

ОТЗЫВ

официального оппонента на диссертационную работу Животовой Алены Анатольевны «Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык», представленную к защите на соискание ученой степени кандидата технических наук по специальности 1.2.2. – «Математическое моделирование, численные методы и комплексы программ»

Структура и объем работы.

Представленная работа состоит из введения, четырех основных глав с выводами, заключения, библиографического списка, состоящего из 124 наименований и двух приложений; изложена на 125 страницах; содержит 13 таблиц и 22 рисунка. В приложение включены: акт внедрения результатов диссертационного исследования в деятельность ООО «Агентство переводов «ФИАС-Амур» № 6/23/1 от 10.06.2023; копии свидетельств о государственной регистрации 6 программ ЭВМ, 2 программных комплексов и 2 баз данных.

Актуальность темы.

Исследование посвящено оптимизационному предредактированию русскоязычных узкоспециальных текстов для повышения качества машинного перевода на английский язык. Несомненная актуальность обусловлена растущей потребностью в сохранении высокой точности перевода с учетом повышения скорости его воспроизводства в современной реальности, в которой системы машинного перевода вытесняют переводчиков в пользу постредакторов и корректоров. При этом, для ряда узкоспециальных предметных областей, таких как нефтегазовый сектор, собрать достаточный объем параллельных тренировочных данных для обучения систем машинного перевода проблематично ввиду ограничений конфиденциальности данных. Среди российских авторов данная тема освещена незначительно, в отличие от зарубежных исследований. Недостаточный уровень анализа методов и алгоритмов предредактирования русскоязычных текстов для повышения качества их машинного перевода определяет теоретическую значимость представленной работы.

Краткая характеристика содержания диссертационной работы.

Анализ содержания диссертационной работы производит исключительно положительное впечатление тщательно проведенного исследования, которое включает все ключевые компоненты, предусмотренные для специальности 1.2.2. – «Математическое моделирование, численные методы и комплексы программ». Структура работы логично организована и охватывает все ключевые аспекты, необходимые для глубокого понимания предложенных методов и алгоритмов.

Основные главы диссертационной работы:

1. *Предредактирование как перспективное направление развития систем машинного перевода.* В этой главе проводится глубокий анализ современных подходов к решению задачи машинного перевода с использованием методов ТРИЗ, оцениваются их преимущества и недостатки, а также проблемы, связанные с переводом узкоспециальных текстов. Автор акцентирует внимание на необходимости предредактирования текстов для повышения качества перевода, что является основным фокусом исследования.

2. *Математическая модель оптимизационного предредактирования.* В данной главе представлена обоснованная математическая модель, которая позволяет формализовать процесс перевода и предредактирования. Автор описывает теоретические аспекты и методологию оценки сложности задач перевода; модель поиска весов значимости признаков текста для вероятностной оценки ожидаемого качества его перевода на целевой язык с использованием метода наименьших квадратов; постановку задачи машинного перевода; постановку задачи оптимизационного предредактирования на основе метода максимизации правдоподобия с использованием элементов нечеткой логики и ее решение методом градиентного спуска.

3. *Методика реализации оптимизационного предредактирования.* В главе рассматриваются разработанные алгоритмы предредактирования, включая вероятностную оценку сложности задач перевода, и их применимость к русскоязычным текстам для улучшения перевода на английский язык. Кроме того, описана методика обучения моделей предредактирования и оценки сложности задачи перевода, предложен алгоритм генерации тренировочных данных для модели предредактора.

4. *Программный комплекс оптимизационного предредактирования узкоспециальных русскоязычных текстов для их перевода на английский язык.* В данной главе представлена архитектура и функциональные возможности разработанного программного комплекса, который реализует предложенные алгоритмы. Описываются результаты тестирования системы на корпусе узкоспециальных текстов, результаты внедрения программного комплекса в операционные процессы компании-поставщика услуг технического перевода со специализацией в нефтегазовой предметной области, подтверждается эффективность предложенных решений.

Каждая глава сопровождается обоснованными выводами. В заключении подводятся итоги диссертационного исследования, формулируются основные результаты работы и выводы по дальнейшему их применению. Результаты и выводы исследования соответствуют поставленным целям и задачам.

Автореферат в полной мере отражает содержание диссертационной работы, в нем раскрыты поставленные цели и задачи исследования, представлены все основные результаты работы, сформулированы положения, выносимые на защиту.

Научная новизна и практическая значимость исследований.

Научная новизна диссертационной работы заключается в разработке и предложении ряда новых методик и алгоритмов, направленных на повышение качества машинного перевода узкоспециальных текстов с русского языка на английский.

В рамках представленного исследования впервые предложена методика оценки сложности переводческой задачи, которая формализует и учитывает понятия компетенции и специализации переводчика, а также параметры исходного текста, что позволяет прогнозировать риски, связанные с некачественным или несвоевременным выполнением задач перевода, и является важным аспектом для управления процессами перевода документации.

Диссертационная работа описывает новые алгоритмы, которые расширяют применение оптимизационного метода градиентного спуска и метода наименьших квадратов. Эти алгоритмы используют элементы нечеткой логики для улучшения оценки сложности задач перевода и поиска весов значимости параметров исходного текста, что способствует более точной вероятностной оценке ожидаемого качества машинного перевода.

В работе также предложена методика обратного перевода для сбора тренировочных данных для модели оптимизационного предредактирования. Важным вкладом является разработка новой архитектуры программного комплекса, который использует ансамбль моделей для оптимизационного предредактирования на основе вероятностной оценки сложности задачи перевода с целью повышения качества машинного перевода текстов с русского языка на английский язык.

Обоснованность и достоверность научных положений и выводов.

Положения, представленные в диссертационном исследовании, являются научно обоснованными, и их достоверность не вызывает сомнений.

Работа основывается на глубоком анализе существующих технологий и обширном литературном обзоре, который включает как зарубежные, так и российские исследования в области машинного перевода и предредактирования. Это позволяет автору обосновать актуальность темы и продемонстрировать глубокое понимание существующих методов и подходов, а также выявить их недостатки. Упоминание значительных исследований, таких как работы S. O'Brien, M. Yamada, R. Miyata и других, создает контекст для предложенных автором решений и подчеркивает их новизну.

В математическом моделировании автор опирается на аппарат апробированных численных методов, таких как метод наименьших квадратов и градиентный спуск, и статистических методов, включая метод максимального правдоподобия, а также теорию множеств.

В работе представлена четкая и обоснованная методология тестирования, которая позволяет объективно оценить эффективность предложенных решений. Разработанные алгоритмы и программный комплекс внедрены на предприятии лингвистической отрасли, это свидетельствует о том, что полученные результаты имеют реальное применение и могут быть использованы для повышения качества перевода машинного перевода русскоязычных узко-

специальных текстов на английский язык в условиях ограниченных объемов эталонных двуязычных корпусов.

Результаты, изложенные в диссертационной работе, отражают личный вклад соискателя, докладывались и обсуждались на множестве научных конференций, включая всероссийские и международные, и признаны научным сообществом, что подтверждается дипломами научно-практических конференций и конкурсов. Квалификационный уровень и научную ценность полученных результатов также подтверждает достаточное количество работ, опубликованных в рецензируемых научных изданиях. Общий список публикаций включает 12 научных работ, в том числе 3 работы в изданиях, рекомендованных ВАК, 3 работы в изданиях, индексируемых в международной базе Scopus.

Значимость полученных результатов для науки и практики.

Научная значимость диссертационного исследования проявляется в разработке математических моделей, которые позволяют формализовать процесс перевода и предредактирования, а также в использовании элементов нечеткой логики для расширения области применения существующих методов оптимизации, что создает основу для дальнейших исследований в области машинного перевода и предредактирования текстов. Предложенные методики оценки сложности переводческой задачи, основанные на компетенции переводчика и параметрах исходного текста, открывают новые горизонты для исследований в области прогнозирования рисков, связанных с качеством перевода. Кроме того, автором предложена новая методика, которая позволит более эффективно решать проблемы, связанные с недостатком двуязычных параллельных данных для обучения генеративных моделей.

Практическая значимость работы состоит в том, что разработанные алгоритмы и программный комплекс позволяют повысить качество машинного перевода узкоспециальных русскоязычных текстов на английский язык, что является актуальным для такой области, как нефтегазовая промышленность, где качество перевода критически важно для коммуникации и обмена технологиями. Внедрение результатов исследования в контур автоматизации процессов переводческой деятельности способствует сокращению затрат на перевод и повышению его качества за счет оптимизации исходных текстов.

Замечания по работе.

Несмотря на высокое качество работы, к ней есть несколько замечаний.

В главе 1 представлен обзор современных систем машинного перевода и их ограничений, однако не вполне ясно, как предлагаемые в работе методы оптимизационного предредактирования и вероятностной оценки сложности задач перевода решают выявленные противоречия, которые приведены в Таблице 1.4 на стр. 26-31, и для которых отсутствует решение согласно столбцу «Описание решения».

В главе 3, посвященной методологии реализации предложенных алгоритмов вероятностной оценки задачи перевода и оптимизационного предредактирования, не уделено достаточного внимания вопросам интеграции предлагаемых решений в существующие системы машинного перевода. Допущения практической реализации на стр. 75 не конкретизируют эту инфор-

мацию. Следовало бы добавить информацию о том, каким образом разработанные алгоритмы могут быть встроены в популярные платформы, такие как Google Translate, DeepL или Yandex.Translate, чтобы обеспечить их широкое практическое применение.

В главе 4 в разделе, посвященном тестированию, не представлено сравнение разработанных алгоритмов с уже существующими методами предредактирования. Полезно было бы провести анализ, в котором будут сопоставлены ключевые характеристики и результаты применения различных методик, чтобы наглядно показать преимущества разработанных автором алгоритмов и методов. Также было бы полезно провести более обширное тестирование на различных языках и текстах, чтобы оценить универсальность предложенных методов.

Вышеперечисленные замечания не меняют общего положительного впечатления от диссертационного исследования.

Общее заключение.

Диссертационная работа Животовой А. А. «Математическая модель, алгоритмы и программный комплекс для повышения качества машинного перевода узкоспециальных технических текстов на английский язык» является законченной научно-квалификационной работой на актуальную тему, в которой предложены и обоснованы методы вероятностной оценки сложности задачи перевода и автоматического оптимизационного предредактирования для повышения качества машинного перевода русскоязычных узкоспециальных текстов на английский язык.

Тема и результаты проведенного исследования соответствуют паспорту специальности 1.2.2. – «Математическое моделирование, численные методы и комплексы программ». Представленная диссертационная работа соответствует требованиям, установленным «Положением о присуждении ученых степеней», утвержденным постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, а ее автор, Животова Алена Анатольевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.2. – «Математическое моделирование, численные методы и комплексы программ».

Официальный оппонент:

Ведущий научный сотрудник ФГБНУ «Федеральный исследовательский центр информационных и вычислительных технологий»,
д-р техн. наук, канд. филол. наук,
Кожемякина Ольга Юрьевна



Дата: 19.09.2024

Почтовый адрес: 630090, г. Новосибирск, пр-т Академика Лаврентьева, д. 6
Тел: +79139462280
Email: olga.yu.gavenko@mail.ru

Я, Кожемякина Ольга Юрьевна, настоящим даю согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета Д 24.1.478.02 и их обработку.

О. Ю. Кожемякина О. Ю.

Личную подпись официального оппонента Кожемякиной О. Ю. заверяю

Ученый секретарь ФИЦ ИВТ



[Handwritten signature]

Н. В. Киланова