УДК 004.891

[1)]С. Лупенко, [1)]О. Пастух, [1)]Б. Хомів, [2)]Ю. Нікольський
[1)]Тернопільський національний технічний університет ім. І. Пулюя,
кафедра комп'ютерних систем та мереж
[2)]Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж

ЗАСТОСУВАННЯ ЛІНГВІСТИЧНИХ ЗМІННИХ ТА ВАГОВИХ КОЕФІЦІЄНТІВ ПІД ЧАС ФОРМУВАННЯ ІНТЕГРАЛЬНОЇ ОЦІНКИ ОБ'ЄКТА У ЗАДАЧАХ OPINION MINING

Описанл процедуру формування інтегрального показника об'єкта згідно з відгуками користувачів під час застосування методів оцінювання опінії текстової інформації у web-документах.

Запропоновано використання лінгвістичних змінних та застосування вагових коефіцієнтів для достовірнішого результату оцінювання емоційного забарвлення текстової інформації.

Ключові слова: опінія, емоційне забарвлення, об'єкт, інтегральний показник, лінгвістична змінна, ваговий коефіцієнт.

Article is devoted to the process of forming the integral indicator of the object according to user reviews, using the opinion mining methods of textual information in the web-documents.

The use of linguistic variables and weighting coefficients for more reliable opinion mining result of text information is proposed in article.

Key words: opinion, object, integral indicator, linguistic variable, weighting coefficient.

Вступ

У зв'язку зі стрімким розвитком інформаційних технологій та інтернет-ресурсів, зокрема, виникла потреба у аналізі надвеликих масивів текстової інформації. Одним із підходів до аналізу текстової інформації є галузь Opinion Mining – виявлення, визначення та добування емоційного забарвлення висловлювань (опінії) у текстових даних. Емоційне забарвлення, тобто опінія, притаманне великій кількості текстової інформації, зокрема, й текстовій інформації, що міститься у web-документах (коментарях Інтернет-магазинів, блогів, форумів та ін.). Опінія [оуп'ін`ійа] – уявлення про якість, характер, значення когось або чогось [1]. Термін «емоційне забарвлення» є синонімом до терміну опінія.

Аналіз останніх досліджень та публікацій. постановка задачі

За останні декілька років різко збільшилась кількість досліджень та публікацій у галузі добування опінії, це пов'язано зі стрімким розвитком інтернет-ресурсів, збільшенням кількості форумів, використанням соціальних мереж. Так сформувалась нова галузь – добування та аналіз опінії в текстових даних інтернет-ресурсів.

У роботі [2] запропоновано структуру задач та методів оцінювання опінії, що зображена на рис. 1. У галузі Opinion Mining виділяють задачі: розроблення лінгвістичних ресурсів, класифікації та узагальнення емоційного забарвлення висловлювань. Слід відзначити, що використання методів, котрі належать до галузі Opinion Summarization (узагальнення опінії) сьогодні не дають цілком достовірних та виправданих результатів, а сама галузь є малодослідженою.

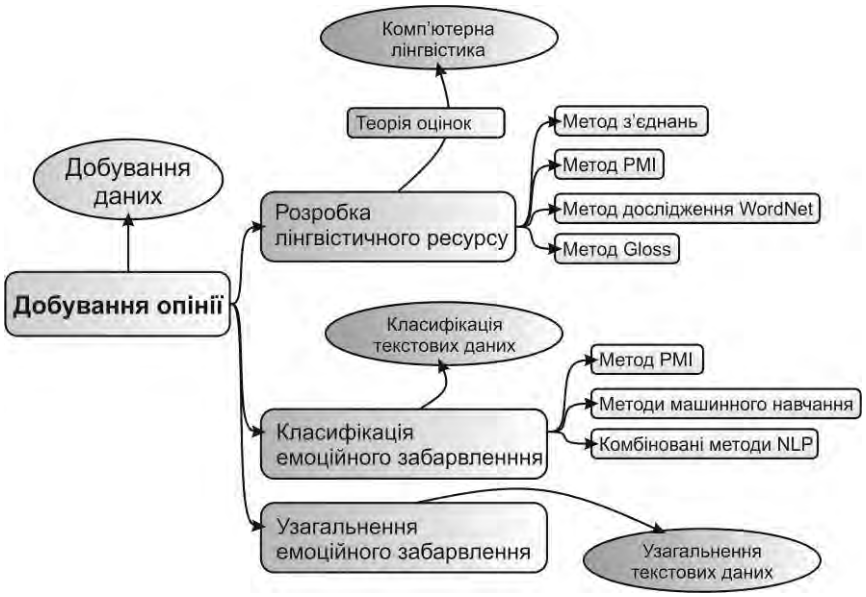


Рис. 1. Структура задач та методів оцінювання опінії

Сьогодні існує вже багато програмних систем та засобів, зокрема: Opinion Observer, OPINE, Review Seer, Red Opal, Web Fountain, Opinion Miner, Bing Shopping, Google Products, Quark Shop [3–11], що використовують наведені вище методи для оцінювання опінії, а саме визначення емоційного забарвлення (позитивного, нейтрального чи негативного) висловлювань щодо певного об'єкта чи його компонент.

Потрібно відзначити, що ці програмні засоби та системи у разі узагальнення результатів оцінювання опінії та порівняння декількох об'єктів між собою, згідно з висловлюваннями користувачів, не завжди об'єднують числові значення емоційного забарвлення компонентів об'єкта до єдиного інтегрального показника або ж об'єднують їх, не враховуючи диференціацію важливості цих компонент. Також класифікація опінії здійснюється, як правило, із використанням трьох класів: позитивного, негативного та нейтрального, що не повною мірою дає змогу оцінити емоційне забарвлення висловлювань.

Тому, для підвищення інформативності та точності програмних систем із використанням засобів та методів Opinion Mining було б доцільно використовувати:

- гнучкіший математичний апарат для задачі оцінювання опінії, котрий дав би змогу отримувати ширший діапазон числових значень емоційного забарвлення висловлювань;
- гнучкіший математичний апарат для задачі узагальнення опінії, котрий дав би змогу отримати інтегральний показник об'єкта згідно з числовими значеннями емоційного забарвлення висловлювань.

Формулювання цілей статті

Метою роботи є розроблення математичного забезпечення та алгоритму оцінювання опінії об'єкта, що дало б змогу підвищити інформативність, достовірність та точність отриманих результатів.

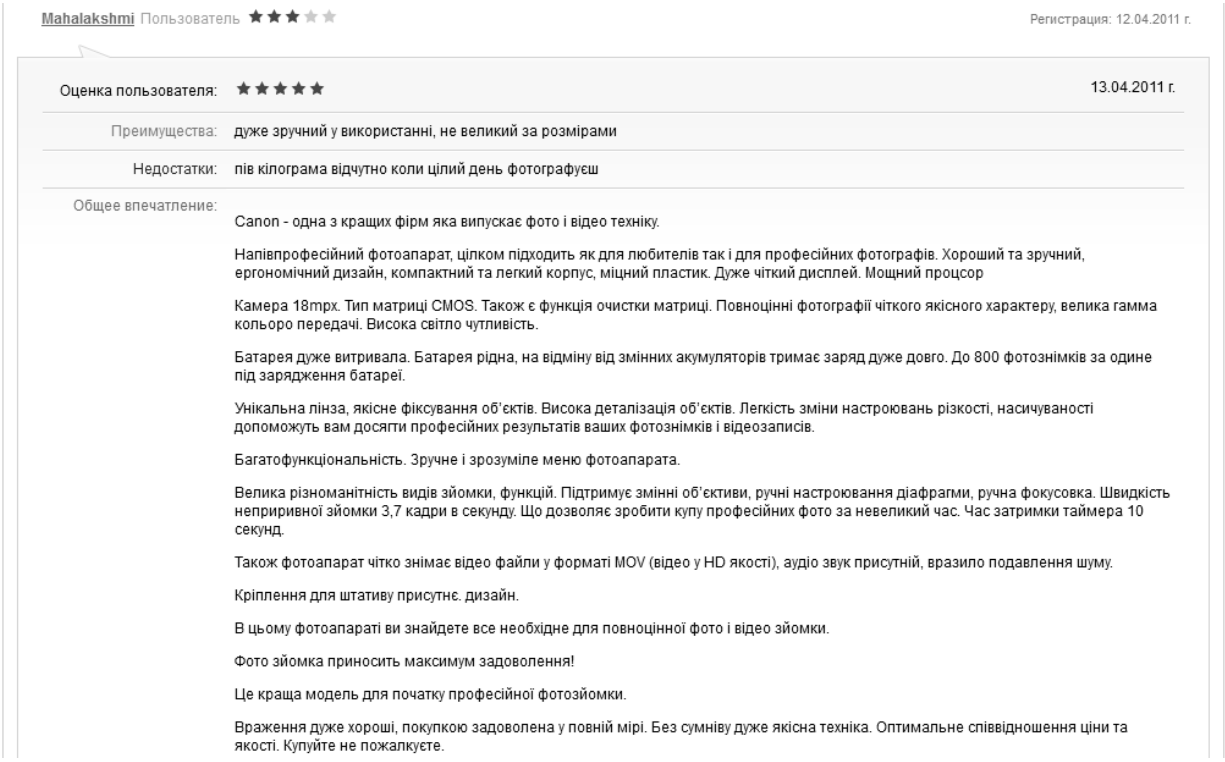
Для досягнення мети, а саме підвищення інформативності, достовірності та точності результатів під час оцінювання опінії текстової інформації необхідно виконати такі завдання:

- проаналізувати математичне представлення опінії в текстових даних та методів побудови ієрархічного графа об'єкта;
- проаналізувати застосування вагових коефіцієнтів, їхню ефективність та важливість;
- описати емоційне забарвлення висловлювань (опінію) у вигляді лінгвістичних змінних, провести їхню фазифікацію та дефазифікацію;
- здійснити формування інтегрального показника об'єкта згідно з відгуками користувачів, враховуючи значення отриманих лінгвістичних змінних та вагових коефіцієнтів.

Основний матеріал

1. Математичний формалізм емоційного забарвлення (опінії) в текстовій інформації

Як було описано вище, емоційне забарвлення притаманне коментарям інтернет-магазинів. Під час аналізу опінії в тестових даних виділяють об'єкти, їх компоненти та обчислюють числові значення емоційного забарвлення висловлювань. На рис. 2 зображено типовий коментар інтернет-магазину стосовно об'єкта «Фотокамера Canon EOS 550D Kit». У цьому коментарі користувач на ім'я Mahalakshmi висловлює свої враження та думки (опінії) стосовно фотокамери Canon EOS 550D Kit, виділяє її переваги та недоліки, описує характеристики, надає емоційного забарвлення коментарю завдяки таким висловлюванням, як «Canon – одна з найкращих фірм», «Дуже чіткий дисплей» та ін.


Mahalakshmi Пользователь ★★★☆☆ Регистрация: 12.04.2011 г.

Оценка пользователя: ★★★★★ 13.04.2011 г.

Преимущества: дуже зручний у використанні, не великий за розмірами

Недостатки: пів кілограма відчутно коли цілий день фотографуєш

Общее впечатление:

Canon - одна з кращих фірм яка випускає фото і відео техніку.

Напівпрофесійний фотоапарат, цілком підходить як для любителів так і для професійних фотографів. Хороший та зручний, ергономічний дизайн, компактний та легкий корпус, міцний пластик. Дуже чіткий дисплей. Мощний процсор

Камера 18mpx. Тип матриці CMOS. Також є функція очистки матриці. Повноцінні фотографії чіткого якісного характеру, велика гамма кольоро передачі. Висока світло чутливість.

Батарея дуже витривала. Батарея рідна, на відміну від змінних акумуляторів тримає заряд дуже довго. До 800 фотознімків за одине під зарядження батареї.

Унікальна лінза, якісне фіксування об'єктів. Висока деталізація об'єктів. Легкість зміни настроювань різкості, насичуваності допоможуть вам досягти професійних результатів ваших фотознімків і відеозаписів.

Багатофункціональність. Зручне і зрозуміле меню фотоапарата.

Велика різноманітність видів зйомки, функцій. Підтримує змінні об'єктиви, ручні настроювання діафрагми, ручна фокусовка. Швидкість неприривної зйомки 3,7 кадри в секунду. Що дозволяє зробити купу професійних фото за невеликий час. Час затримки таймера 10 секунд.

Також фотоапарат чітко знімає відео файли у форматі MOV (відео у HD якості), аудіо звук присутній, вразило подавлення шуму.

Кріплення для штативу присутнє. дизайн.

В цьому фотоапараті ви знайдете все необхідне для повноцінної фото і відео зйомки.

Фото зйомка приносить максимум задоволення!

Це краща модель для початку професійної фотозйомки.

Враження дуже хороші, покупкою задоволена у повній мірі. Без сумніву дуже якісна техніка. Оптимальне співвідношення ціни та якості. Купуйте не пожалкуєте.

Рис. 2. Коментар інтернет-магазину стосовно об'єкта «Фотокамера Canon EOS 550D Kit»

Проаналізувавши наведений вище коментар, програмні системи Opinion Mining формують граф, який відображає структуру об'єкта, що містить компоненти, підкомпоненти та атрибути з відповідними значеннями емоційного забарвлення. Частина графа наведена на рис. 3.

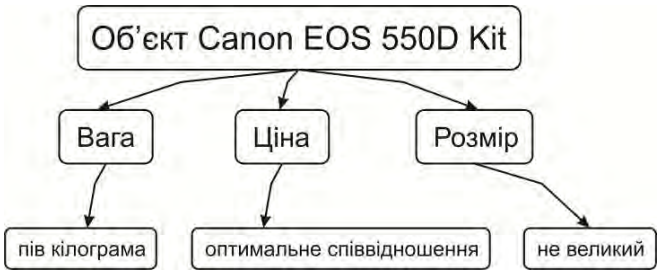


Рис. 3. Граф об'єкта

Як відомо [12], опінія текстової інформації в коментарях інтернет-магазинів подається п'ятіркою

$$O = (e_j, a_{j,k}, so_{j,k,i,l}, h_i, t_l), \quad (1)$$

де e_j – цільовий об'єкт (сутність), $a_{j,k}$ – аспект/компонента/властивість цільового об'єкта e_j, $so_{j,k,i,l}$ – значення емоційного забарвлення опінії, висловленого власником опінії h_i стосовно

компоненти $a_{j,k}$ цільового об'єкта e_j в час t_l ($so_{j,k,i,l}$ може набувати таких числових значень: +1 – позитивне висловлювання, -1 – негативне висловлювання, 0 – нейтральне висловлювання, або приймати інші числові значення залежно від вибраної шкали).

У цій роботі пропонується уточнення моделі опінії (1), а саме:

- введення ваги $w_{j,k}$ для компоненти об'єкта $a_{j,k}$, у зв'язку з тим, що різні властивості/компоненти/характеристики можуть бути більш чи менш важливими;
- запровадження лінгвістичної змінної для подання емоційного забарвлення $so_{j,k,i,l}$, що дасть змогу отримувати точніше числове значення опінії висловлювань;
- введення ваги w_i для власника опінії h_i, у зв'язку з тим, що різні люди можуть мати різні пріоритети, наприклад, висловлювання експерта в цій галузі повинно бути важливішим, ніж інші висловлювання;
- введення ваги w_l для часу t_l, коли була висловлена опінія, у зв'язку з тим, що в певний період часу висловлювання може бути актуальнішим, ніж в інший.

Отже, враховуючи наведені вище міркування, модель (1) для формалізованого опису опінії об'єкта можна зобразити у такому вигляді:

$$Opinion = \{O_j, (F_{j,k}, W_{j,k}), SO_{j,k,i,l}, (H_i, W_i), (T_l, W_l)\}, \quad (2)$$

де $\{O_j, j = \overline{1,J}\}$ – множина об'єктів. O_j – об'єкт (сутність), J – кількість об'єктів; $\{F_{j,k}, j = \overline{1,J}, k = \overline{1,K_j}\}$ – множина компонент. $F_{j,k}$ – аспект/компонента/властивість цільового об'єкта O_j, K_j – кількість компонент j-го об'єкта. Якщо $K_j = 1$, то $O_j = F_{j,1}$ – у випадку, коли у висловлюванні відсутнє згадування певної компоненти об'єкта, проте описується об'єкт загалом; $\{W_{j,k}, j = \overline{1,J}, k = \overline{1,K_j}\}$ – вага компоненти $F_{j,k}$; $\{SO_{j,k,i,l}, j = \overline{1,J}, k = 1,K_j, i = \overline{1,I_{j,k}}, l = \overline{1,L_{j,k,i}}\}$ – множина емоційного забарвлення висловлювань. $SO_{j,k,i,l}$ – значення емоційного забарвлення опінії, висловленого власником опінії H_i стосовно компоненти $F_{j,k}$ цільового об'єкта O_j в час T_l. $SO_{j,k,i,l}$ представляється у вигляді лінгвістичної змінної та може набувати числових значень відповідно до її терм-множини; $\{W_i, i = \overline{1,I_{j,k}}\}$ – вага власника опінії H_i. $I_{j,k}$ – кількість власників опінії, що прокоментували k-й компонент j-го об'єкта; $\{W_l, l = \overline{1,L_{j,k,i}}\}$ – вага часової мітки T_l. $L_{j,k,i}$ – кількість часових міток, в які i-й власник опінії прокоментував k-й компонент j-го об'єкта.

Введення таких уточнень для математичної формалізації опінії дасть змогу отримати достовірніший результат під час узагальнення емоційного забарвлення висловлювань.

Коментар, зображений на рис. 2, можна описати за допомогою вищенаведеної математичної моделі (2) та зобразити у вигляді графа об'єкта, що зображений на рис. 4, де O_1 – фотокамера Canon EOS 550D Kit; $F_{1,1}$ – вага, $F_{1,2}$ – ціна, $F_{1,3}$ – розмір, $F_{1,4}$ – ергономіка; H_1 – Mahalakshmi; T_1 – 13.04.11, $SO_{1,4,1,1}$ – дуже зручний у використанні, $SO_{1,3,1,1}$ – невеликий за розмірами, $SO_{1,2,1,1}$ – пів кілограма.

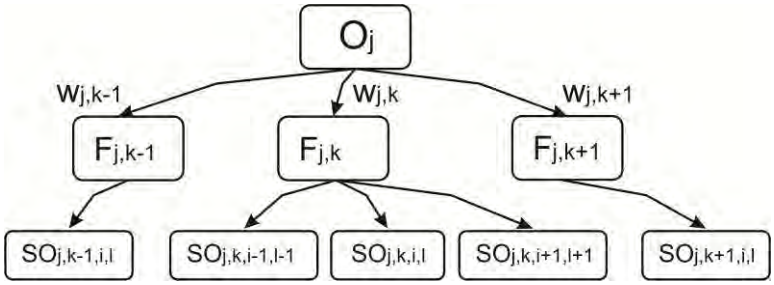


Рис. 4. Граф об'єкта

2. Застосування вагових коефіцієнтів

Під час інтегрального оцінювання опінії об'єкта «фотокамера» його характеристики (компоненти) такі, як «ціна» та «якість» у моделі (1), матимуть однакову вагу, що є не повною мірою виправданим та коректним з точки зору покупця. Проте, якщо ввести вагові коефіцієнти між такими компонентами, як «батарея», «зображення», «ергономіка» та «ціна», то може виявитися, що компонента «ціна» важливіша за компоненту «батарея», а компонента «ергономіка» важливіша за компоненту «ціна» під час порівняння двох, чи більше об'єктів, що дасть змогу достовірніше та точніше оцінити та порівняти ці об'єкти.

Важливою задачею є врахування вагових коефіцієнтів $W_{j,k}$ між компонентами $F_{j,k}$, W_i між власниками опінії H_i та W_l відносно часової мітки T_l в математичній моделі опінії (2) з урахуванням відгуків користувачів для покращення систем підтримки прийняття рішень під час ранжування чи порівняння об'єктів.

Розглянемо для прикладу такий випадок: нехай є два об'єкти – фотокамера 1 та фотокамера 2 з компонентами «батарея», «якість зображення», «оптичний зум» та «ціна». У табл. 1 наведено кількість позитивних, негативних, нейтральних відгуків користувачів та обчислено загальну кількість відгуків покомпонентно (стовпець 7), рейтинг компоненти об'єкта (стовпець 8) та відсоткове співвідношення рейтингу компоненти (стовпець 9).

На рис. 5, а зображено діаграму відгуків користувачів, що побудована згідно з табл. 1. Виникає запитання, яку з фотокамер вважати кращою? Теоретично можна вважати кращою фотокамеру 2, тому що згідно з діаграмою відсоткового співвідношення емоційного забарвлення коментарів в неї переважають компоненти «батарея» та «якість зображення». Проте, якщо ввести вагові коефіцієнти між такими компонентами, як «батарея», «якість зображення», «оптичний зум» та «ціна», то може виявитися, що компонента «ціна» важливіша за компоненту «батарея», а компонента «оптичний зум» важливіша за компоненту «ціна».

Отже, виникає потреба у введенні вагових коефіцієнтів $W_{j,k}$ для компонент об'єкта $F_{j,k}$. У табл. 1 (стовпець 10) наведено значення вагових коефіцієнтів, котрі обчислюються за формулою (3), враховуючи стовпець 7 табл. 1.

$$W_{j,k} = \frac{n_в}{n_{в.k}}, \tag{3}$$

де $n_в$ – загальна кількість висловлювань, що обчислюється за формулою (4); $n_{в.k}$ – кількість висловлювань щодо певного компоненту всіх об'єктів, що обчислюється за формулою (5).

$$n_в = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \sum_{i=1}^{I_{j,k}} L_{j,k,i}, \tag{4}$$

$$n_{в.k} = \sum_{j=1}^{J} \sum_{i=1}^{I_{j,k}} L_{j,k,i}. \tag{5}$$

Обчислити кількість висловлювань про компонент певного об'єкта $n_{в.\,j,k}$ та кількість висловлювань про об'єкт $n_{в.\,j}$ можна за такими формулами:

$$n_{в.\,j,k} = \sum_{i=1}^{I_{j,k}} L_{j,k,i}, \tag{6}$$

$$n_{в.\,j} = \sum_{k=1}^{K_j} \sum_{i=1}^{I_{j,k}} L_{j,k,i}, \tag{7}$$

де $j = \overline{1, J}$, J – кількість об'єктів; $k = \overline{1, K_j}$, K_j – кількість компонент j-го об'єкта; $i = \overline{1, I_{j,k}}$, $I_{j,k}$ – кількість власників опінії, що прокоментували k-й компонент j-го об'єкта; $l = \overline{1, L_{j,k,i}}$, $L_{j,k,i}$ – кількість часових міток, в які i-й власник опінії прокоментував k-й компонент j-го об'єкта.

Таблиця 1

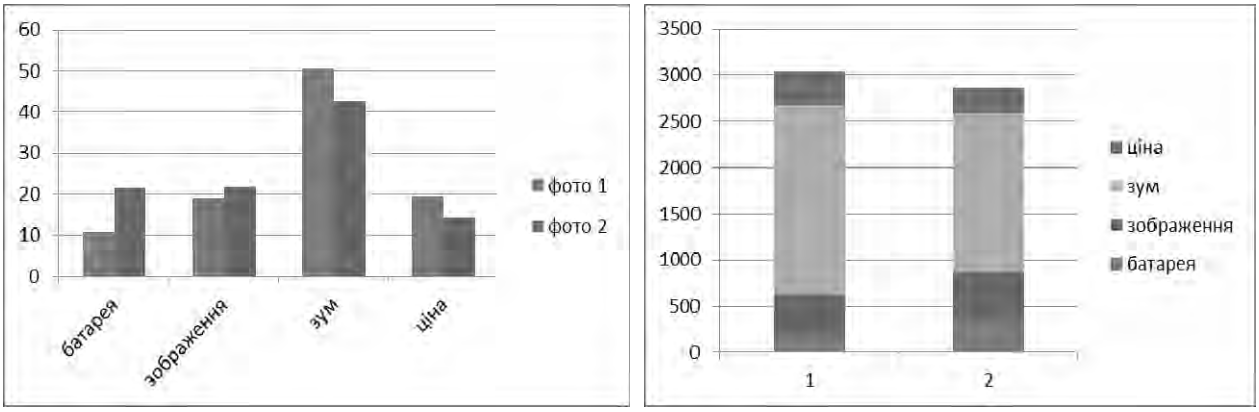
Співвідношення кількості відгуків користувачів

1	2	3	4	5	6	7	8	9	10	11
E	A	SO_{pos}	SO_{neg}	SO_{neu}	SO	SO	SO	SO	SO	SO
фото 1		поз (+1)	нег (-1)	нейтр (0)	всього	заг. к-ть	рейтинг	%	коеф.	% * коеф
	батарея	125	25	3	153	513	100	10,88139	16,9980119	184,962
	зображення	243	69	2	314	706	174	18,93362	23,3929755	442,9138
	зум	500	34	1	535	1218	466	50,70729	40,3578529	2046,437
	ціна	232	53	4	289	581	179	19,47769	19,2511597	374,9682
				всього	1291	3018	919	100	100	3049,281
фото 2		поз (+1)	нег (-1)	нейтр (0)	всього	заг. к-ть	рейтинг	%	коеф.	% * коеф
	батарея	350	8	2	360	513	342	21,49591	16,9980119	365,3878
	зображення	367	21	4	392	706	346	21,74733	23,3929755	508,7347
	зум	680	2	1	683	1218	678	42,61471	40,3578529	1719,838
	ціна	256	31	5	292	581	225	14,14205	19,2511597	272,2508
				всього	1727	3018	1591	100	100	2866,211

Завдяки введенню вагових коефіцієнтів для компонент об’єкта, отримано діаграму співвідношення кількості відгуків користувачів із застосування вагових коефіцієнтів, що зображена на рис. 5, б. Як видно з рис. 5, б, фотокамера 1 буде кращою від фотокамери 2 згідно з відгуками користувачів, (див. табл. 1, стовпець 11).

Проаналізувавши методи підрахунку позитивних та негативних висловлювань [13–15], було виявлено такі принципи знаходження вагових коефіцієнтів:

1) за кількістю згадувань (використовується у цій роботі);
2) за кількістю згадувань із застосуванням лінійної регресії;
3) за кількістю покупок із застосуванням економетричних моделей.



а *б*

Рис. 5. Діаграми співвідношення кількості відгуків користувачів:
а) без застосування вагових коефіцієнтів, б) із застосуванням вагових коефіцієнтів

У роботі пропонується ранжувати об’єкти з врахуванням вагових коефіцієнтів, оскільки їхнє застосування уможливить отримати релевантніший, точніший та достовірніший результат для отримання інтегрального показника об’єкта під час оцінювання опінії текстової інформації.

3. Застосування лінгвістичних змінних

З робіт [3–11] слідує, що значення емоційного забарвлення $SO_{j,k,i,l}$, як правило, характеризується лише трьома значеннями (+1 – позитивне висловлювання, -1 – негативне висловлювання, 0 – нейтральне висловлювання, хоча можуть бути й інші варіації, напр. від -5 до 5), що своєю чергою, не повною мірою відображає опінію користувача щодо певного об’єкта.

У роботі пропонується описати $SO_{j,k,i,l}$ відповідно до (2) у вигляді лінгвістичної змінної, що дасть змогу представити емоційне забарвлення висловлювань за допомогою нечітких значень відповідно до її терм-множини.

Таким чином $SO_{j,k,i,l}$ може бути подане у вигляді набору [16]:

$$SO_{j,k,i,l} = (b, P, X, G, M), \tag{8}$$

де: b – найменування лінгвістичної змінної; P – множина її значень (терм-множина), що являють собою імена нечітких змінних, областю визначення, кожної з яких є множина X; G – синтаксична процедура, що дає змогу оперувати елементами терм-множини T, зокрема, генерувати нові терми (значення); M – семантична процедура, що дає змогу перетворити кожне нове значення лінгвістичної змінної, утвореною процедурою G, у нечітку змінну, тобто сформувати відповідну нечітку множину.

Таким чином власник опінії H_i може виступати експертом, а емоційне забарвлення $SO_{j,k,i,l}$ можна описати у вигляді лінгвістичної змінної (8).

3.1. Фазифікація

Нехай експерт визначає роздільну здатність фотокамери за допомогою поняття «непогана», також можуть бути поняття «хороша», «погана», «задовільна», при цьому мінімальна роздільна здатність $\min F_{j,k}$ дорівнюватиме 0 МПікселів, а максимальна $\max F_{j,k}$ – 15 МПікселів (у табл. 3 наведено інші мінімальні та максимальні значення, що можуть приймати компоненти об’єкта та зазначено, в яких одиницях вони вимірюються).

Формалізацію висловлювання експерта пропонується здійснювати за допомогою лінгвістичної змінної (4), де b – роздільна здатність; P – {«погана», «задовільна», «хороша»}; X – [0, 15]; G – процедура утворення нових термів за допомогою зв'язувань «і», «або» і модифікаторів типу «дуже», «не», «злегка» та ін. Наприклад, «мала або непогана роздільна здатність», «дуже мала роздільна здатність» тощо; M – процедура завдання на X = [0, 15] нечітких підмножин A_1 = «погана роздільна здатність», A_2 = «задовільна роздільна здатність», A_3 = «хороша роздільна здатність», а також нечітких множин для термів з $G(T)$ відповідно до правил трансляції нечітких зв'язувань і модифікаторів «і», «або», «не», «дуже», «злегка» та інших операцій над нечіткими множинами.

Цю процедуру можуть виконувати як експерти, так і користувачі, безпосередньо під час коментування, при цьому їм потрібно буде ввести декілька уточнюючих відповідей, наприклад вказати, яка, на їхню думку, «хороша» роздільна здатність фотокамери в числовому значенні від 0 до 15 МПікселів.

Таблиця 3

Мінімальні та максимальні значення компонент об’єкта та їхня розмірність

Характеристики				
	батарея, к-ть фото	**зображення, Мпікселі**	**зум, кратність**	**ціна, грн**
мін	0	0	0	0
макс	1400	15	25	6800

Разом із розглянутими вище базовими значеннями лінгвістичної змінної «роздільна здатність» (P = {«погана роздільна здатність», «задовільна роздільна здатність», «хороша роздільна здатність»}) існують можливі значення, що залежать від області визначення X. У цьому випадку значення лінгвістичної змінної «роздільна здатність» можуть бути визначені як «близько 10 МПікселів», «близько 5 МПікселів», «близько 7 МПікселів», тобто у вигляді нечітких чисел.

Отже, ми зможемо провести фазифікацію та побудувати функції належності лінгвістичної змінної (8).

3.2. Дефазифікація

У теорії нечітких множин процедура дефазифікації аналогічна знаходженню моментів (математичного сподівання, моди, медіани) розподілів випадкових величин у теорії ймовірності. Найпростішим способом виконання процедури дефазифікації є вибір чіткого числа, відповідного максимуму функції приналежності. Однак придатність цього способу обмежується лише одноекстремальними функціями належності. Для багатоекстремальних функцій належності використовують такі методи дефазифікації:

1) Centroid - центр ваги;
2) Bisector – медіана;
3) LOM (Largest Of Maximums) – найбільший з максимумів;
4) SOM (Smallest Of Maximums) – найменший з максимумів;
5) Mom (Mean Of Maximums) – центр максимумів.

У разі дискретної універсальної множини дефазифікація нечіткої множини за методом центра ваги здійснюється за такою формулою:

$$a = \frac{\sum_{q=1}^{Q} u_q \cdot \mu_A(u_q)}{\sum_{q=1}^{Q} \mu_A(u_q)}, \tag{9}$$

де $\mu_A(u_q)$ – ступінь належності терму до нечіткої підмножини; u_q – числове значення терму, $q = \overline{1,Q}$, Q – кількість значень терм-множини P.

4. Розробка алгоритму оцінювання опінії об'єкта

Оскільки об'єкт складається з компонент, а для опису більшості компонент є можливість застосування лінгвістичної змінної, то процедура формування інтегрального показника об'єкта виглядатиме так.

1. Аналіз відгуків користувачів. Побудова графа об'єкта. Для цього етапу потрібно мати сформований лінгвістичний ресурс (рис. 1).

2. Введення лінгвістичних змінних пооб'єктно.

Функції належності можна отримувати двома способами:

- безпосередньо від експертів;
- безпосередньо від користувачів, що пишуть коментарі, шляхом введення декількох допоміжних запитань під час написання коментаря (наприклад, користувач вибиратиме від 0 до 6800 «доступну ціну»).

3. Дефазифікація лінгвістичної змінної за методом центра ваги проводиться згідно з формулою (9).

Отже, ми отримуємо числове значення емоційного забарвлення висловлювань, котре залежить від терм-множини лінгвістичної змінної та може приймати значення не лише +1, 0 чи -1, а прийматиме значення, яке буде в діапазоні, зазначеному в лінгвістичній змінній.

4. Об'єднання дефазифікованих лінгвістичних змінних у межах одного компоненту об'єкта та зведення їх до середнього значення (табл. 4, 3-й стовпець).

5. Нормування середніх значень лінгвістичних змінних (табл. 4, 4-й стовпець), при цьому значення зводяться до одного діапазону від 0 до 1 відповідно до їхніх максимальних та мінімальних значень лінгвістичних змінних компонент об'єкта згідно з табл. 3.

6. Домноження нормованих значень лінгвістичних змінних на ваговий коефіцієнт компоненти (табл. 4, 5-й стовпець).

7. Підсумовування нормованих значень лінгвістичних змінних, домножених на вагові коефіцієнти для отримання інтегрального показника об'єкта згідно з формулою

$$I_j = \sum_{k=1}^{K_j} \left(F_{j,k} \cdot W_{j,k}\right) \cdot W_i \cdot W_l, \tag{10}$$

де I_j – інтегральний показник *j*-го об'єкта; W_i – вага власника опінії H_i, W_l – вага часової мітки T_l, $W_{j,k}$ – вага компоненти $F_{j,k}$, що обчислюється за формулою (3), $F_{j,k}$ – числове значення емоційного забарвлення компоненти об'єкта O_j, що обчислюється за формулою

$$F_{j,k} = \frac{\sum_{i=1}^{I_{j,k}} \sum_{l=1}^{L_{j,k,i}} SO_{j,k,i,l}}{n_{в.k} \cdot \max F_{j,k}}, \tag{11}$$

де $SO_{j,k,i,l}$ – числове значення емоційного забарвлення висловлювання, що обчислюється за формулою (9).

Порівняння двох фотокамер згідно з наведеним вище алгоритмом зображено у табл. 4. Нехай після дефазифазифікації та зведення лінгвістичних змінних до середнього значення є числові еквіваленти опінії таких компонент, як «батарея», «якість зображення», «оптичний зум» та «ціна».

Таблиця 4

Співвідношення кількості відгуків користувачів із застосуванням вагових коефіцієнтів

1	2	3	4	5
Об'єкт	Компонент	Лінгв. зм., число	Нормув. ЛЗ	ЛЗ * Коеф.
фото 1				
	батарея	800	0,571428571	9,713149673
	зображення	10	0,666666667	15,59531699
	зум	15	0,6	24,21471173
	ціна	3500	0,514705882	9,908685144
	Загальне число		2,35280112	59,43186353
фото 2				
	батарея	1100	0,785714286	13,3555808
	зображення	12	0,8	18,71438038
	зум	10	0,4	16,14314115
	ціна	3800	0,558823529	10,75800101
	Загальне число		2,544537815	58,97110335

Після нормування лінгвістичних змінних та зведення їх до інтегрального показника, бачимо, що фотокамера 2 є кращою, проте після застосування вагових коефіцієнтів – фотокамера 1 буде кращою від фотокамери 2, тому що в неї рейтинг більший згідно з числовими значеннями емоційного забарвлення висловлювань користувачів.

Висновки

1. Удосконалено математичну модель подання опінії в текстових даних інтернет-ресурсів, що уможливило врахування вагових коефіцієнтів та лінгвістичних змінних.

2. Обґрунтовано узагальнений алгоритм оцінювання опінії об'єкта та зведення числових значень емоційного забарвлення висловлювань до єдиного інтегрального показника з використанням лінгвістичних змінних та вагових коефіцієнтів, що дало змогу отримати достовірніший та адекватніший результат.

У перспективі автори здійснюватимуть роботу над розробленням прототипу системи підтримки прийняття рішень під час оцінювання опінії текстової інформації із використанням одержаних результатів.

1. Електронне видання "Словники України" – 3.2 [Електронний ресурс] : за даними "Орфографічного словника української мови", 7-е видання. – К.: Довіра, 2007. Режим доступу: http://lcorp.ulif.org.ua/dictua. 2. Dongjoo Lee, Ok-ran Jeong, Sang-goo Lee. Opinion Mining of Customer Feedback Data on the Web / Lee Dongjoo, Jeong Ok-ran, Lee Sang-goo // Proceedings of the 2nd international conference on Ubiquitous information management and communication. – New York, NY, USA, 2008. – Pp. 230-235. 3. Bing Liu. Opinion observer: Analyzing and comparing opinions on the web / Liu Bing // Proceedings of the 14th international conference on World Wide Web. – In WWW, 2005. – Pp. 342-351. 4. Xiaowen Ding, Bing Liu, Philip S. Yu. A holistic lexicon-based approach to opinion mining / Ding Xiaowen, Liu Bing, S. Yu. Philip // Proceedings of the international conference on Web search and web data mining. – New York, NY, USA, 2008. – Pp. 231-240. 5. Ana-Maria Popescu, Oren Etzioni. Extracting product features and opinions from reviews / Popescu Ana-Maria, Etzioni Oren // Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. – Stroudsburg, PA, USA, 2005. – Pp. 339-346. 6. Ana-Maria Popescu, Bao Nguyen, Oren Etzioni. OPINE: Extracting Product Features and Opinions from Reviews / Popescu Ana-Maria, Nguyen Bao, Etzioni Oren // Proceedings of HLT/EMNLP 2005 Demonstration Abstracts. – Vancouver, 2005. – Pp. 32-33. 7. Sasha Blair-Goldensohn , Tyler Neylon , Kerry Hannan , George A. Reis , Ryan Mcdonald , Jeff Reynar. Building a sentiment summarizer for local service reviews / Blair-Goldensohn Sasha, Neylon Tyler, Hannan Kerry, A. Reis George, Mcdonald Ryan, Reynar Jeff // WWW Workshop on NLP in the Information Explosion Era (NLPIX). – Beijing, China, 2008. 8. Wei Jin, Hung Hay Ho, Rohini K. Srihari. OpinionMiner: a novel machine learning system for web opinion mining and extraction / Jin Wei, Hay Ho Hung, K. Rohini // Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. – New York, NY, USA, 2009. – Pp. 1195-1204. 9. Dave K. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification in Product Reviews / K. Dave, S. Lawrence, D. Pennock // Proceedings of ACM WWW2003. – Budapest, 2003. – Pp. 519–528. 10. Google Products [Електронний ресурс]. – Режим доступу: http://www.google.com/shopping 11. Bing Shopping [Електронний ресурс]. – Режим доступу: http://www.bing.com/cashback 12. Bing L. Web Data Mining. Exploring Hyperlinks, Contents and Usage Data, Second Edition : / Liu Bing. – Springer, 2011. – 622 p. 13. Suke Li, Zhi Guan, Liyong Tang, Zhong Chen. Exploiting consumer reviews for product feature ranking / Li Suke, Guan Zhi, Tang Liyong, Chen Zhong // Proceedings of the SWSM'11. – Beijing, China., July 28, 2011. 14. Narisa Zhao, Yuan Li. Research on multi-affective fuzzy computing and product competitive advantage of online product reviews / Zhao Narisa, Li Yuan // Advanced Materials Research Vol. 186 . – Trans Tech Publications, Switzerland, 2011. – Pp. 464-468. 15 Nikolay Archak, Anindya Ghose, Panagiotis G. Ipeirotis. Show me the Money! Deriving the Pricing Power of Product Features by Mining Consumer Reviews / Archak Nikolay, Ghose Anindya, G. Ipeirotis Panagiotis // Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. – New York, NY, USA, 2007. – Pp. 56-65. 16. Штовба С.Д. Введение в теорию нечетких множеств и нечеткую логику. / С.Д. Штовба. – Винница: Изд-во винницкого госуд. техн. ун-та, 2001. – 198 с.