

Класифікація тексту

Misha Beherksy

1 червня 2017 р.

This is abstract

1 Теоретичні основи е-е

На відміну від штучно створених мов, наприклад мов програмування чи математичних нотацій, мови, які ми використовуємо для спілкування, розвивалися з покоління в покоління, постійно видозмінюючись, а тому досить складно відслідкувати і встановити набір чітких конкретно визначених правил. Розробка алгоритмів, що дозволяють "розуміти" людські висловлювання дає змогу покращити велику кількість аспектів взаємодії людини та комп'ютера: передбачення вводу, розпізнавання тексту, пошук інформації в неструктурованому тексті, переклад з однієї мови на іншу, аналіз емоційного забарвлення тексту та багато іншого. Створюючи інтерфейси, що дозволяють людині більш ефективно використовувати комп'ютер, ми прискорюємо розвиток багатомовного інформаційного суспільства.

The text classification problem Doe [2100]

In text classification, we are given a description \mathbb{X} of a document, where \mathbb{X} is the document space ; and a fixed set of classes \mathbb{C} Classes are also called categories or labels . Typically, the document space \mathbb{X} is some type of high-dimensional space, and the classes are human defined for the needs of an application, as in the examples China and documents that talk about multicore computer chips above.

1.1 Exploratory data analysis

Візуалізація для наступних цілей: * Комунікативна - представлення даних та ідей - проінформувати - підтримати і аргументувати - вплинути і переконати * Дослідницька - вивчити (дослідити) дані - проаналізувати ситуацію - визначити наступні кроки - прийняти рішення стосовно деякого питання

$$\alpha = \sqrt{\beta} \quad (1)$$

1.2 Класифікація тексту

Метою класифікації текстів є розподіл документів на групи наперед визначених категорій. *-*

1.3 Учи матчасть

Результати показують, що стабільно показують чудові результати для завдань класифікації текстів, суттєво перевищуючи показники інших методів.

2 Розробка алгоритму

Оптимізація алгоритмів для використання у предметній галузі

3 Програмна реалізація

Особливості і деталі програмної реалізації

4 Аналіз рішення

Анализ, согласно критериям как работает, пути улучшения (таблица сравнения с существующими подходами, графики, диаграммы)

| Зміст ідеї | Напрямки застосування | Вигоди для користувача |
|---|--|---|
| Система побудови універсальних пропозитів для підвищення їх ефективності та продуктивності роботи загалом | 1. Використання науковцями та підприємствами однакових даних для підвищення їх ефективності та продуктивності роботи загалом | Зручний та зрозумілий вихідний код дозволяє новим програмістам писати ефективніше, тим самим зосереджуючись на прикладних задачах, замість деталей реалізації |
| 2. Узагальнення алгоритмів для роботи з різними типами даних | Універсальність моделі дозволить не перемикати контексти під час роботи з різними типами даних, використовуючи однаковий підхід для вхідної інформації | |
| 3. Отримання кращих результатів передбачень для даних, що змінюються з часом | Допомога під час роботи з величинами, що залежать від часу: курси валют, показники біржі, зміни клімату | |

Табл. 1: Опис ідеї стартап-проекту

| № п/п | Показники стану ринку (найменування) | Характеристика |
|-------|--|---|
| 1 | Кількість головних гравців, од | 1 |
| 2 | Загальний обсяг продаж, грн/ум. од | 914 218 млн грн |
| 3 | Динаміка ринку (якісна оцінка) | Спадає |
| 4 | Наявність обмежень для входу (вказати характер обмежень) | Висока доля невизначеності, відсутність попереднього досвіду та необхідних статистичних даних |
| 5 | Специфічні вимоги до стандартизації та сертифікації | - |
| 6 | Середня норма рентабельності в галузі (або по ринку), % | 18-20% |

Табл. 2: Попередня характеристика потенційного ринку стартап-проекту

5 Стартап

Опис ідеї стартап-проекту

| | | | |
|------------------|-------------|----------------|------------------|
| numeric literals | integers | in decimal | 8743 |
| | | in octal | 0o7464 |
| | | | 0O103 |
| | | in hexadecimal | 0x5A0FF |
| | 0xE0F2 | | |
| | fractionals | in decimal | 140.58 |
| | | | 8.04e7 |
| | | | 0.347E+12 |
| 5.47E-12 | | | |
| 47e22 | | | |
| char literals | | | 'H' |
| | | | '\n' |
| | | | '\x65' |
| string literals | | | "bom dia" |
| | | | "ouro preto\nmg" |

Ринок є доволі привабливим для входження: пристойна середня норма рентабельності, що трохи вище за середній банківський відсоток на вклади у гривні, а спадання ринку потенційно відкриває його для нестандартних інноваційних рішень, оскільки існує дуже висока необхідність в розробці універсального методу для відновлення зображень.

Обрано альтернативу 2 як таку, що має на увазі довше життя проекту.

В якості цільових груп обрано: 1 та 2.

5.0.1 Тут субсекція з висновками

Проведений детальний аналіз ринку та перспектив розвитку проекту дав змогу отримати такі результати:

- Існує можливість ринкової комерціалізації проекту, на ринку наявний попит на пропонований продукт.

| № п/п | Потреба, що формує ринок | Цільова аудиторія (цільові сегменти ринку) | Відмінності у поведінці різних потенційних цільових груп клієнтів | Вимоги споживачів до товару |
|-------|---|--|---|---|
| 1 | Необхідність для інвесторів знайти перспективний метод для вкладень | Люди, які мають фінансову можливість та зацікавленість робити інвестиції у інноваційні проекти | Люди, які мають фінансову можливість та зацікавленість робити інвестиції у інноваційні проекти мають на меті збільшення свого капіталу, підвищення свого іміджу, а також долучитися до новітніх технологій, щоб бути у тренді | Необхідно розробити методику оцінювання та рекомендації, які б з високою ймовірністю розраховували потенційні необхідні інвестиції та шляхи попередження ключових ризиків |
| 2 | Необхідність команди для побудови цього | Активні люди, які бажають втілити у життя свій проект | Необхідність проаналізувати всі ключові фактори, щоб визначити, чи доцільно реалізовувати проект та чи вдасться залучити спонсорів | Високоточний метод оцінки відновлення зображень, щоб визначити доцільність реалізації відновлення зображень |

Табл. 3: Характеристика потенційних клієнтів стартап-проекту

| № п/п | Фактор | Зміст загрози | Можлива реакція компанії |
|-------|-------------|---|---|
| 1 | Попит | Не вдасться розробити унікальний метод, який би можна було застосовувати для будь-яких алгоритмів та адаптувати для роботи з різними типами даних | Розробка максимально універсального методу |
| 2 | Конкуренція | Можливість появи конкурентів з дуже схожими функціями, їх вихід на ринок раніше за нас | Доопрацювання якості розробленого методу з фокусом на зручність та простоту використання, розробка нових властивостей, яких немає у конкурента. Розгляд можливості об'єднання компаній для подальшої спільної роботи. |
| 3 | Економічні | Зменшення доходу інвесторів, що призведе до зменшення кількості інвестицій | Моніторинг економічної ситуації у країні, пошук закордонних користувачів та адаптація для світового ринку |

Табл. 4: Фактори загроз

| № п/п | Фактор | Зміст можливості | Можлива реакція компанії |
|-------|---------------------|--|---|
| 1 | Попит | Унікальність пропонованого функціоналу та додаткових можливостей при умові високої конкуренції дозволить захопити велику частку ринку, особливо зацікавивши додатком невеликих інвесторів (бізнес-ангелів) та команди проєктів, які не потребують значних інвестицій | Адаптація до ринку, що розширюється, моніторинг новітніх розробок та ризиків, які тільки нещодавно з'явилися |
| 2 | Науково-технічні | Поява нових технологій, виникнення нових ринкових умов та факторів, які виявлять значний вплив на розвиток алгоритмів класифікації | Активне використання використання рішення; у випадку, якщо наше рішення буде одним з перших та матиме суттєві відмінності від аналогів, захист інтелектуальної власності розробників, патентування цієї технології та надання її до інтелектуальних активів проєкту |
| 3 | Соціально-культурні | Велика популярність сфери роботи з даними та їх аналізу | Адаптація системи до розширення ринку, появи нових умов та технологій |

Табл. 5: Фактори можливостей

- Ринок відкритий для інновацій, прослідковується позитивна динаміка ринку.
- Рентабельність роботи на ринку вища за прибутковість банківських вкладів, а отже приваблює як інвесторів, так і розробників для роботи над перспективним проєктом.
- З огляду на потенційні групи клієнтів існує потенціал та перспектива входу на ринок.
- Істотні бар'єри для входження відсутні.
- В якості варіанту для впровадження для ринкової реалізації проєкту доцільно обрати довгострокову роботу та утримання клієнтів, роботу над покращенням розробленого методу з використанням багатовимірного статистичного аналізу.
- Конкуренція практично відсутня, а конкурентноспроможність самого продукту достатньо висока. ...

Враховуючи описані вище ключові моменти, можна зробити висновок, що подальша імплементація даного проєкту є доцільною та обґрунтованою.

| | | |
|---|--|--|
| Особливості конкурентного середовища | В чому проявляється дана характеристика | Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентноспроможною) |
| 1. Тип конкуренції - чиста конкуренція | Велика кількість методів відновлення зображень, частина з яких є запатентованою інтелектуальною власністю | Звертати увагу на якість та універсальність методу відновлення зображень |
| 2. За рівнем конкурентної боротьби - національний | Відновлення зображень не буде прив'язуватися до географічних показників | Акцент в рекламі на потреби жителів великих міст (столиці), таргетування на науковців та молодих дослідників, а також на високозабезпечених людей - потенційних інвесторів |
| 3. За галузевою ознакою - внутрішньогалузева | Конкуренцію складають подібні методики розробки прогностичних моделей | Акцентувати увагу на незвичайність подачі послуг, а також зручність у використанні та надійність, яку вони забезпечують |
| 4. Конкуренція за видами товарів - між бажаннями | Потенційні клієнти роблять вибір між звичними методами побудови моделей (яких дуже велика кількість) і відчують складність у виборі найбільш доцільного методу | Чітко зрозуміти потреби та бажання кожної з груп цільової аудиторії та розробляти гнучку систему, яка задовольнятиме потреби всіх груп користувачів |
| 5. За характером конкурентних переваг - нецінова | Акцент знаходиться на унікальності та якості послуг, що надаються, а також на перевагах, які отримує клієнт під час використання наших послуг | Робота над покращення методики побудови прогностичних моделей та підвищенням її універсальності |
| 3. За інтенсивністю - не марочна | Продається втілення ідеї, а не певний бренд | Просування ідеї у соціальних мережах |

Табл. 6: Ступеневий аналіз конкуренції на ринку

| Складові аналізу | Прямі конкуренти в галузі | Потенційні конкуренти | Постачальники | Клієнти | Товари-замінники |
|------------------|--|---|---|---|---|
| | Прямах конкурентів немає, непрямі - різноманітні методи побудови прогностичних моделей | Нові розробки у галузі | Інвестори диктують умови розвитку ринку: ключова умова - проект повинен бути потрібним користувачам та приносити користь | Кількість зацікавлених клієнтів, рівень зацікавленості в такому типі послуг | Поява схожих дешевших або якісніших продуктів-конкурентів |
| Висновки | Прямах конкурентів немає | - можливості входу в ринок присутні, необхідно вирішити проблему пошуку та адаптації статистичних даних - необхідність розробки універсального методу, який може бути використаний як інвесторами, так і командою проекту | Успіх нашого проекту залежить від рівня довіри інвесторів та команд проекту до новітнього методу побудови прогностичних моделей | Клієнти формують попит на таку послугу | Універсальних методів, які могли б замінити запропонований проект немає |

Табл. 7: Аналіз конкуренції в галузі за М. Портером

| № п/п | Фактор конкурентноспроможності | Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим) |
|-------|--|--|
| 1 | Фактор часу | Ідея є частково новою, для перейняття ідеї та втілення її у життя потенційним конкурентам знадобиться час |
| 2 | Фактор новизни товару | Початковий успіх продукту очікується через його новизну та інтерес цільової аудиторії до нових інноваційних рішень |
| 3 | Фактор якості послуг та надання інформації | Науковці та експерти з обробки даних потребують універсальний метод побудови прогностичних моделей |

Табл. 8: Обґрунтування факторів конкурентноспроможності

| № п/п | Фактор конкурен-тності | Бали 1-20 Рейтинг товарів-конкурентів у порівнянні з іншими методами оцінювання | 2 | 3 | 4 | 5 | 6 |
|-------|--|--|---|---|---|---|---|
| 1 | Фактор часу | 15 | | | + | | |
| 2 | Фактор новизни товару | 20 | | + | | | |
| 3 | Фактор якості послуг та надання інформації | 17 | | + | | | |

Табл. 9: Порівняльний аналіз сильних та слабких сторін методу

| | |
|---|---|
| Сильні сторони: Якість послуг, що надаються Новизна послуг Можливість використання як інвесторами, і командою з розробки | Слабкі сторони: Відсутність статистичних даних та попереднього досвіду в реалізації подібних рішень |
| Можливості: Створення нової ринкової ніші Потреба у ефективному та компактному методі створення прогностичних моделей Необхідність закладати у бюджет можливі ризики та зміни ринкових умов | Загрози: Різка зміна ринку, поява нових стартапів, економічна криза |

Табл. 10: SWOT-аналіз стартап-проекту

| № п/п | Альтернатива (орієнтовний комплекс заходів) ринкової поведінки | Ймовірність отримання ресурсів | Строки реалізації |
|-------|--|---|---|
| 1 | - Ціль: отримання прибутку в короткостроковій перспективі - Конкуренція: цінова та партнерська (пропонуємо свої нові послуги розповсюдження інформації про партнерів - рекламні послуги) - Взаємодія з фірмами: активна боротьба за долю ринку, що належить конкурентам | В короткостроковому плані - велика В довгостроковому плані - значний ризик втрати частини ринку, якщо займатися лише ціновою конкуренцією | 8-12 місяців після запуску проекту |
| 2 | - Ціль: захоплення частини ринку, підтримання її розміру та поступове нарощення об'ємів - Конкуренція: нецінова (акцент на тому, що пропонуємо інноваційні послуги) - Взаємодія з конкурентами: співпраця, активний моніторинг їх діяльності, при можливій появі реальних конкурентів можна запропонувати злиття компаній/проектів | Висока ймовірність отримання ресурсів та утримання їх протягом довгого проміжку часу. Більш ймовірний розвиток компанії та постійне покращення продукту | 8-12 місяців після запуску проекту - для отримання перших фінансових надходжень від розповсюдження інформації про акції магазинів-партнерів, та їх реклама. Далі фінансові надходження прогнозовано регулярними |

Табл. 11: Альтернативи ринкового впровадження стартап-проекту

| № п/п | Опис профілю цільової групи потенційних клієнтів | Готовність споживачів сприйняти продукт | Орієнтовний попит в межах цільової групи (сегменту) | Інтенсивність конкуренції в сегменті | Простота входу у сегмент |
|-------|---|---|--|--------------------------------------|--|
| 1 | Високозабезпечені люди, які зацікавлені у пошуку перспективних проектів для інвестування | Споживачі слідкують за найновітнішими технологіями, бажають бути в тренді та готові сприйняти новий продукт | Потенційно високий, інвестори хочуть бути впевненими у доцільності своїх інвестицій та подальшому отриманні прибутку | Практично відсутня | При наявності достойної та до-ручної реклами - досить просто |
| 2 | Ініціативні люди та науковці, які мають хорошу ідею в схожій сфері та хочуть втілити її у життя | Споживачі готові сприйняти продукт, так як зацікавлені у глибинному аналізі ситуації | Високий попит | Практично відсутня | При наявності достойної та до-ручної реклами - досить просто |

Табл. 12: Вибір цільових груп потенційних споживачів

| № п/п | Обрана альтернатива розвитку проекту | Стратегія охоплення ринку | Ключові конкурентноспроможні позиції відповідно до обраної альтернативи | Базова стратегія розвитку |
|-------|--|--------------------------------------|---|---------------------------|
| 1 | Захоплення, підтримання та захист частки ринку | Стратегія концентрованого маркетингу | - Новизна послуг - Доступність продукту - Простота в користуванні продуктом - Додаткові зручні аспекти, які враховуються під час розрахунку ефективності та інвестиційної привабливості побудови прогностичних моделей, що вигідно виділяють наш продукт серед конкурентів | Стратегія диференціації |

Табл. 13: Визначення базової стратегії розвитку

| № п/п | Чи є проект "першо-прохідцем" на ринку | Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів | Чи буде компанія копіювати основні характеристики товару конкурента і які? | Стратегія конкурентної поведінки |
|-------|--|---|--|----------------------------------|
| 1 | Частково | Нові споживачі, частково забиратиме споживачів конкурентів | Частково. Новий метод оцінювання ефективності побудови прогностичних моделей буде агрегувати декілька методик аналізу, що дозволить оцінювати проекти більш точно з використанням більшої кількості факторів, що впливають на проект | Стратегія лідера |

Табл. 14: Визначення базової стратегії конкурентної поведінки

| № п/п | Вимоги до товару цільової аудиторії | Базова стратегія розвитку | Ключові конкурентноспроможні позиції власного стартап-проекту | Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових) |
|-------|--|---------------------------|---|---|
| 1 | Універсальність методу оцінювання з точки зору інвестора | Стратегія диференціації | Врахування всіх аспектів оцінювання проекту з точки зору інвестиційної привабливості | - Ваші гроші ефективно працюють у інноваційному прогресивному проекті |
| 2 | Універсальність методу оцінювання з точки зору команди проекту | Стратегія диференціації | Врахування всіх аспектів оцінювання проекту з точки зору інвестиційної привабливості та життєздатності і проекту, доцільність реалізовувати інноваційний проект | - Реальна можливість втілити у життя ідею завдяки глибинному аналізу ключових аспектів та пошуку інвесторів |
| 3 | Необхідність враховувати ризики проекту, ринкові та економічні умови, що швидко змінюються | Стратегія диференціації | Врахування ключових ризиків та ринкових умов завдяки розробленій системі коефіцієнтів | - Детальний облік ризиків та моніторинг ринкових умов дозволять уникнути передчасного закриття проекту |

Табл. 15: Визначення стратегій позиціонування

| № п/п | Потреба | Вигода, яку пропонує товар | Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити) |
|-------|---|---|---|
| 1 | Універсальний метод оцінювання ефективності та інвестиційної привабливості проекту, який буде корисний як для інвесторів, так і для команди проекту | Методика оцінювання дозволить уникнути передчасного закриття проекту та перевитрат бюджету завдяки високоточній оцінці на ранніх етапах проекту | Оцінювання проекту як з точки зору витрат та ефективності їх використання командою стартапу, так і з урахуванням потенційних ризиків, прихованих стратегічних переваг на недоліків. |

Табл. 16: Визначення ключових переваг концепції потенційного товару

| № п/п | Рівень цін на товари-замінники | Рівень цін на товари-аналоги | Рівень доходів цільової групи споживачів | Верхня та нижня межі встановлення ціни на товар/послугу |
|-------|--------------------------------|------------------------------|--|---|
| 1 | Безкоштовно | Безкоштовно | Більше 10000 грн/місяць | - |

Табл. 17: Визначення меж встановлення ціни

| № п/п | Специфіка закупівельної поведінки цільових клієнтів | Функції збуту, які має виконувати поставальник товару | Глибина каналу збуту | Оптимальна система збуту |
|-------|---|--|----------------------|--------------------------|
| 1 | Купують право на використання методики | Зберігання, сортування, встановлення контакту інформування | Однорівневий | Залучена |

Табл. 18: Формування системи збуту

| № п/п | Специфіка поведінки цільових клієнтів | Канали комунікацій, якими користуються цільові клієнти | Ключові позиції, обрані для позиціонування | Завдання рекламного повідомлення | Концепція рекламного звернення |
|-------|--|---|--|---|--|
| 1 | Позитивне відношення до інновацій та швидкий розвиток технологій призводять до появи великої кількості нових методів, росту кількості даних і побудова прогностичних моделей стає більш актуальною | Соціальні мережі (facebook, twitter), тематичні ресурси | Універсальний метод побудови прогностичних моделей | Впевнити клієнта у тому, що метод є унікальним та універсальним | Повідомлення у соціальних мережах, статті на веб-ресурсах, короткі демонстраційні ролики |

Табл. 19: Концепція маркетингових комунікацій

6 Висновки

Тут багато незрозумілих слів, ще більше води, ніж у всіх інших частинах диплому

Результати показують, що стабільно показують чудові результати для завдань класифікації текстів, суттєво перевищуючи показники інших методів.

Література

John Doe. The Book without Title. Dummy Publisher, 2100.

Robert W. Fairlie. Kauffman Index of Entrepreneurial Activity. Kansas City: Ewing Marion Kauffman Foundation, 2014.

Brad Feld. Startup communities: Building an entrepreneurial ecosystem in your city. Hoboken, NJ: John Wiley & Sons, 2012.

Young-Hoon Kwak. A brief history of Project Management. Greenwood Publishing Group, 2005.

Dane Stangler. The Economic Future just Happened. Kansas City: Ewing Marion Kauffman Foundation., 2009.

Martin Stevens. Project Management Pathways. Association for Project Management. APM Publishing Limited, 2002.

Шмидт С. Бирман Г. Капиталовложения. Экономический анализ ин-вестиционных проектов. М.: ЮНИТИ-ДАНА, 2003.

Лапыгин Ю. Н. Управление проектами: от планирования до оценки эффективности. М.: Омега-Л, 2008.