

## МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИДОБУВАННЯ СТРУКТУРОВАНОЇ ІНФОРМАЦІЇ ПРО ТОВАР З ТЕКСТІВ КОРИСТУВАЧІВ

Хасцький М.В.<sup>1)</sup>, Микитюк В.А.<sup>2)</sup>, Ділай П.Р.<sup>3)</sup>, Костик Б.П.<sup>4)</sup>, Поворозник Б.С.<sup>5)</sup>

*Західноукраїнський національний університет*

*<sup>1)</sup>магістрант;<sup>2)</sup>аспірант;<sup>3)</sup>магістрант;<sup>4)</sup>аспірант;<sup>5)</sup>магістрант*

### І. Постановка проблеми

Виявлення ставлення користувачів до товарів і послуг, що купуються, є важливою частиною роботи маркетингового відділу будь-якого бізнесу. Спираючись на отриману інформацію, компанії можуть керувати асортиментом продукції, враховувати переваги окремих соціальних груп при формуванні рекламних повідомлень, діагностувати проблеми, що виникають при експлуатації, і вчасно реагувати на них, використовувати негативний зворотний зв'язок для формування бачення майбутніх продуктових лінійок. Доступним джерелом великої кількості вихідних даних для подібного аналізу є численні сайти, де споживачі можуть обговорювати товари, висловлюючи свою думку про різні аспекти товару [1,2].

У цій роботі пропонується витягувати з текстів відгуків окремі оціночні висловлювання покупців, які містили б у собі основний зміст повідомлення, у формі, корисній при вирішенні виникаючих на етапі супроводу продукту завдань.

### II. Мета роботи

Метою дослідження є розробка математичного та програмного забезпечення для видобування структурованої інформації про товар з текстів користувачів.

### III. Математичне забезпечення для вилучення та аналізу думок користувачів про споживчі властивості товарів

Дослідження текстів відгуків про товари показало, що найчастіше у своїх відгуках користувачі згадують три джерела проблем: логістичний фактор (швидкість доставки, якість упаковки, стан товару), якість товарів, досвід спілкування з продавцем. Часто спостережувану поведінку можна пояснити тим, що споживачі з обережністю ставляться до покупок у даному магазині і частіше вибирають дешеві товари без високих вимог до якості, турбуючись переважно про терміни та якість доставки. Опис якості, як правило, досить небагатослівний і обмежується загальною оцінкою товару. Це дозволяє зробити висновок про те, що аналіз буде корисним маркетологам для покращення якості до ставки та пошуку шляхів підвищення довіри у користувачів. Можна сказати, що у майбутньому ситуація почне змінюватися, оскільки інтернет-магазини активно покращують логістичний аспект свого бізнесу. У міру того як проблеми із доставкою стануть менше, користувачі почнуть більше уваги приділяти безпосередньо якості товару [3,4].

Для об'єднання множини типів думок користувачів у групи з метою виявлення найбільш важливих напрямів розвитку продукту та супровідних сервісів, в які потрібно вкласти додаткові ресурси, запропоновано оригінальний класифікатор, заснований на елементах універсальної моделі діяльності: суб'єкт діяльності, об'єкт на який спрямована діяльність, засоби використовувані у процесі діяльності, взаємозв'язку між елементами діяльності. Стосовно особливостей розв'язуваного завдання суб'єктом діяльності є продавець, об'єктом діяльності – продукт, що підлягає продажу.

До основних засобів можна віднести сайт інтернет-магазину, логістичні служби, відділ маркетингу. Виходячи із запропонованих елементів декомпозиції, пропонується асоціювати з кожним її елементом типи оціночних висловлювань, які користувачі можуть використовувати для висловлювання своєї думки під час написання текстів відгуків. Вибір типів думок здійснювався у відповідності з критерієм інформативності/значимості висловлювання для вирішення основних завдань, що виникають на етапах експлуатації та супроводу продуктів [3,4].

Споживчі властивості товару: думки про товар в цілому та його експлуатаційні характеристики, такі як якість виконання, виконання заявлених функцій, зовнішній вигляд товару.

Споживчі властивості товарів конкурентів: думки цілком аналогічні класу "Споживчі властивості товару", однак вони стосуються характеристик товарів конкурентів, висловлених споживачами у відгуках.

Якість обслуговування: думка споживачів про те, як продавець враховує їхні побажання, зокрема щодо конкретних товарів. У цю категорію також можна віднести згадки про готовність продавця робити спеціальні пропозиції, такі як знижки, безкоштовна доставка, подарунки.

Здатність вирішувати конфліктні ситуації: оцінка споживачами того, як продавець поводить себе у разі конфлікту інтересів, такого як виявлення дефектів, проблеми з доставкою товару тощо.

Гарантійне обслуговування: оцінка справедливості гарантійних термінів та умов обслуговування, якість та оперативність ремонту товарів.

Оплата: форма та якість процедури оплати товарів та/або послуг.

Доставка: швидкість відправлення товару, якість упаковки товару, цілісність товару після прибуття. До цієї групи включаються лише ті аспекти доставки, на які може вплинути продавець.

На основі проведеного аналізу текстів відгуків про продукти користувачів інтернет-магазину для організації розмітки даних та проведення експериментального дослідження було обрано типи думок і об'єднані в наступні класи:

- товар (споживчі якості товару);
- продавець (загальна якість обслуговування, здатність вирішувати конфліктні ситуації);
- доставка (швидкість та якість доставки).



Рисунок 1 –Класифікатор типів думок користувачів

З урахуванням вищезазначеного завдання вилучення та аналізу думок користувачів про товар може бути подане в наступному вигляді. Нехай задано множину текстів відгуків покупців  $T$ , що містять думки з множини  $O$  про набір продуктів  $P$ . Думкою називатимемо четвірку (Аспект, Опис, Тональність, Тип). Аспектом називатимемо послідовність слів, що позначає важливу характеристику або згадку про продукт. Описом будемо називати послідовність слів, що містить висловлену користувачем думку про певний аспект.

Тональність буде визначатися категоріальною шкалою з трьох рівнів: позитивна (+), нейтральна (0), негативна (-). Множина типів думок буде подана відповідно до введеного ієрархічного класифікатора, проте для формування навчального набору даних та практичної апробації будемо використовувати узагальнені класи: {Товар, Продавець, Доставка}.

Маючи вихідні множини  $T, P, O$  і задану структуру класифікатора думок, необхідно витягти з нових текстів відгуків  $T'$  множини думок користувачів  $O'$  про новий набір товарів  $P'$ . Для вирішення цього завдання потрібно адаптувати нейромережу у модель шляхом визначення складу множин  $Lbl, A, V(a)$  згідно смислового наповнення, введеного вище поняття думки.

$$\begin{aligned}
 Lbl &= \{ \text{Аспект}, \text{Опис} \}, \\
 A &= \{ \text{Тональність}, \text{Мета} \}, \\
 V(\text{Тональність}) &= \{ +, 0, - \}, \\
 V(\text{Мета}) &= \{ \text{Товар}, \text{Продавець}, \text{Доставка} \}.
 \end{aligned} \tag{1}$$

У функцію  $A(C_t)$  потрібно додати нову умову, що дозволяє будувати зв'язок тільки в тих випадках, коли вона поєднує аспект та опис. Виходячи з даної формалізації, множину переходів визначено таким чином:

$$Y = \{ \text{Shift}, \text{Start}(\text{Аспект}), \text{Start}(\text{Опис}), \\ \text{Add}(\text{Аспект}), \text{Add}(\text{Опис}), \text{Link}(n_1, n_2), \text{End}, \\ \text{Attr}(\text{Тональність}, +), \text{Attr}(\text{Тональність}, 0), \\ \text{Attr}(\text{Тональність}, -), \text{Attr}(\text{Мета}, \text{Товар}), \\ \text{Attr}(\text{Мета}, \text{Продавець}), \text{Attr}(\text{Мета}, \text{Доставка}) \}. \quad (2)$$

Загальна архітектура нейронної мережі для вилучення думок з текстів з урахуванням заданих  $Lbl$ ,  $A$ ,  $V(a)$  та  $A(C_t)$ , отримана із загальної моделі, наведена на рисунку 2.

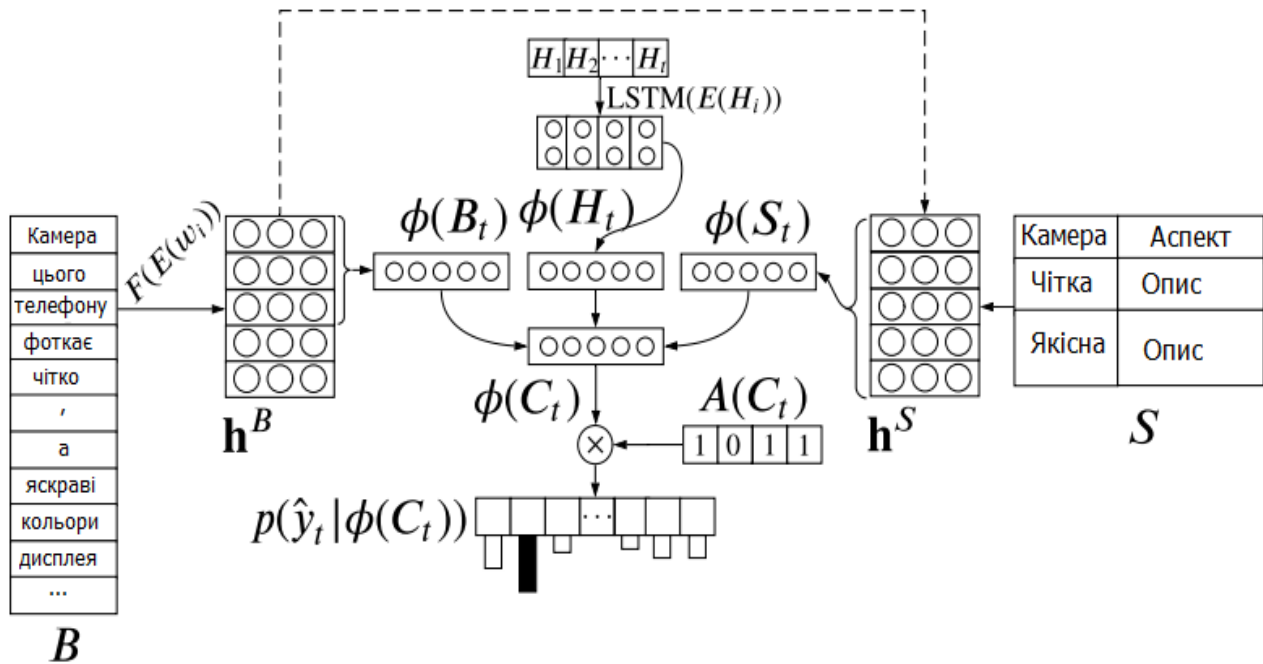


Рисунок 2—Архітектура нейронної мережі для отримання думок

Запропонована загальна нейромережева модель на основі системи переходів для вилучення складових об'єктів та їх атрибутів дозволяє адаптувати її для вилучення об'єктів у різних предметних областях шляхом конкретизації складу множин типів фрагментів та атрибутів.

### Висновок

В рамках цього дослідження запропонована модельна основі системи переходів для вилучення складових об'єктів та їх атрибутів. Вона має більш високу порівняно з альтернативами якість визначення значень атрибутів при збільшенні довжини речення, що свідчить про здатність краще враховувати контекстну інформацію при здійсненні передбачень. Використана методика оцінки моделей та отримані результати дозволяють зробити висновок про практичну придатність моделі для вилучення думок про товари з категорій, не включених до навчальної вибірки без додаткового навчання.

### Список використаних джерел

1. Зайченко Ю. П. Основи проектування інтелектуальних систем / Ю.П. Зайченко. - Київ: Видавничий дім «Слово», 2004. - 352 с.
2. Терейковський І.А., Заріцький О.В., Терейковська Л.О., Погорелов В.В. Метод розробки архітектури глибокої нейронної мережі, призначеної для розпізнавання комп'ютерних вірусів. Захист інформації. 2018. Т. 20, № 3, С. 188–199.
3. -Zaragoza J., Toselli A. H., Vidal E. Hand written Music Recognition for Mensural Notation: Formulation, Data and Baseline Results. 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017. Pp. 1081–1086.
4. Tapia E., Rojas R. Recognition of on-line handwritten mathematical formulas in the E-chalk system. Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, 2003. Pp. 980–984.