

По сравнению с другими методами вейвлет-очистка ИС от шумов имеет ряд преимуществ, заключающихся в первую очередь в оптимальной частотно-временной локализации, что позволяет обнаружить компоненты шума вблизи некоторой точки и, не искажая при этом самого сигнала, удалить их, тем самым очистив полезный сигнал от шума. Математическое обоснование метода очистки речевого сигнала от шумов с использованием комплексного вейвлета Морле (КВМ) и сравнение его с другими методами нами было рассмотрено в работе [0]. Применение КВМ для анализа тонкой структуры ИС, принятого на выходе КУИ, позволило выявить наличие разномасштабной периодичности, содержащейся в анализируемых зависимостях, а также выявить появившиеся частотные составляющие, не соответствующие его собственным частотам. Это открыло реальную возможность осуществить дополнительную очистку ИС от шумов и, тем самым, существенно повысило точность оценки защищенности речевой информации в КУИ.

Исследования показали, что вейвлет-анализ целесообразно проводить при определенно заданных отношениях мощности ИС к мощности шума.

### **Литература**

1. Бураченок, И. Б. Анализ вейвлет-преобразованием тонкой структуры гласных звуков речевого сигнала / И. Б. Бураченок, В. К. Железняк // Теоретические и прикладные аспекты информационной безопасности : материалы междунар. науч.-практ. конф., Минск, 19 июня 2014 г. / УО «Акад. М-ва внутр. дел Респ. Беларусь»; редкол.: В. Б. Шабанов (отв. ред.) [и др.]. – Минск, 2015. – С. 124–128.

## **ОПРЕДЕЛЕНИЕ ТОНАЛЬНОСТИ ТЕКСТА МЕТОДОМ ОБУЧЕНИЯ БЕЗ УЧИТЕЛЯ**

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

Для определения тональности текста в настоящее время активно применяются методы машинного обучения с учителем, однако для их применения необходимо наличие обучающей выборки. Сбор данных для обучающей выборки может быть либо невозможен, либо неоправданно затратен. В таких случаях целесообразно применять методы обучения без учителя.

Данный метод определения тональности состоит из трех шагов. На первом шаге из текста выделяются фразы, которые могут выражать мнение автора. Для этого для всех слов текста определяются их часть речи, затем в тексте осуществляется поиск фраз, которые соответствуют одному из синтаксических шаблонов, часто используемых для выражения мнения (например, существительное + прилагательное либо глагол + наречие). На втором шаге определяется тональность фраз, выделенных на первом шаге. Для этого из поточечной взаимной информации фразы и слова «хороший» вычитается поточечная взаимная информация этой же фразы и слова «плохой». Поточечная взаимная информация – логарифм отношения вероятности появления двух терминов вместе к вероятности появления этих терминов независимо друг от друга. В качестве оценки вероятности используется количество результатов поисковой системы на соответствующий запрос. В результате для каждой фразы будет получено значение тональности: положительное число будет означать положительную тональность, отрицательное – негативную. На третьем шаге определяется тональность текста путем нахождения среднего значения тональностей фраз, входящих в данный текст.

Подходы, основанные на обучении без учителя, хоть и показывают меньшую точность в работе, но получаемые модели являются доменно-независимыми и верно определяют тональность отдельных слов, значение которых зависит от контекста.

### **Литература**

1. Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.  
2. 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.