Stochastic optimisation
- ▶ Spatial and time series alignment