

УДК 004.056

А.А. Воробьева, М.Ю. Федосенко

Методы интеллектуального анализа данных и обработки естественного языка в управлении роботизированными производственными системами

Представлен сравнительный анализ методов обработки естественного языка и интеллектуального анализа данных, используемых при обработке данных в управлении роботизированными производственными системами. Разработана концепция реализации процесса интеграции методов интеллектуального анализа данных и обработки естественного языка в роботизированные производственные системы, описаны необходимые компоненты. Основная фокусировка сделана на использование методов обработки естественного языка (natural language processing). Материал работы имеет ценность для проведения экспериментальных исследований в области применения методов обработки естественного языка и искусственного интеллекта в роботизированных производственных системах.

Ключевые слова: умные фабрики, машинное обучение, большие данные, роботизированные производственные системы, обработка естественного языка, text mining, natural language processing, кибербезопасность.

DOI: 10.21293/1818-0442-2023-26-3-65-71

Современный уровень развития технологий искусственного интеллекта (ИИ) позволяет автоматизировать решение множества задач, таких как выявление закономерностей, принятие решений, цифровая имитация поведенческих характеристик. Однако их качественное решение требует применения эффективных алгоритмов обработки разнородных данных, полученных с различных производственных систем. Такие данные могут быть представлены в различных форматах, на различных языках, а также могут быть неструктурированными. Эти данные впоследствии используются для обучения моделей машинного обучения (МО), на основе которых строятся интеллектуальные системы управления.

Автоматизация ряда человеческих задач является необходимым условием для практической реализации современных робототехнических систем [1]. По данному направлению сегодня в научном сообществе ведется активная работа, большое число публикаций посвящено вопросам обработки данных с использованием методов ИИ и МО.

В данной работе выполнено сравнение существующих решений и научных публикаций в области обработки неструктурированных и слабоструктурированных данных, текстов на естественном языке (ЕЯ), методов обработки ЕЯ (англ. natural language processing, NLP). Представлены результаты анализа перспектив применения интеллектуального анализа текстов на ЕЯ для управления роботизированными системами производства.

Анализ основных направлений исследований в области роботизированных производственных систем

Сегодня одним из главных направлений развития робототехнических систем является создание технологий для умных фабрик (англ. smart factory). На таких фабриках методы ИИ и МО используются для анализа данных, управления автоматизированными процессами и самообучения. Здесь производ-

ственные системы действуют в значительной степени автономно, используя данные, поступающие в режиме реального времени.

Современная концепция «Индустрия 4.0» направлена именно на поддержку сбалансированного и/или симбиотического взаимодействия между людьми и машинами в производственной среде.

Предусмотрено использование киберфизических систем (КФС) и передовых технологий взаимодействия «человек–машина». При этом акцент делается именно на симбиозе человека и КФС, что должно обеспечивать синергетический эффект от взаимодействия. Концепция не подразумевает замены человека роботами и интеллектуальными системами. Основная её цель – обеспечить повышение эффективности и результативности работы операторов с помощью новейших технологий.

В научной работе [2] представлена концепция применения подхода интеграции знаний (англ. knowledge integration) на умной фабрике, где производственные системы могут автономно адаптироваться к событиям и действовать с учетом изменений в производстве, цепочке поставок и требований клиентов. Рассмотрена возможность интеграции знаний о предыдущем опыте в различные виды взаимодействия умной фабрики, такие как «человек–человек», «человек–машина» и «машина–машина».

Интеграция знаний рассмотрена с организационной, кадровой и технологической точек зрения как перспективный подход к управлению знаниями в «Индустрии 4.0». Основной идеей является преобразование качественных показателей знаний в количественные, что позволяет выполнять их обработку в системах ИИ.

Авторы предлагают использование цифровых двойников для объединения данных и имитационных моделей, что позволяет выполнять моделирование и прогнозирование в режиме реального времени. Данный подход обеспечивает возможность раннего обна-

ружения потенциальных проблем, а также более эффективное и надежное функционирование производственных систем, обеспечивая при этом высокую гибкость и адаптивность.

В статье [3] авторами представлено исследование в области использования так называемых софт-ботов (англ. *softbots*) в интеллектуальных производственных системах. Софт-бот представляет собой интеллектуального персонального помощника, представляющего оператору возможность взаимодействовать с интеллектуальным оборудованием, роботами, компьютерами и другими информационными системами, как человек с человеком. В работе предложена эталонная структура, которая позволяет создавать софт-ботов для заданных условий. Авторы отмечают, что разработанное решение является перспективным, однако имеет ряд ограничений, связанных с безопасностью, семантической интероперабельностью, обработкой голоса и распознаванием ЕЯ, адаптивностью.

В работе [4] выполнен систематический обзор литературы для определения перспективных направлений исследований и критических областей развития в области технологий для умных фабрик.

Авторами определено, что исследования могут быть сгруппированы в пять кластеров: радиочастотная идентификация (англ. *RFID*) для управления производством в режиме реального времени (отслеживание объектов, сбор производственных данных); беспроводные технологии (*RFID*, сенсорные сети и пр.) для функционирования, контроля и планирования процессов производства в режиме реального времени; мультиагентные системы, онтологии для контроля и планирования процессов, ускорения принятия решений; облачные вычисления; методы оптимизации, планирования и моделирования для снижения издержек производства.

Анализ представленных выше работ показал, что качественная интеграция перечисленных подходов способна вывести умные фабрики и роботизированные системы на новый уровень, минимизировав участие оператора. Эффективное использование приведенных технологий может быть обеспечено только в случае применения корректных процедур управления и обмена данными по всей цепочке поставок.

Такого рода обмен, в свою очередь, порождает риски информационной безопасности. Вопросы обеспечения безопасности на умных фабриках в настоящий момент остаются открытыми, доступно лишь ограниченное число публикаций по данной тематике.

Большинство проанализированных работ требуют интеграции знаний в умные фабрики и выделяют необходимость наличия релевантных данных, прошедших качественную предобработку. Выделены следующие перспективные методы работы с данными: облачные, туманные, граничные вычисления; описательная, предиктивная, предписывающая аналитика; графы знаний; анализ и генерация текста, обработка полуструктурированных и неструктурированных данных; использование цифровых двойников.

Методы предобработки данных для управления техническими изменениями в роботизированных производственных системах

Вопросы управления техническими изменениями (ТИ) являются особенно актуальными при производстве сложной продукции и изделий [5].

Работа [6] описывает применение интеллектуальных технологий для обработки информации прошлых ТИ, с акцентом на её важности для оптимизации процессов разработки текущих изменений и принятия решений. Авторы применяют метод обнаружения знаний в базах данных (*knowledge discovery in databases, KDD*) с целью оценить возможность его использования для решения задачи управления техническими изменениями.

В работе приведено описание эксперимента по анализу данных производства автомобилей, включающих более 50 тыс. предыдущих инженерных изменений за пять лет [7].

Также выполнено сравнение достоинств и недостатков более ранних подходов [8–11]. Ключевыми составляющими анализа текстовых данных является использование правил ассоциации, *KDD* и применение методологии по исследованию данных (англ. *Cross-industry standard process for data mining*). Полученные результаты формулируют необходимость оптимизации временных и вычислительных ресурсов путём сокращения количества атрибутов данных.

Подобные задачи также исследуются в работе [12], которая представляет объемное описание применения интеллектуальных методов анализа данных ТИ. Она содержит в себе исследование возможностей использования цепей Маркова для анализа и сравнения запросов на ТИ в промышленных комплексах, смоделированных при помощи матрицы структуры проектирования.

Эксперименты по обработке текстов запросов на ТИ проведены на данных проекта по разработке грузовых автомобилей компании *Volvo Group Trucks Technology*. Основной акцент работы сделан на анализе исторических данных доработок из базы данных с целью выявления закономерностей в аналитических моделях проектирования. Основные этапы процесса включают: выборку, подготовку данных, трансформацию, поиск ассоциативных связей, интерпретацию результатов. Схематически они представлены на рис. 1.

Методы обработки естественного языка в управлении техническими изменениями

Запросы на технические изменения представляют собой письменные тексты на ЕЯ, которыми обмениваются операторы процесса. Ключевым этапом процесса обнаружения знаний является применение технологий анализа данных и текста (англ. *data и text mining*) к документам на ТИ.

В работе [13] авторы описывают последовательное применение технологий обработки ЕЯ к данным ТИ проектов. Подход позволяет сократить время простоя из-за сбоев. Полученные формализованные знания также могут быть использованы в качестве основы для оптимизации производственных процессов.

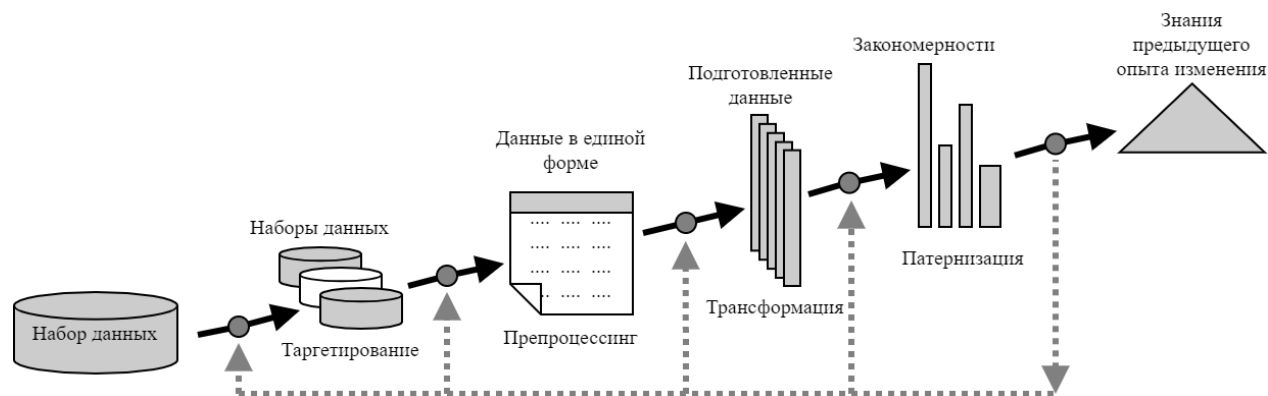


Рис. 1. Основные этапы процесса обнаружения знаний в text mining [12]

Конвейер обработки неструктурированных текстовых данных, поступающих от оборудования и операторов, включает: очистку и подготовку данных, токенизацию (разделение текста на токены), удаление стоп-слов, лемматизацию (приведение слов к нормальной морфологической форме), векторизацию (конвертация текста в числа) и статистический анализ (выделение математических закономерностей для последующего интеллектуального анализа). Для прогнозирования времени простоя оборудования на основе текстовых комментариев операторов станков в сочетании с данными оборудования производственных линий применяются алгоритмы XGBoost и LigthGBM в сочетании со следующими методами векторизации: Bag-of-Words, TF-IDF, Word2Vec. Наиболее высокая точность достигается с использованием LigthGBM и TF-IDF.

В работе [14] исследуется применение методов кластеризации текстов документов на ТИ, написанных на итальянском языке. Анализируемые тексты включали описания причин ТИ.

Для поиска связей между текстами документов на ТИ использовались нейронные сети, а именно алгоритм самоорганизующихся карт Кохонена (англ. SOM) [15]. На этапе предобработки предлагается использовать метод TF-IDF. Количество терминов, связанных с каждым документом в наборе данных после фильтрации и стемминга, было равно 361. Далее выполнялись кластеризация и визуализация посредством использования U-матриц. Результаты показали, что использование нейронных сетей SOM при поиске закономерностей и обработке текстов документов на ТИ является эффективным инструментом для улучшения процесса анализа ТИ. Также он позволяет решить проблему неконтролируемой кластеризации.

В рамках данного исследования проведен анализ достоинств и недостатков применения SOM в text mining, представленных в табл. 1.

Результаты анализа методов обработки ЕЯ в управлении техническими изменениями показали, что данные методы используются для бинарной (в случае с автоматизацией процесса принятия решения) и многоклассовой классификации (в случае определения категории технической доработки в автоматизированной системе). Помимо этого, применение методов NLP решает задачи анализа тональности

текста, выделения именованных сущностей (например, наименований и имен), извлечение из текста фактов и взаимоотношений между словами (отношение конкретных терминов к конкретным задачам автоматизации процессов).

Таблица 1

Характеристика самоорганизующихся карт Кохонена

№	Достоинства	Недостатки
1	Высокая скорость обучения	Работа только с вещественными числовыми векторами
2	Устойчивость к зашумленным данным	Требует явного указания количества кластеров
3	Визуализация многомерных данных	Отсутствие чётких границ кластеров, объективно существующие кластеры могут разрываться на фрагменты
4	Снижение размерности исходного пространства, проецирование многомерного пространства в пространство с более низкой размерностью	Сильная зависимость от инициализации (начального распределения весов)

Одной из наиболее важных задач NLP в рамках обеспечения эффективного производства и интеллектуального управления роботизированными системами производства является задача суммаризации-минимизации большого объёма текста с сохранением ключевой сути изложенной информации. Применение данного подхода позволяет анализировать большие объёмы технической документации и выделять из неё конкретный опыт, необходимый для улучшения систем.

На рис. 2 представлены подходы и приведена классификация задач обработки ЕЯ [16].

Интеллектуальные методы обеспечения функциональной и информационной безопасности в производственных системах

Помимо задач по организации интеллектуальных систем управления, роботизации и автоматизации производственных процессов, а также использования методов МО и обработки ЕЯ, важной задачей является обеспечение функциональной и информационной безопасности роботизированных систем.

В работе [17] изложены основные подходы к обеспечению информационной и функциональной безопасности микросетей, которые обладают потенциалом для автономной и отказоустойчивой работы в процессе их электрообеспечения. Автор приводит

описание наиболее значимых атак на КФС за последние 10 лет, делает выводы, что целью злоумышленников являются не сами микросети, а именно системы управления.



Рис. 2. Классификация задач обработки естественного языка [16]

Применение умных сетей электроснабжения (англ. smart grid) в настоящее время затруднено, так как вопросы обеспечения безопасности остаются нерешенными, существуют лишь отдельные частные решения. В качестве решения предлагается прежде всего отталкиваться от архитектуры сети и структуры системы управления, а также присущих ей уязвимостей, поскольку на данный момент обеспечение кибербезопасности не является встроенной функцией в микросетях.

В качестве универсальных методов защиты автор предлагает непрерывный мониторинг состояния, проактивные методы защиты, методы кодирования, использование интеллектуальных счётчиков состояния системы (IoT). Кроме того, авторы отмечают технологию блокчейн как перспективную для обеспечения безопасности киберфизических систем за счёт встроенных алгоритмов хеширования, смарт-контрактов и распределённого характера хранения и обработки данных [18].

Отметим, что решение задачи по обеспечению кибербезопасности в интеллектуальных производственных системах прежде всего заключается в использовании криптографических алгоритмов (в том числе легковесных) и интеллектуальных методов обнаружения и предотвращения классических атак с учётом имеющихся знаний об уязвимостях системы. Основным методом защиты от потери данных и обеспечения отказоустойчивости является резервирование,

на реализацию которого необходимо заложить определённое количество вычислительных мощностей на этапе проектирования. В целом в роботизированных производственных системах должна применяться концепция «secure by design», которая заключается в обеспечении безопасности на всех этапах производственного процесса – от этапа проектирования до финальной фазы поставки.

Стоит подчеркнуть, что в большинстве случаев атаки нацелены на системы управления и принятия решений, где часто используются технологии ИИ, что ведёт к необходимости также учитывать и атаки, специфичные для систем ИИ: состязательные атаки, отравляющие атаки, атаки в обход, атаки уклонения и др. [19]. Исследования в данном направлении только начинаются, и в настоящий момент данный вид атак может быть отнесен к наиболее опасным, так как не существует общепринятых методов противодействия.

Необходимые технологии в виде инструментов и алгоритмов МО, необходимых для решения задач разработки систем, представлены в табл. 2.

На рис. 3 представлена сравнительная характеристика системы в контексте моделирования на VAD-диаграмме. Данная диаграмма отражает цели, направления и подходы, необходимые при сборе требований и последующей разработке интеллектуальных роботизированных систем управления.

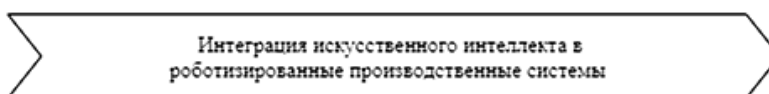
Таблица 2

Подходы, методы и инструменты для интеллектуального управления роботизированными производственными системами

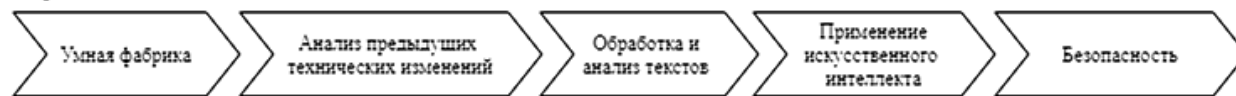
№	Подход	Инструменты
1	2	3
1	Вычисления	–
1.1	Облачные	IaaS, SaaS, PaaS, FaaS, CaaS
1.2	Туманные	OpenFog, IFCIoT
1.3	Граничные	AR, VR, SCADA, PAC, IoT
2	Аналитика	Описательная, предиктивная, предписывающая аналитика
3	Предобработка естественного языка	–
3.1	Векторизация	Подходы: TF-IDF, Bag-of-Words, Word2Vec, регулярные выражения, One-hot encoding. Библиотеки: Keras, NLTK, YouTokenToMe

1	2	3
3.2	Лематизация	Библиотеки: <i>Mystem, Stemka, Treetagger, UDPipe, Wordnet Lemmatizer, Metaphraz, Coda, Cadenza</i>
3.3	Стемминг	Подходы: анализ <i>n</i> -грамм, стемминг на основе корпуса текстов, стохастические алгоритмы, <i>аффикс-стеммеры, kipple-down rules, стеммер Портера</i>
3.4	Статистический анализ текстов	Эмбедингу, <i>Scoring, TF-IDF, Word2Vec, регулярные выражения</i>
4	Графы знаний	Подходы и библиотеки: <i>онтологии, матричная факторизация, Random Walk, DeepWalk, Node2vec, DNNGR, SDNE</i>
5	Машинное обучение	–
5.1	Классификация	Методы: <i>наивный Байесовский классификатор, деревья решений, логистическая регрессия, k-ближайших соседей, метод опорных векторов, нейронные сети</i> . Библиотеки: <i>NumPy, SciPy, Matplotlib, Seaborn, Scikit-learn, Pandas, TensorFlow, Theano, Keras, PyTorch, Matplotlib</i>
5.2	Регрессия	Методы: <i>линейная регрессия, полиномиальная регрессия</i>
5.3	Кластеризация	Методы: <i>k-средних, сдвиг среднего значения, DBSCAN, SOM</i> . Библиотеки: <i>PyCaret, PyClustering, NumPy, Pandas, Scikit-learn, Matplotlib</i>
6	Правила ассоциации	Подходы: <i>Generalized Association Rules, Quantitative Association Rules, Apriori, DHP, Partition, DIC</i>
7	Корреляция	Подходы: <i>множественная корреляция, КПК Пирсона, КПК Спирмена, ККЗ Фехнера, критерий Стьюдента</i>
8	Обнаружение знаний в базах данных	Методы и подходы: <i>деревья решений, нейронные сети, генетические алгоритмы, методы статистического анализа (дескриптивный, регрессионный, корреляционный, факторный, дисперсионный, компонентный, дискриминантный)</i>
9	Методы обработки естественного языка	Инструменты и библиотеки: <i>AlchemyAPI, Expert System S.p.A., General Architecture for Text Engineering (GATE), Modular Audio Recognition Framework, MontyLingua, Natural Language Toolkit (NLTK)</i>
10	Информационная и функциональная безопасность	–
10.1	Управление уязвимостями	Банки данных угроз и уязвимостей: <i>CWE, CVE, OWASP, ФСТЭК БДУ</i>
10.2	Автоматизация контроля кибератак (технологии противодействия)	Инструменты: <i>SIEM, SOAR, EDR, UEBA, NGFW, SOC, сканеры уязвимостей, IDS/IPS</i>
10.3	Кодирование и криптография	Код Хемминга, <i>блочное шифрование, асимметричное шифрование, хеширование, ГОСТ 28147-89 и пр.</i>
10.4	Резервирование	<i>Backup, Snapshot, PBBA target systems, PBBA integrated systems, VRRP, Ethernet-channel</i>

Цель



Направления



Подходы

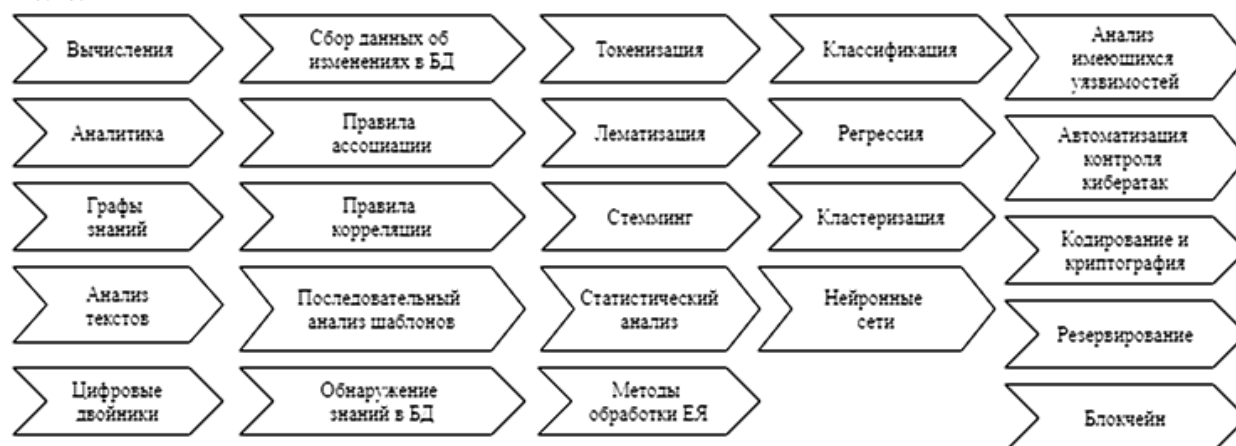


Рис. 3. VAD-диаграмма интеллектуального управления роботизированными системами производства

Заключение

В статье представлены обзор и анализ научно-технической литературы в области методов интеллектуального анализа данных и обработки естественного языка в управлении роботизированными производственными системами. Рассмотрены особенности применения данных методов на умных фабриках и при управлении техническими изменениями. Одной из основных составляющих данного процесса является анализ предыдущего опыта.

В силу большого объёма данных (в том числе слабо- и неструктурированных) анализ целесообразно выполнять при помощи методов искусственного интеллекта. Акцент сделан на методах обработки естественного языка (токенизации, лематизации, стемминга, статического анализа текстовых данных). Однако необходимо учитывать вычислительно-ресурсные возможности системы и подходы по оптимизации их использования. Наиболее применимыми подходами под задачи интеллектуального управления роботизированными системами являются облачные, туманные и граничные вычисления.

Важнейшей задачей является обеспечение кибер- и функциональной безопасности роботизированных производственных систем, а именно систем управления и принятия решений. Её решение прежде всего заключается в использовании криптографических методов, а также интеллектуальных методов противодействия атакам с учётом имеющихся знаний об уязвимостях. Основным методом защиты от потери данных и обеспечения отказоустойчивости является резервирование. При проектировании роботизированных систем и интеллектуальных методов управления рекомендуются использование предиктивной и предписывающей аналитики, а также реализация идеологии «secure by design».

Представленная в работе концепция интеграции методов интеллектуального управления в роботизированные производственные системы может служить основой для выбора различных подходов, методов и технологий интеллектуального анализа данных и обработки естественного языка для управления роботизированными производственными системами.

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, № 2019-0898.

Литература

1. How natural language processing helps manufacturing sector? [Электронный ресурс]. – Режим доступа: <https://www.analyticsinsight.net/how-natural-language-processing-helps-manufacturing-sector/>, свободный (дата обращения: 01.11.2022).
2. Knowledge Integration in Smart Factories / J. Zenkert, C. Weber, M. Dornhöfer, H. Abu-Rasheed, M. Fath // *Encyclopedia*. – 2021. – Vol. 1, No. 3. – P. 792–811.
3. Rabelo R.J. Softbots supporting the operator 4.0 at smart factory environments / R.J. Rabelo, D. Romero, S.P. Zambiasi // *IFIP International Conference on Advances in Production Management Systems*. – 2018. – P. 456–464.
4. Literature review on the ‘Smart Factory’ concept using bibliometric tools / F. Strozzi, C. Colicchia, A. Creazza, C. Noè // *International journal of production research*. – 2017. – Vol. 55, No. 22. – P. 6572–6591.

5. What is an engineering change request (ECR)? [Электронный ресурс]. – Режим доступа: <https://www.arena-solutions.com/resources/articles/engineering-change-request/> (дата обращения: 01.11.2022).
6. Wickel M.C. How to integrate information about past engineering changes in new change processes? / M.C. Wickel, U. Lindemann // *Proceedings of the 20th International Conference on Engineering Design (ICED 15)*. – 2015. – Vol. 3. – P. 229–238.
7. Sharafi A. Knowledge discovery in databases. – Springer Fachmedien Wiesbaden, 2013. – P. 51–108.
8. A posteriori design change analysis for complex engineering projects [Электронный ресурс]. – Режим доступа: <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/133/10/101005/467343/A-Posteriori-Design-Change-Analysis-for-Complex>, свободный (дата обращения: 02.11.2023).
9. Change propagation analysis in complex technical systems [Электронный ресурс]. – Режим доступа: <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/131/8/081001/466974/Change-Propagation-Analysis-in-Complex-Technical>, свободный (дата обращения: 02.11.2023).
10. Modeling engineering change management process in virtual collaborative design environments [Электронный ресурс]. – Режим доступа: <https://spectrum.library.concordia.ca/id/eprint/9094/>, свободный (дата обращения: 02.11.2023).
11. Knowledge-based methods for evaluation of engineering changes [Электронный ресурс]. – Режим доступа: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/78926/mehtacr_1.pdf?sequence=1, свободный (дата обращения: 02.11.2023).
12. Amarrsson I.O. Systematic Analysis of Engineering Change Request Data: Applying Data Mining Tools to Gain New Fact-Based Insights. – Sweden, Goeteborg: Chalmers Tekniska Hogskola, 2020. – 80 p.
13. Natural Language Processing for Manufacturing Industry [Электронный ресурс]. – Режим доступа: <https://www.stridelysolutions.com/resources/blog/natural-language-processing-for-manufacturing-industry/>, свободный (дата обращения: 01.11.2022).
14. Grieco A. On the application of text clustering in engineering change process / A. Grieco, M. Pacella, M. Blaco // *Procedia Cirp*. – 2017. – Vol. 62. – P. 187–192.
15. T. Kohonen Self-Organizing Maps // *Artificial Neural Networks*. – New York, 2001. – 501 p.
16. DEVOPEDIA for developers by developers [Электронный ресурс]. – Режим доступа: <https://devopedia.org/images/article/187>, свободный (дата обращения: 01.11.2022).
17. Canaan B. Microgrid cyber-security: Review and challenges toward resilience [Электронный ресурс]. – Режим доступа: <https://www.mdpi.com/2076-3417/10/16/5649>, свободный (дата обращения: 01.11.2023).
18. Gimenez-Aguilar M. Achieving cybersecurity in blockchain-based systems: A survey / M. Gimenez-Aguilar, J.M. de Fuentes, L. Gonzalez-Manzano, D. Arroyo // *Future Generation Computer Systems*. – 2021. – Vol. 124. – P. 91–118.
19. Adversarial attacks on machine learning cybersecurity defences in industrial control systems / E. Anthi, L. Williams, M. Rhode, P. Burnap, A. Wedgbury // *Journal of Information Security and Applications*. – 2021. – Vol. 58. – P. 102717.
20. Вигерс К. Разработка требований к программному обеспечению / К. Вигерс, Д. Битти. – М.: Русская редакция, 2004. – 576 с.
21. Wickel M.C. A retrospective analysis of engineering change orders to identify potential for future improvements / M.C. Wickel, U. Lindemann // *Proceedings of NordDesign*. – 2014. – P. 692–701.

Воробьева Алиса Андреевна

Канд. техн. наук, доцент фак. безопасности информационных технологий (ФБИТ) Национального исследовательского университета ИТМО (Университет ИТМО) Кронверкский пр-т, 49, А, г. Санкт-Петербург, Россия, 197101
 ORCID: 0000-0001-6691-6167
 Тел.: +7-921-947-21-14
 Эл. почта: vorobeva@itmo.ru

Федосенко Максим Юрьевич

Аспирант ФБИТ Университета ИТМО Кронверкский пр-т, 49, А, г. Санкт-Петербург, Россия, 197101
 ORCID: 0000-0001-8786-5661
 Тел.: +7-921-751-41-52
 Эл. почта: fedosenkomaksim98@gmail.com

Vorobeva A.A., Fedosenko M.Y.

Methods for data mining and natural language processing in the management of robotic production systems

The paper presents a comparative analysis of natural language processing and data mining methods used in data processing in the industrial robotic systems. A concept for implementing the process of integrating artificial intelligence methods into production management systems has been developed, and the necessary components have been described. The focus is on the use of natural language processing methods. The work could be useful for conducting experimental research in the field of application of natural language processing and artificial intelligence methods in robotic production systems.

Keywords: smart factories, machine learning, big data, robotic production systems, natural language processing, text mining, natural language processing, cybersecurity.

DOI: 10.21293/1818-0442-2023-26-3-65-71

References

1. *How natural language processing helps manufacturing sector?* Available at: <https://www.analyticsinsight.net/how-natural-language-processing-helps-manufacturing-sector/>, free (Accessed: November 01.11.2022).
2. Zenkert J., Weber C., Dornhöfer M., Abu-Rasheed H., Fath M. Knowledge Integration in Smart Factories, *Encyclopedia*, 2021, vol. 1, no. 3, pp. 792–811.
3. Rabelo R.J., Romero D., Zambiasi S.P. Softbots supporting the operator 4.0 at smart factory environments, *IFIP International Conference on Advances in Production Management Systems*, 2018, pp. 456–464.
4. Strozzi F., Colicchia C., Creazza A., Noè C. Literature review on the ‘Smart Factory’ concept using bibliometric tools, *International Journal of Production Research*, 2017, vol. 55, no. 22, pp. 6572–6591.
5. *What is an engineering change request (ECR)?* Available at: <https://www.arenasolutions.com/resources/articles/engineering-change-request/>, free (Accessed: November 01.11.2022).
6. Wickel M.C., Lindemann U. How to integrate information about past engineering changes in new change processes? *Proceedings of the 20th International Conference on Engineering Design (ICED 15)*, 2015, vol. 3, pp. 229–238.
7. Sharafi A. Knowledge discovery in databases. *Springer Fachmedien Wiesbaden*, 2013, pp. 51–108.
8. *A posteriori design change analysis for complex engineering projects.* Available at: <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/133/10/101005/>

467343/A-Posteriori-Design-Change-Analysis-for-Complex, free (Accessed: 02.11.2023).

9. *Change propagation analysis in complex technical systems.* Available at: <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/131/8/081001/466974/Change-Propagation-Analysis-in-Complex-Technical>, free (Accessed: 02.11.2023).

10. *Modeling engineering change management process in virtual collaborative design environments.* Available at: <https://spectrum.library.concordia.ca/id/eprint/9094/>, free (Accessed: 02.11.2023).

11. *Knowledge-based methods for evaluation of engineering changes.* Available at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/78926/mehtacr_1.pdf?sequence=1, free (Accessed: 02.11.2023).

12. Arnarsson I.O. *Systematic Analysis of Engineering Change Request Data: Applying Data Mining Tools to Gain New Fact-Based Insights.* Sweden, Goteborg, Chalmers Tekniska Hogskola, 2020. 80 p.

13. *Natural Language Processing for Manufacturing Industry.* Available at: <https://www.stridelysolutions.com/resources/blog/natural-language-processing-for-manufacturing-industry>, free (Accessed: 01.11.2022).

14. Grieco A., Pacella M., Blaco M. On the application of text clustering in engineering change process, *Procedia CIRP*, 2017, vol. 62, pp. 187–192.

15. Kohonen T. *Self-Organizing Maps, Artificial Neural Networks (Third Extended Edition)*, New York, 2001, 501 p.

16. *DEVOPEDIA for developers. by developers.* Available at: <https://devopedia.org/images/article/187/4433.1560446395.png>, free (Accessed: 01.11.2022).

17. *Microgrid cyber-security: Review and challenges toward resilience.* Available at: <https://www.mdpi.com/2076-3417/10/16/5649>, free (Accessed: 01.11.2023).

18. Gimenez-Aguilar M., de Fuentes J.M., Gonzalez-Manzano L., Arroyo D. Achieving cybersecurity in blockchain-based systems: A survey, *Future Generation Computer Systems*, 2021, vol. 124, pp. 91–118.

19. Anthi E., Williams L., Rhode M., Burnap P., Wedgbury A. Adversarial attacks on machine learning cybersecurity defenses in industrial control systems, *Journal of Information Security and Applications*, 2021, vol. 58, P. 102717.

20. Wieggers K., Beatty J. *Software requirements*, Pearson Education, 2013, 576 p.

21. Wickel M.C., Lindemann U. A retrospective analysis of engineering change orders to identify potential for future improvements, *Proceedings of NordDesign*, 2014, pp. 692–701.

Alisa A. Vorobeva

Candidate of Sciences in Engineering, Associate professor, Faculty of Secure Information Technologies ITMO University 49 bldg. A, Kronverksky pr., St. Petersburg, Russia, 197101
 ORCID: 0000-0001-6691-6167
 Phone: +7-921-947-21-14
 Email: vorobeva@itmo.ru

Maksim Yu. Fedosenko

Postgraduate student, Faculty of Secure Information Technologies ITMO University 49 bldg. A, Kronverksky pr., St. Petersburg, Russia, 197101
 ORCID: 0000-0001-8786-5661
 Phone: +7-921-751-41-52
 Email: fedosenkomaksim98@gmail.com