

## **МИНОБРНАУКИ РОССИИ**

Федеральный исследовательский центр «Информатика и управление» РАН  
Национальный комитет при президиуме РАН по распознаванию образов  
и анализу изображений

Институт информационных технологий Национальной  
академии наук Азербайджана

Институт проблем передачи информации им. А.А. Харкевича РАН  
Издательство «Наука и технологии»

Национальный исследовательский Томский государственный университет  
Федеральный исследовательский центр «Карельский научный центр РАН»  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Юго-Западный государственный университет»

## **ОПТИКО-ЭЛЕКТРОННЫЕ ПРИБОРЫ И УСТРОЙСТВА В СИСТЕМАХ РАСПОЗНАВАНИЯ ОБРАЗОВ И ОБРАБОТКИ ИЗОБРАЖЕНИЙ**

**Распознавание – 2021**

Сборник материалов XVI Международной  
научно-технической конференции

14–17 сентября 2021 года

Редакционная коллегия:

С. Г. Емельянов, В. С. Титов (отв. ред.),

Т. А. Ширабакина, Э. И. Ватутин,

В. С. Панищев

Курск 2021

<i>Волков Д.А.</i> Метод обработки и распознавания изображения оптико-электронным вычислительным устройством .....	81
<i>Волков Е.О.</i> Разработка системы распознавания вязкой составляющей на изломах металла.....	83
<i>Гаджирагимова М.Ш., Исмаилова М.И.</i> Общий подход к анализу тональности текстов.....	85
<i>Гвоздева С.Н., Ватутин Э.И.</i> Оценка вероятности досрочного прерывания процесса умножения бинарных матриц в зависимости от их размера и плотности.....	87
<i>Гершкович М.М., Бирюкова Т.К., Синицин В.И.</i> О возможностях распознавания информационных объектов с неточно известными атрибутами и биометрическими характеристиками в информационных системах, получающих данные из разнородных источников .....	90
<i>Голубов Д.А., Гвоздева С.Н.</i> Идентификация n-элементных двухполюсников на основе обобщенных параметров .....	92
<i>Горбачев С.В., Белозеров А.П., Бхаттачарья С.</i> Объединение решений с помощью мета-классификатора в ансамблевой модели коллективного распознавания .....	95
<i>Губарь М.Н., Калгин К.В.</i> Оптимизация параметров оптико-электронной системы для осуществления маловысотного полета.....	97
<i>Гурбанова К.Ш.</i> Исследование этапов и методов моделирования для распознавания жестов .....	99
<i>Дегтярев С. В., Селихова К. Н.</i> Цифровая антропометрия .....	101
<i>Дегтярев С.В., Криушин Е.А., Иванова Е.Н.</i> Аппаратный ускоритель обработки изображения для систем распознавания образов .....	103
<i>Добрица В.П., Плугатарев А.В., Золотухин А.В.</i> Технологии выявления и скрытия конфиденциальных данных в составе графических объектов .....	105
<i>Добрица В.П., Волокитина Т.С.</i> Обнаружение использования стеганографии в изображениях при помощи нейросетей .....	107
<i>Дубинин В.Н., Дубинин А.В., Климкина Л.П.</i> Представление и преобразование онтологических описаний киберфизических систем с использованием раскрашенных сетей Петри.....	109
<i>Егошина И.Л.</i> Комплексирование разноспектральных изображений на основе вейвлет-преобразования.....	111
<i>Ефанов С.В., Иванова Е.Н.</i> Система определения и извлечения данных из различных типов файлов ....	114

Объём обучающей выборки составил 174 фотографии, на которых находились по два излома, имеющих различные показатели хрупкости. На ней была обучена нейронная сеть U-NET [6] на сформированной выборке для решения задачи поиска излома металла на изображении. Точность построенной модели составляет 95%. Была обучена нейронная сеть Mask-RCNN [5], которая решала задачи поиска зоны контроля, на которой нужно вычислять хрупкую составляющую, а также на определение зон 100% хрупкости и 50%, когда чередуются зоны хрупкости-вязкости. Точность обученной модели составила 90%. В результате исследований они были интегрированы в программное обеспечение.

### СПИСОК ЛИТЕРАТУРЫ

1. Федотов А. Физическое материаловедение. Ч. 2: Фазовые превращения в металлах и сплавах. Минск: Вышэйшая школа, 2012. 448 с.
2. ГОСТ 30456-97 Прокат листовой и трубы стальные. Методы испытания на ударный изгиб; введ. 01.01.2000. М. : Изд-во стандартов, 2003. 11 с.
3. API RP 5L3. Drop-Weight Tear Tests on Line Pipe. American Petroleum Institute, 2014. 16 с.
4. Shapiro Linda G., Stockman George C. Computer Vision. New Jersey, Prentice-Hal, 2001. 609 p.
5. He Kaiming, Gkioxari Georgia, Dollár Piotr, Girshick Ross. Mask R-CNN. 2017. arXiv 1703.06870.
6. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. arXiv:1505.04597.

УДК 004.89

**М. Ш. Гаджирагимова<sup>1</sup>, М. И. Исмаилова<sup>1</sup>**  
e-mail: hmakrufa@gmail.com, imarziya@gmail.com

<sup>1</sup> Институт информационных технологий НАНА, Баку, Азербайджан

### ОБЩИЙ ПОДХОД К АНАЛИЗУ ТОНАЛЬНОСТИ ТЕКСТОВ

*Рассматривается архитектура классификации тональности коротких текстов, основанной на методах машинного обучения.*

Анализ тональности (Sentiment Analysis – SA) можно определить как процесс, который автоматизирует анализ отношений, мнений, взглядов и эмоций из текста, речи, твитов и источников баз данных с помощью обработки естественного языка, интеллектуального анализа текста и компьютерной лингвистики. Анализ тональности включает в себя классификацию мнений в тексте на такие категории, как «положительные», «отрицательные» или «нейтральные» [1–3].

С быстрым распространением социальных сетей сбор и анализ реакции людей по различным темам на покупку продукта, социальных и общественных услуг и т.д. имеют жизненно важное значение. Чтобы классифицировать короткие тексты в социальных сетях (Facebook, Twitter и т.д.) по определенному тону, используются методы анализа текста. Анализ тональности имеет несколько трудностей. Традиционным программным инструментом интеллектуального анализа данных сложно управлять из-за увеличения их величины в петабайтах.

Как указано на рисунке, основные задачи SA включают: сбор данных (набор данных); предварительную обработку данных; выбор/извлечение признаков; алгоритмическую обработку классификации настроений; представление результатов, то есть представление полярности мнений.



Рис. Архитектура SA

**Набор данных.** Сбор данных в SA – это первый шаг, в частности, собирается объем данных для анализа. Поскольку Twitter в основном используется широкой аудиторией для обмена различными мнениями, набор данных собирается из Twitter с помощью Twitter API. Подход машинного обучения, применимый к анализу настроений, в основном относится к контролируемой классификации, поэтому требуются два набора данных: учебный набор; тестовый набор.

**Предварительная обработка данных.** Этот шаг можно отметить как наиболее важный этап полного процесса SA. Для извлечения полезных мнений из необработанного набора данных предлагаемая система должна сначала отфильтровать необработанный набор данных, чтобы удалить бесполезную информацию. В системе предусматривается выполнение следующих методов предварительной обработки набора данных: удаление стоп-слов; удаление цифр, специальных символов и URL-адресов; токенизация; стемминг.

**Извлечение признаков.** Из предварительно обработанных данных предлагаемая система идентифицирует и извлекает признаки множеств. Признак описывается как атрибут объекта. Bag-of-words – традиционный метод извлечения элементов текста, которые используются для классификации. Он создает векторное представление о каждом документе на основе словаря всего текстового набора и различных методов подсчета очков (например, Term Frequency – Inverse term frequency –  $tf \cdot idf$ ). Вектор признаков преобразует слова в числовые значения  $.tf$ , измеряет частоту появления термина  $t$  в документе, а  $idf$  – обратная частота документов, показывает важность терминов. Их формула:

$$tf(t) = \text{No. of times term } t \text{ occurs in a document} / \text{Total no. of terms in the document}$$

$$idf(t) = \log_e(\text{Total no. of documents} / \text{No. of document with term } t \text{ in it})$$

**Классификация и тестирование.** Последний модуль предлагаемого подхода, в котором сгенерированный вектор признаков выступает в качестве входных данных для классификаторов контролируемого машинного обучения, таких как SVM, NaiveBayes, MaxEntropy, Decision Tree, XGBoost, Random Forest. На этом этапе обучаемся классификаторам на сгенерированном векторе признаков из предыдущего этапа и делаем выводы на тестовом наборе.

Мы представили подход анализа настроений и обнаружили чувства людей в форме полярности (негативный и позитивный). В будущих исследованиях планируется применить классификаторы, основанные на углубленном обучении (LSTM, CNN, RNN и т.д.), для анализа настроений и построить ансамбль классификаторов для рассматриваемой задачи.

## СПИСОК ЛИТЕРАТУРЫ

1. Liu B. Sentiment analysis: mining opinions, sentiments, and emotions. The Cambridge University Press, 2015.
2. Ravi K., Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications // Knowledge-Based Systems. 2015. Vol. 89. P. 14-46.
3. Twitter API. URL: <https://developer.twitter.com/en/docs/tweets/search/overview>.

УДК 621.397.01

**С. Н. Гвоздева<sup>1</sup>, Э. И. Ватутин<sup>1</sup>**  
e-mail: svetka-gvozdeva@yandex.ru

<sup>1</sup> ФГБОУ ВО «Юго-Западный государственный университет», Курск, Россия

## ОЦЕНКА ВЕРОЯТНОСТИ ДОСРОЧНОГО ПРЕРЫВАНИЯ ПРОЦЕССА УМНОЖЕНИЯ БИНАРНЫХ МАТРИЦ В ЗАВИСИМОСТИ ОТ ИХ РАЗМЕРА И ПЛОТНОСТИ

*Представлены зависимости вероятности досрочного прекращения умножения бинарных векторов от размера  $N$  умножаемых матриц и их плотности  $d$ .*

Задача умножения бинарных матриц возникает в ряде частной подзадачи в ряде фундаментальных и прикладных задач (классификация бинарных отношений в граф-схемах параллельных алгоритмов [1], транзитивное замы-