

ОПРЕДЕЛЕНИЕ ТОНАЛЬНОСТИ ТЕКСТА МЕТОДОМ ОБУЧЕНИЯ С УЧИТЕЛЕМ

А.И. Гербик, Е.А. Макович, М.В. Аксамит

Целью анализа тональности является нахождение мнений в тексте и определение позиции автора относительно упомянутой темы. При наличии выборки размеченных данных данную задачу можно рассматривать как задачу классификации, эффективными методами решения которой являются методы машинного обучения с учителем.

Для того, чтобы методы решения задач классификации можно было применить для анализа тональности текста, необходимо текст представить в виде математического вектора. В настоящее время одной из самых распространенных в различных областях лингвистических исследований является векторная модель «мешок слов». Текст в данной векторной модели рассматривается как неупорядоченное множество слов. Вектор, являющийся модельным представлением текста в векторном пространстве, образуется упорядочением весов всех слов (включая те, которых нет в конкретном тексте). Размерность этого вектора равна количеству различных слов во всей коллекции, и является одинаковой для всех текстов коллекции.

Полученную задачу классификации можно решить различными методами машинного обучения: наивный байесовский классификатор, логистическая регрессия, метод опорных векторов, методы нейронных сетей и т.д. Сравнив их временную сложность, качество полученных моделей, масштабируемость можно выбрать наиболее подходящий для конкретных данных и конкретной задачи.

Литература

1. Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2002. P. 79–86.

2. Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of WWW, pages 519–528, 2003.

ИССЛЕДОВАНИЕ РАЗБОРЧИВОСТИ РЕЧИ С ИСПОЛЬЗОВАНИЕМ СВЯЗНЫХ ТЕКСТОВ НА АРАБСКОМ ЯЗЫКЕ

С.М. Горошко, С.Н. Петров

Исследования разборчивости речи (Н.Б. Покровский, Ю.С. Быков, М.А. Сапожков) начались в середине прошлого века в связи с возникшей необходимостью оценки качества средств связи. Факторов, ухудшающих передачу речи, может быть несколько: фоновый шум, длинные поздние отражения и эхо. В результате были получены базовые зависимости между различными видами разборчивости: слоговой, словесной, фразовой. Методы измерения разборчивости речи можно разделить на: субъективные (чисто субъективный метод; объективизированный; тональный); объективные (формантные, AI (индекс артикуляции), SII (индекс разборчивости речи); модуляционные: (STI (Speech transmission index – индекс передачи речи), RASTI (быстрый STI), STIPA (STI для систем звукоусиления), STITEL (STI для телекоммуникационных систем). Однако в реальности мы имеем дело только со связными текстами. Существующая методика оценки защищенности речевой информации от утечки по техническим (акустическим) каналам нуждается в корректировке с учетом специфики задач защиты информации и, прежде всего, в определении методических и косвенных погрешностей.

Целью работы является усовершенствование и общепринятой инструментальной методики оценки разборчивости речи на основе модернизации зависимостей коэффициентов восприятия и словесной разборчивости от формантной.

Испытания проводились с использованием связных текстов в произношении дикторов, свободно владеющих арабским языком. Необходимо установить различия методов при оценке разборчивости связной речи и отдельных фраз, слов, а также выявить факторы, существенно усложняющие разборчивость речи на арабском языке. Целесообразно сформировать речеподобный и формантный шум для оценки эффективности защищенности речевой информации.