Министерство образования и науки Российской Федерации

Государственное образовательное учреждение высшего профессионального образования «Московский физико-технический институт (государственный университет)»;

Центр «ПУСК» МФТИ

Программа «Науки о данных»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

по направлению 01.04.02 «Прикладная математика и информатика»

Тема: «ML-моделирование перевода с русского жестового на русский словесный»

(рабочее название)

Студент:

Кудрявцева П. Д.

Научный руководитель:

Каприелова М. С.

Москва 2024

**Содержание**

[Введение 4](#_Toc134231723)

[Список литературы 8](#_Toc134231724)

Введение

Общение – важная составляющая в жизни любого человека. Все мы испытываем потребность делиться с другими своими мыслями, чувствами, точками зрения, хотим быть понятыми и услышанными. Также общение важно и для развития человека – мы многому учимся именно через общение с другими людьми: родными, близкими, учителями, друзьями, коллегами.

Глухие и слабослышащие люди зачастую сталкиваются с определёнными сложностями в общении с людьми слышащими. Основное средство коммуникации между глухими и слабослышащими – это жестовые языки. Далеко не каждый слышащий человек знаком с жестовым языком своих соотечественников. Для глухих же научиться понимать (посредством чтения по губам) и говорить (не слыша самих себя) – весьма сложная задача, которая требует кропотливой работы как с его стороны, так и со стороны преподавателя. Более того, в некоторых случаях (например, при общении с врачом в маске) человек оказывается лишённым даже возможности прочитать по губам.

Статистические данные свидетельствуют о постоянном росте числа новорожденных с нарушениями слуха. По данным Всемирной федерации глухих, более 300 жестовых языков используются более чем 700 миллионами людей по всему миру. По данным ВОЗ, количество глухих и слабослышащих людей в мире увеличивается год от года:

* порядка 278 миллионов человек в 2005 году,
* порядка 360 миллионов человек в 2015 году (+14%),
* порядка 466 миллионов человек в 2019 году (+13%),
* к 2050 году ВОЗ прогнозирует, что это число удвоится.

Традиционным посредником в общении между глухими/слабослышащими людьми и слышащими людьми выступает сурдопереводчик. По законодательству РФ, услуги сурдоперевода частично оплачиваются за счёт федерального бюджета. Однако объём услуги, которую глухой человек может получить за счёт государства относительно небольшой – максимум 40 часов в год. Кроме того, сурдопереводчики – достаточно редкая профессия: по данным Министерства просвещения РФ[[1]](#footnote-1), в 2020-м году в России насчитывалось около 900 сурдопереводчиков, при этом на каждого из них приходилось до 100 тысяч глухих. Также посредничество постороннего человека, может быть не всегда желательно для глухого человека: например, в вопросах, касающихся его здоровья.

Со стороны медицины также постоянно ведутся разработки технологических решений, помогающим глухим и слабослышащим людям ориентироваться в мире звуков. На настоящее время самым эффективным решением в этом направлении является кохлеарная имплантация. Однако она подходит не всем. Кроме того, слышат люди с имплантами несколько иначе, чем слышащие, и вынуждены какое-то время учиться воспринимать и различать звуки с помощью импланта.

Ещё одним мостиком для общения между глухими и слышащими людьми мог бы стать автоматический переводчик с жестового языка на словесный и в обратном направлении. Несмотря на впечатляющие результаты, достигнутые учёными и разработчиками в этом направлении, необходимое качество такого перевода в настоящий момент не достигнуто. Глухие предъявляют следующие требования к такой системе автоматического перевода [Гриф и др., 2021]:

* точность перевода не менее 90%,
* качество визуализации должно быть максимально приближено к движениям человека-сурдопереводчика.

Почему же при текущем уровне развития машинного перевода с одного жестового языка на другой, для жестовых языков удовлетворительный результат пока ещё не достигнут? С какими сложностями сталкиваются исследователи, работающие с жестовыми языками?

Человек, говорящий на жестовом языке, использует для передачи своего сообщения не только движения рук (которые сами по себе допускают определённую вариативность), но также и движение корпусом (например, при сравнении), выражение лица (в РЖЯ, например, при вопросе необходимо нахмурить брови), «проговаривание» слов губами (чтение по губам может помочь различить похожие между собой жесты, однако не все глухие задействуют губы), движения головой (например, в РЖЯ при вопросе голова как правило движется немного вперед в сторону собеседника), позиции тела (например, при пересказе разговора двух и более лиц). Таким образом, каждый из указанных компонентов несёт в себе определённую смысловую нагрузку, а значит, при упрощении задачи только до движений руками, теряется существенная часть информации, а необходимость учитывать все факторы приводит к усложнению задачи.

С развитием ИИ, и особенно глубокого обучения, исследователи всё больше внимания стали уделять построению именно интеллектуальных систем перевода для жестовых языков. Среди работ последнего десятилетия чётко прослеживается популярность методов глубокого обучения. Например, работы на базе набора данных RWTH-PHOENIX-Weather 2014 чаще всего строят распознавание жестов на базе свёрточной нейронной сети (CNN). Те же работы, которые сфокусированы не на распознавании отдельных жестов, а на переводе непрерывной жестовой речи, чаще всего дополняют CNN лингвистической моделью (как правило, BLSTM) [Koller, 2020]. Тут следует отметить, что на данный момент большая часть работ всё-таки ориентирована на распознавание отдельных изолированных жестов, а не на перевод непрерывной жестовой речи.

В качестве входных данных сейчас наиболее популярны RGB-изображения или видео. Также есть варианты получения данных с сенсоров: электронные перчатки, технологии захвата движения, Kinect), но они существенно более дорогие, и сложно представить их массовое использование при работе с автопереводчиком.

Степень популярности того или иного языка у исследователей определяется в основном наличием общедоступного размеченного набора данных. Лидером по количеству публикаций в настоящий момент выступает американский жестовый язык. Но это верно только для небольших наборов данных (со словарём менее 200 жестов). На более крупных наборах лидирует уже китайский жестовый, а на наборах со словарём более 1000 жестов (RWTH-PHOENIX-Weather 2014) – немецкий жестовый [Koller, 2020].

Немануальные составляющие жестового языка (такие как: движения головы, корпуса, губ, бровей) чаще всего учитываются в работах именно по немецкому жестовому языку [Koller, 2020]. Исследователи РЖЯ также довольно часто принимают во внимание немануальные параметры. Однако в последние годы в мировой практике есть тренд к работе с полнокадровыми изображениями (когда говорящий на жестовом языке полностью попадает в кадр и все его движения рассматриваются как единое целое).

Подробнее о существующих в настоящий момент наборах данных, основных наработках и о трендах в подходах к распознаванию и переводу с жестовых языков (в том числе РЖЯ) см. главу 1.

Исследований по РЖЯ в настоящий момент существенно меньше, чем по американскому, немецкому или китайскому жестовым языкам. Одна из главных проблем тут – отсутствие общедоступных размеченных наборов данных, достаточно больших по объему и словарному запасу. Но это не значит, что нужно отказываться от исследований по тем наборам, которые доступны на данный момент.

**Объект исследования настоящей работы:** возможности автоматизации перевода с русского жестового языка (РЖЯ) на словесный русский язык.

**Предмет исследования:** разработка подходов к автоматическому распознаванию непрерывной речи на РЖЯ и её переводу на словесный русский язык методами глубокого обучения.

**Цель исследования:** обучить нейронную сеть (или набор нейронных сетей), которая будет показывать приемлемою точность перевода непрерывной жестовой речи в видеоформате на русский словесный язык.

**Задачи исследования:**

1. Изучить международную практику в области интеллектуального машинного перевода с жестовых языков, а также наработки отечественных исследователей в отношении РЖЯ.
2. Выбрать несколько наиболее популярных/показывающих хорошую точность архитектур для перевода непрерывной жестовой речи и применить их к поставленной задачи. И применить их к выбранному набору данных.
3. Провести сравнительный анализ полученных результатов.
4. Выбрать модель с лучшим результатом и постараться определить, за счёт каких факторов результат получился лучше.

В качестве набора данных для исследования возьмём Корпус русского жестового языка [Электронный ресурс][[2]](#footnote-2), разработанный под руководством С. И. Бурковой в Новосибирском государственном техническом университете. Подробнее о характеристиках корпуса и его сравнение с другими доступными наборами данных, подходящими для решения поставленной задачи, см. главу 1.

* методологический аппарат исследования: TBD;
* основные пункты новизны, выносимой на защиту: TBD;
* теоретическая и/или практическая значимость исследования: TBD;
* апробация исследования (если таковая имеется) : TBD;
* краткая характеристика структуры диссертации: TBD.

Глава 1. Обзор публикаций по теме

* 1. Особенности жестовых языков и РЖЯ в частности

1. Жестовый язык - та же функция, что и у звучащего. Но есть отличия - свой словарь, своя грамматика, своя морфология и фонология. 5 базовых параметров: форма руки, локаця руки, ориентация ладони, движение руки, выражение лица. Значения жеста может поменяться при изменении одного из этих параметров. Жестовые языки менее универсальны, различаются по географическим локациям, национальным, социальным группам и словарям. Страны с общим звучащим языком, можут иметь разные жестовые языки (развивались независимо, так как не общались между собой).
2. Сложности: скорость жестикуляции различается у людей. Внешняя среда влияет на то, как жест выглядит на изображении.
3. Диалекты РЖЯ
4. Особая сложность - задача распознавания непрерывной жестовой речи. При выделение жестов необходимо учитывать комбинаторные изменения параметров жестов, а также эпентезы.
   1. Наборы данных в области распознавания и перевода с жестовых языков

RWTH-PHOENIX-Weather 2014 – стандартная контрольная задача

* 1. 1.3 Подходы к решению задачи в зарубежной и отечественной литературе

**[Koller, 2020]**

Начало работы над темой – 1983 год ([Grimes, 1983])

Смена подходов к решения с 2015 года

Переход от интрузивных к неинтрузивным методам

До 2015 не было больших наборов данных. После 2015 работа ведётся преимуществоенно с двумя базовыми корпусами: 1080 и 178 жестов.

Большинство работ нацелены на распознавание изолированных жестов (почти в два раза больше, чем для непрерывной жестовой речи).

Входные данные:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Метод | Описание | Интрузивность | Популярность | Языки | Стоимость | Ключевые даты |
| RGB изображения | Зрение: фичи, относящиеся к ладоням, положению пальцев и углов между ними оцениваются и потом используются для распознавания. Здесь необходимы изображения или видео, а также обработка этих изображений. | - | Самый популярный | Американский жестовый,  Китайский жестовый  Немецкий жестовый |  | Набирает популярность с 2005 года |
| RGB-D изображения |  | - |  | Китайский жестовый |  | Глубина стала популярна после выхода сенсоров Kinect в 2010 году. |
| разноцветные перчатки |  | + | Почти не используется |  |  |  |
| электронные перчатки |  | + |  |  |  |  |
| захват движения |  | + |  |  |  |  |

С 2005 года переход от интрузивных к неинтрузивным. Всё больше и больше используются изображения.

Мануальные параметры:

* Форма руки – самый популярный параметр (на маленьких датасетах, до 1000)
* Движение – 2 популярноть (на маленьких датасетах, до 1000)
* Локация – 2 популярность (на маленьких датасетах, до 1000)
* Ориентация

Немануальные параметры:

* Голова
* Губы
* Глаза
* Брови
* Моргание и взгляд

Полное представление говорящего на жестовом языке (суставы тела, RGB изображение, захватывающее всего человека целиком, такие же полнорамочные изображения с глубиной, полнорамочные движущиеся изображения с использованием оптического потока).

На больших датасетах (более 1000 знаков) чаще всего анализируются полнорамочные изображения.

С 2005 полнорамочный подход набирает популярность и приближается к масштабам исследований формы руки.

С ростом датасетов тренд от мануальных параметров к глобальным фичам. Возможные причины:

- Доступность суставов тела и полнорамочных изображений с выходом Kinect в 2010

- Сдвиг в сторону глубокого обучения и тренд к вводу полнорамочных изображений вместо ручного фиче инжиниринга.

[Гриф]

Препятствия:

неполнота описания грамматической системы РЖЯ

отсутсвие "плавности" показа жестов аватаром-переводчиком РЖЯ, система управления которого использует систему нотаций языков глухих

перевод осуществляется преимущественно на калькирующую жестовую речь, а не на РЖЯ, обладающий выразительными возможностями

высокий процент ошибок при переводе многозначных слов и омонимов на жесты РЖЯ (более 20%)

* Основные технологи основаны на: компьютерном зрении и на сенсорных-перчатках (и их аналогов)
* Перчатки: механические или оптические сенсоры, приреплённые к перчаткам, которые надевает пользователь. движения пальцев трансформируются в электрический сигнал, почзволяющий определь положение рук для распознавания.

Список литературы

**Нормативно-правовые акты**

Постановление Правительства РФ от 25 сентября 2007 г. N 608 "О порядке предоставления инвалидам услуг по переводу русского жестового языка (сурдопереводу, тифлосурдопереводу)".

**Статьи**

Гриф М.Г., Элаккия Р., Приходько А.Л., Бакаев М.А., Раджалакшми Е. Распознавание русского и индийского жестовых языков на основе машинного обучения // Системы анализа и обработки данных. - 2021. - №3(83). - С. 53-74.

URL: <https://journals.nstu.ru/vestnik/catalogue/contents/view_article?id=27523>

Гриф М.Г., Королькова О.О., Приходько А.Л. Распознавание жестовой речи с учётом комбинаторных изменений жестов // Материалы XXI Международной научно-методической конференции «Информатика: проблемы, методы, технологии». - В.: Вэлборн, 2021. С. 1387-1393.

Кагиров И.А., Рюмин Д.А. База данных русского жестового языка поликлинического предназначения: лингвистические особенности материала и аннотирования. // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2022 - №20(3). – С. 90-108.

URL: <https://doi.org/10.25205/1818-7935-2022-20-3-90-108>

Adeyanju I.A., Bello O.O., Adegboye M.A. Machine learning methods for sign language recognition: A critical review and analysis // Intelligent Systems with Applications – 2021. – Vol.12. – P.2667-3053.

URL: <https://doi.org/10.1016/j.iswa.2021.200056>

Koller O. Quantitative survey of the state of the art in sign language recognition // <https://arxiv.org/abs/2008.09918v2> - 2020.

**Интернет-источники**

«В России отмечают День сурдопереводчика» / Пресс-служба [Минпросвещения России](https://edu.gov.ru/), 2020. URL: <https://edu.gov.ru/press/3084/v-rossii-otmechayut-den-surdoperevodchika/>

**Наборы данных**

Корпус русского жестового языка [Электронный ресурс] / Руководитель проекта С. И. Буркова – Новосибирск – 2012–2015 – Режим доступа: [http://rsl.nstu.ru/](http://rsl.nstu.ru/site/signlang)

1. «В России отмечают День сурдопереводчика» / Пресс-служба [Минпросвещения России](https://edu.gov.ru/), 2020. URL: https://edu.gov.ru/press/3084/v-rossii-otmechayut-den-surdoperevodchika/ [↑](#footnote-ref-1)
2. http://rsl.nstu.ru//site/signlang [↑](#footnote-ref-2)