МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВОЛОДИМИРА ДАЛЯ

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ МОРСЬКИЙ УНІВЕРСИТЕТ

ЗБІРНИК НАУКОВИХ ПРАЦЬ

V-ї МІЖНАРОДНОЇ НАУКОВО-ПРАКТИЧНОЇ КОНФЕРЕНЦІЇ

«ТРАНСПОРТ: НАУКА ТА ПРАКТИКА»

21-22 травня 2026 р.



Україна, Київ – Одеса

очікування прибутку E[D], дозволяє не лише мінімізувати потенційні фінансові втрати, а й оптимізувати коефіцієнт використання місткості судна в умовах волатильного попиту. Встановлено, що досягнення динамічного балансу між провізною здатністю судна та ймовірнісним розподілом пасажиропотоку є критичним фактором забезпечення фінансової стійкості поромної лінії в довгостроковій перспективі.

Використана література:

1. Notteboom, T., Pallis, A. *Ferry and Ro-Ro Shipping: Operations and Markets.* — Routledge, 2020.
2. Maritime Economics: 3rd ed. Routledge, 2009. hapter 5–7: Demand, freight markets, shipping economics (ст. 120)
3. Ferry service network design under demand uncertainty. *Transportation Research Part E*, 2013. Sections 3–5 (probabilistic demand, service reliability, expected value models)
4. Introduction to Probability Models. 11th ed. Academic Press, 2014. Chapter 4–5: Normal distribution, expectation, truncated expectation E[min(X,C)]E[min(X,C)]
5. Kirillova, Y., Meleshenko, E.Development of an economic and mathematical model of loading a freight and passenger ferry. *Eastern-European Journal of Enterprise Technologies*, 2016.

УДК 656.25:004.89

Сотник В.О., к.т.н., доц., Щебликіна О.В., PhD, доц.

Український державний університет залізничного транспорту, Україна

ВИКОРИСТАННЯ NLP (NATURAL LANGUAGE PROCESSING) ДЛЯ АНАЛІЗУ ТЕХНІЧНОЇ ДОКУМЕНТАЦІЇ ТА ЖУРНАЛІВ ВІДМОВ

Анотація. *У доповіді розглядається концепція трансформації неструктурованих текстових звітів електромеханіків у стратегічний інформаційний ресурс за допомогою технологій обробки природної мови (NLP). Автори пропонують архітектуру інтелектуальної системи, що базується на модулях нормалізації, векторного представлення та аналітичного ядра з використанням нейромереж BERT і RoBERTa для автоматизації діагностики відмов. Впровадження запропонованої моделі дозволяє виявляти латентні закономірності в роботі*

систем залізничної автоматики, мінімізувати вплив людського фактору та забезпечити перехід до предиктивного обслуговування інфраструктури за її фактичним станом.

Проблематика дослідження. Сучасна залізнична інфраструктура генерує великі обсяги даних, проте найцінніша інформація про реальний стан систем СЦБ залишається закритою у неструктурованих текстових звітах, журналах та повідомленнях електромеханіків. Головна проблема полягає у неможливості оперативного ручного аналізу цих масивів через використання професійного сленгу, специфічних скорочень та суб'єктивізму в описах відмов, що створює ефект інформаційного хаосу. Відсутність автоматизованих інструментів нормалізації та семантичного аналізу таких текстів не дозволяє вчасно виявляти приховані системні дефекти та переходити від реактивного ремонту до предиктивного обслуговування за станом. Таким чином, виникає гостра науково-практична потреба у створенні спеціалізованих NLP-моделей, адаптованих до термінології Укрзалізниці, які здатні трансформувати інженерний досвід у структуровані цифрові дані для прийняття управлінських рішень.

В своїй доповіді ми зосереджуємо увагу на одному з найменш структурованих, але найбільш інформативних ресурсів нашої залізниці — текстових звітах і записах електромеханіків. Ми обговоримо впровадження технологій Natural Language Processing (NLP) для аналізу технічної документації та журналів відмов. У сучасних системах залізничної автоматики ми маємо справу з колосальним обсягом даних. Проте значна частина інформації про реальний стан техніки зберігається у вільному текстовому форматі: це записи в різних формах журналів, таких, як ДУ-46, журнали реєстрації вимірів параметрів та виконаних робіт, звіти про відновлювальні роботи та пояснювальні записки. Ручний аналіз тисяч таких записів для виявлення системних дефектів є фізично неможливим, а проста фільтрація за ключовими словами не враховує контексту та специфічного професійного жаргону.

Використання технології NLP (Natural Language Processing) у нашому секторі дозволить вирішити цю проблему. Методи обробки природної мови дозволять перетворити текст на структуровані дані. Для реалізації цього завдання слід виділити три ключові етапи автоматизації: вилучення сутностей, класифікація причин несправностей та виявлення латентних(прихованих) закономірностей.

На етапі вилучення сутностей алгоритм розпізнає в тексті конкретні об'єкти: "стрілочний електропривод", "тип реле", "рейкове коло", а також локацію та час.

Далі, на етапі класифікації причин несправностей за допомогою нейромереж (наприклад, архітектури BERT або RoBERTa) автоматично відбувається віднесення запису до певної категорії, скажімо зовнішній вплив (атмосферні явища), механічний знос, помилки програмного забезпечення або людський фактор.

На етапі виявлення латентних (прихованих) закономірностей технологія NLP дозволяє знайти зв'язки, які не очевидні при лінійному читанні. Наприклад, частота згадок про окислення контактів у звітах по конкретній дистанції сигналізації може вказувати на системний брак партії комплектуючих або порушення технології регламентного обслуговування.

Запропонована модель автоматизації складається з наступних модулів:

Модуль нормалізації. Це фундамент всієї системи NLP. Його завданням є перетворення хаотичного набору символів на стандартизований текст, який нейромережа (BERT чи RoBERTa) зможе адекватно проаналізувати. На цьому етапі ми видаляємо "шум", який не несе смислового навантаження для аналізу причин відмов. Видаляються зайві пробіли, спецсимволи та дати (якщо вони вже зафіксовані в метаданих системи), виправляються типові помилки розпізнавання тексту (якщо журнали оцифровувалися зі сканів). Стандартні токенізатори часто розбивають технічні назви на незрозумілі частини. Модуль нормалізації повинен розпізнавати технічні одиниці як цілісні об'єкти. Наприклад: Замість розбиття

“СП-6М” на “СП”, “-“, “6”, “М”. Система повинна бачити це як єдиний токен “електропривод_тип_СП6М”.

Далі в цьому модулі відбувається дебрівіація та складання словника професійного сленгу. Це найвідповідальніша частина для залізничної автоматики. Модуль замінює скорочення на повні форми, щоб уникнути неоднозначності. Кожній абревіатурі присвоюється її конкретне значення: РК - рейкове коло, ЕЦ – електрична централізація, АЛС - автоматична локомотивна сигналізація, ШН – електромеханік СЦБ (важливо для ідентифікації виконавця в тексті) та інше.

Враховуючи специфіку експлуатації в Україні, записи часто містять змішування мов або технічні жаргонізми (наприклад, “фрикція” замість “фрикційна муфта”, “напільний” замість “підлоговий” або “колійний”). Модуль використовує алгоритми відстані Левенштейна або векторні підстановки для приведення цих слів до літературної та технічної норми.

Далі відбувається лематизація та морфологічний аналіз – приведення всіх слів до початкової форми (наприклад, “недотягування”, “недотягнув”, “недотягує” → “недотягування”). Для української мови це складно через багату флексію, тому тут застосовуються спеціалізовані бібліотеки (наприклад, Stanza або NLP-UK), навчені на технічних текстах.

Без якісної нормалізації ми отримаємо ефект “GIGO” (Garbage In, Garbage Out – сміття на вході, сміття на виході). Якщо ми хочемо виявити системну проблему, наприклад, масовий вихід з ладу певної серії реле в конкретній кліматичній зоні, модуль нормалізації має гарантувати, що всі згадки цього реле (навіть з помилками чи скороченнями) будуть зведені до одного ідентифікатора.

Створення якісного модуля нормалізації для Укрзалізниці вимагає залучення не лише програмістів, а й фахівців служб сигналізації та зв'язку та департаменту Автоматики та телекомунікацій Укрзалізниці, які можуть скласти вичерпний словник експлуатаційних термінів. Саме такий синергетичний підхід ми зараз закладаємо в основу наших практичних посібників з впровадження ШІ.

Модуль векторного представлення. Питання модуля векторного представлення слід розглядати не просто як програмування, а як спосіб оцифрування інженерної інтуїції. У нашій справі цей модуль є важливою ланкою. Він перетворює нормалізовані слова (токени), про які ми говорили раніше, у багатовимірні математичні вектори. Це дозволяє машині розуміти не просто букви, а технічний сенс та контекст експлуатаційної ситуації.

При цьому кожне слово або термін перетворюється на точку у просторі з сотнями координат. Головна мета тут у тому, що схожі за змістом терміни опиняються поруч у цьому математичному просторі. Наприклад, вектор слова “електропривід” буде математично ближчим до вектора “електродвигун”, ніж до вектора “баласт”. Це дозволяє ШІ розуміти синоніми та професійний жаргон, навіть якщо в журналах вони записані по-різному. В цьому модулі доцільно виділити три рівні його функціонування. На рівні семантичного кодування будуються контекстні вектори. Наприклад: якщо в звіті написано “обрив у колі”, вектор слова “коло" буде вказувати на електричний ланцюг. Якщо ж написано “колія 1а”, вектор того ж слова зміститься в бік колійного господарства. Модуль враховує оточення слова на 360 градусів.

Далі проводиться професійна адаптація – донавчання моделей на корпусі технічних текстів Укрзалізниці та специфікацій виробників. Стандартна нейромережа сприймає абревіатуру “ЕЦ” як щось загальне. Модуль векторного представлення має чітко позиціонувати її поруч із "централізацією", "релейним стативом" та "керуванням рухом". Стандартна нейромережа сприймає абревіатуру "ЕЦ" як щось загальне. Модуль векторного представлення має чітко позиціонувати її поруч із “централізацією”, “релейним стативом” та “керуванням рухом”.

На заключному, третьому рівні, проводиться формування векторів цілих звітів. Модуль не просто створює вектори окремих слів, а агрегує їх у єдиний вектор цілого запису в журналі. Це дозволяє порівнювати цілі інциденти. Система обчислює подібність між новим записом та тисячами архівних. Якщо

вектор сьогоднішньої відмови на станції А збігається за напрямком із вектором відмови, що сталася місяць тому на станції Б, система сигналізує про системний д–ефект конкретного типу обладнання.

Аналітичне ядро. Ми розглядаємо аналітичне ядро не як набір алгоритмів, а як "мозок" системи, що приймає відповідальні рішення. Якщо модуль нормалізації готує дані, а модуль векторного представлення їх розуміє, то аналітичне ядро – це саме та інстанція, яка формулює висновки для керівництва дистанції (ШЧ) або служби (Ш). Це ядро можна структурувати як багаторівневу систему, що вирішує три стратегічні завдання: класифікацію, діагностику та прогнозування. Аналітичне ядро базується на ансамблі нейромереж, де кожна відповідає за свій сектор аналізу. Класифікатор інцидентів на основі векторів від BERT/RoBERTa відносить подію до певної категорії (наприклад, "Хибна зайнятість рейкового кола через порушення ізоляції"). Детектор системних аномалій використовує методи навчання без учителя (Unsupervised Learning) для пошуку нетипових записів, які не вкладаються в стандартні сценарії відмов. Механізм логічного виведення поєднує результати нейромережі з жорсткими правилами технічної експлуатації (ПТЕ). Це гарантує, що висновки ШІ не суперечитимуть залізничним стандартам безпеки. Ядро аналізує текстовий опис і автоматично виділяє першопричину. Часто електромеханік пише наслідок ("немає контролю стрілки"), але ядро, аналізуючи тисячі подібних записів, ідентифікує, що причиною є, наприклад, "зниження ізоляції кабелю" при певних погодних умовах. Ядро також проводить виявлення латентних (прихованих) трендів. Це те, що можна назвати цифровою інтуїцією. Воно здатне помітити, наприклад, що на певній дільниці за останній місяць кількість згадок про нагрів контактів зросла на 15%. Це ще не призвело до аварії, але ядро вже формує попередження про необхідність позапланової ревізії.

А далі в роботу вступає предиктивна аналітика (прогнозування). На основі динаміки записів у журналах відмов ядро розраховує ймовірність виходу з ладу

конкретного вузла протягом наступних періодів. Це дозволяє перейти від регламентного обслуговування до обслуговування за станом.

У наших дослідженнях ми застосовуємо метод “людина в контурі”. Якщо ядро помилилося в класифікації, досвідчений інженер вносить корективу. Ядро миттєво перенавчається, адаптуючи свою вагову сітку під реалії залізниці. Враховується також географічний та технічний контекст. Ядро знає різницю між експлуатацією систем на гірських дільницях Карпат та на швидкісних магістралях, де динамічні навантаження зовсім інші. Таким чином аналітичне ядро - це інструмент, який дозволяє нам подолати інформаційний хаос. Впровадження такого ядра дозволяє скоротити час на розслідування причин затримок поїздів орієнтовно на 30-40% і практично виключити людський фактор при первинному аналізі технічних інцидентів. Це і є перехід до Smart Railway – залізниці, яка вміє думати і попереджати проблеми.

Впровадження технологій обробки природної мови (NLP) дозволяє трансформувати накопичений десятиліттями інформаційний хаос журналів та звітів електромеханіків у структурований стратегічний ресурс. Застосування розроблених модулів нормалізації, векторного представлення та аналітичного ядра вже зараз дозволяє оптимізувати процес розслідування відмов, усунути елементи суб’єктивізму, застосувати інтелектуальну діагностику та цифрову стандартизацію.

Висновок. Інтелектуальний аналіз тексту – це інструмент глибокого розуміння реального стану систем автоматики. Поєднання практичного досвіду експлуатації з потужністю нейромереж типу BERT/RoBERTa є ключовим етапом створення надійної та безпечної залізничної інфраструктури майбутнього.

sotnyk.va@kart.edu.ua
sov@kart.edu.ua

Сільванська Г.М., Деордєєва А.Є. СТОХАСТИЧНЕ МОДЕЛЮВАННЯ ОПЕРАЦІЙНОЇ ЕФЕКТИВНОСТІ ТА УПРАВЛІННЯ РЕНТАБЕЛЬНІСТЮ RO-PAX ПЕРЕВЕЗЕНЬ: КЕЙС СУДНА «SANTOÑA» 189
Сотник В.О., Щебликіна О.В. ВИКОРИСТАННЯ NLP (NATURAL LANGUAGE PROCESSING) ДЛЯ АНАЛІЗУ ТЕХНІЧНОЇ ДОКУМЕНТАЦІЇ ТА ЖУРНАЛІВ ВІДМОВ..... 195
Сумець О.М. ІДЕНТИФІКАЦІЯ ФАКТОРІВ ВПЛИВУ НА ЛОГІСТИЧНІ ПРОЦЕСИ ТОРГОВЕЛЬНИХ ПІДПРИЄМСТВ 202
Супринович В.П., Кічкіна О.І. АДАПТИВНИЙ ПІДХІД ДО ПЛАНУВАННЯ ПРОЦЕСУ ПОСТАЧАННЯ МЕТАЛОВИРОБІВ НА ЕКСПОРТ ЧЕРЕЗ ПОРТ..... 205
Тарасенко Д.В. ОПТИМІЗАЦІЯ МІЖПОЇЗДНИХ ІНТЕРВАЛІВ НА ЗАЛІЗНИЧНИХ ДІЛЬНИЦЯХ УКРАЇНИ НА ОСНОВІ ВПРОВАДЖЕННЯ ВІРТУАЛЬНИХ БЛОК-ДІЛЬНИЦЬ..... 209
Тихонін В. І., Тихоніна І.І. ЗАСТОСУВАННЯ ДЕРАТИЗАЦІЇ ПРИ ПЕРЕВЕЗЕННІ ВАНТАЖІВ МОРЕМ 216
Ушаков А.М., Баркалова Н.О., Вернигора Р. В. ПРОБЛЕМИ ПЕРЕРОБКИ КОНТЕЙНЕРІВ В МОРСЬКИХ ПОРТАХ УКРАЇНИ В УМОВАХ ВОЄННОГО СТАНУ 221
Харченко Д.Р., Прохорченко А.В. РОЗРОБКА ФУНКЦІЇ КОРИСНОСТІ ПЕРЕВІЗНИКА ПРИ РОЗПОДІЛІ ПРОПУСКНОЇ СПРОМОЖНОСТІ ЗАЛІЗНИЧНОЇ ІНФРАСТРУКТУРИ В МОДЕЛІ ШТАКЕЛЬБЕРГА 225
Хітров І.О. ПІДВИЩЕННЯ БЕЗПЕКИ ПЕРЕВЕЗЕНЬ ВАНТАЖІВ ЗА РАХУНОК ОПТИМІЗАЦІЇ СИСТЕМ КРІПЛЕННЯ ТА РОЗМІЩЕННЯ 230
Худяков І.В., Грицук І.В., Погорлецький Д.С. МОНІТОРИНГ ЕНЕРГЕТИЧНИХ УСТАНОВОК СУДЕН ЗМІШАНОГО РАЙОНУ ПЛАВАННЯ «РІКА-МОРЕ»..... 233
Шандер О.Е. НАПРЯМКИ УДОСКОНАЛЕННЯ ІНТЕРМОДАЛЬНИХ ПЕРЕВЕЗЕНЬ В УМОВАХ ВІЙНИ І ПІСЛЯВОЄННОГО ВІДНОВЛЕННЯ .. 242