УДК 004.912

АВТОМАТИЧЕСКАЯ ПРОСОДИЧЕСКАЯ РАЗМЕТКА АНГЛИЙСКОЙ УСТНОЙ РЕЧИ НА ОСНОВЕ АКУСТИЧЕСКИХ И ЛИНГВИСТИЧЕСКИХ ПРИЗНАКОВ

Вавилов М. А.[1], Оганов А. О.[1]
Научный руководитель – канд. пед. наук, доцент Толстых О. М.[1]
[1]НИТУ МИСИС
m2509820@edu.misis.ru

Введение

В последние десятилетия развитие технологий автоматической обработки речи существенно расширило возможности анализа устных корпусов и создания интеллектуальных языковых систем. Одной из задач стала просодическая разметка, позволяющая автоматически выделять интонационные и акцентные характеристики речи. Это направление особенно важно для лингвистов, поскольку ручная разметка остаётся трудоёмкой, дорогостоящей и плохо масштабируемой. Однако большинство существующих решений опирается либо на акустические признаки, либо на текстовую информацию, что ограничивает точность таких методов. Цель данной работы заключается в анализе современных технологий просодической разметки, основанные на акустических и лингвистических признаках. Анализ существующих решений позволит определить модели, наиболее подходящие для последующей интеграции в гибридную систему автоматической разметки английского устного корпуса.

Основная часть

Просодическая разметка английской речи предполагает выявление таких просодических структур как границы интонационных групп (синтагм), фразового и ядерного ударений. В лингвистических исследованиях такая разметка используется в экспериментальной фонетике, корпусной лингвистике, методике обучения фонетике.

Однако, существуют трудности в реализации автоматической разметки. Глобальная – выбор алгоритмов для автоматизации. Необучаемые алгоритмы не могут быть использованы, в связи с непрерывным ростом объема данных [1]. Локальная же трудность заключается в существовании множества видов классификаций интонации, на которую можно поделить человеческую речь [2].

С учетом указанных ограничений целесообразно использовать алгоритмы машинного обучения. Это позволит не только обрабатывать большие данные, но и адаптировать саму модель автоматизации ко входным данным. Предпочтение стоит отдавать эффективным методам, так как данные придется делить на много интонационных классов. В данной работе рассматриваются ансамблевые методы, которые впервые были предложены Джоном Тьюки [3].

Для обучения модели просодической классификации из аудиофайлов извлекаются акустические признаки, такие как: интенсивность, тембр и спектр, темпоральные признаки, гармонические признаки и прочие. Затем эти данные подаются на вход ансамблевым методам. Основополагающий принцип таких алгоритмов – мудрость толпы [4], то есть использование совокупности нескольких более простых методов. Выделяют три основные идеи построения: стекинг, бэггинг и бустинг. Основная идея бустинга, который был впервые предложен Робертом Шапире [5], заключатся в последовательном обучении однородных моделей, причем последующая модель должна исправлять ошибки предыдущей. Бэггинг, созданный Б.Эфроном [6], подразумевает параллельное обучение независимых моделей, результат работы которых усредняется. Стекинг же, разработанный Дэвидом Вольпертом [7], рассматривает разнородные отдельно взятые модели, которые подаются на вход метамодели, то есть более сложной, обобщающей.

Кроме того, в рамках работы был проведён сравнительный анализ моделей

автоматической разметки частей речи, которые впоследствии будут использоваться для создания предварительной просодической разметки. В исследовании рассмотрены несколько популярных инструментов: Flair, Stanza, UDPipe, две модели spaCy. Дополнительно были оценены трансформерные модели из библиотеки HuggingFace, ориентированные на задачу POS-тэггинга. Рассматриваемые решения различаются по архитектуре, скорости работы, качеству разметки и устойчивости к ошибкам автоматического распознавания речи.

Для оценки точности определения частей речи использовались размеченные корпуса английского текста. Тестовые данные включали UD English EWT и корпус Treebank из библиотеки NTLK. В ходе эксперимента измерялись следующие показатели: точность разметки (Accuracy) и время обработки данных каждой модели. Также, для каждой модели был проведен анализ наиболее частых ошибок, что позволило выявить их ограничения. Полученные результаты показали, что наибольшую точность продемонстрировали модели Flair и Stanza, тогда как самую высокую скорость обработки показали UDPipe и spaCy. При этом стоит отметить, что даже самые точные модели в отдельных случаях могут по-разному определять части речи, это зависит от конкретного примера, структуры предложения и контекста. Такой анализ подчёркивает важность рассмотрения не только характеристик текста и речи по отдельности, но и их совместного использования.

Выводы

В результате проведенного исследования были отобраны основные методы машинного обучения для задачи просодической классификации и различные акустические признаки, как входные данные для обучения модели. Также было установлено, что наибольшую точность разметки демонстрируют модели Flair и Stanza, а наивысшую скорость UDpipe и spaCy. Эти результаты позволяют выбрать модель с сильными сторонами для дальнейшей интеграции в гибридную систему автоматической просодической разметки.

Перспектива дальнейших исследований видится в изучении методов извлечения аудиопризнаков, в оптимальном отборе признаков, и в последующем переходе к нейросетевым алгоритмам.

Эффективность вышеупомянутых методов следует оценивать на практике, так как существует множество факторов, влияющих на нее, таких как: методы извлечения аудиопризнаков, вид просодической классификации и многие другие.

Литература

1. Интеллектуальные голосовые помощники: современное состояние и перспективы развития: доклад // Л.Б. Нарусова, федеральный-справочник.рф [Электронный ресурс]. – URL: https://федеральный-справочник.рф/files/SVAYZ/saderzhanie/Tom%207/I/Narusova.pdf (дата обращения: 09.02.2026).
2. Pitch and Scales // Smotri Uchis [Электронный ресурс]. – URL: https://smotriuchis.ru/blog/pitch-and-scales (дата обращения: 09.02.2026).
3. Tukey J. W. Exploratory Data Analysis. – Reading, MA: Addison-Wesley, 1977. – 688 p.
4. Мудрость толпы // Cyclowiki [Электронный ресурс]. – URL: https://cyclowiki.org/wiki/%D0%9C%D1%83%D0%B4%D1%80%D0%BE%D1%81%D1%82%D1%8C_%D1%82%D0%BE%D0%BB%D0%BF%D1%8B (дата обращения: 09.02.2026).
5. Schapire R.E. The Strength of Weak Learnability // Proceedings of the Second Annual Workshop on Computational Learning Theory (COLT). 1990. P. 197–227. – URL: https://dl.acm.org/doi/10.5555/100230.100234 (дата обращения: 09.02.2026).
6. Efron B. Bootstrap Methods: Another Look at the Jackknife // The Annals of Statistics. 1979. Vol. 7, № 1. P. 1–26.
7. Wolpert, D.H.: Stacked generalization. Neural Networks 5(2), 241-259 (1992) https://doi.org/10.1016/S0893-6080(05)80023-1