PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 427

Taksonomia 27
Klasyfikacja i analiza danych – teoria i zastosowania

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania", zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 427 • 2016

Taksonomia 27 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski
Uniwersytet Ekonomiczny w Krakowie
e-mails: {lulap; wojcikk; tuchowsj}@uek.krakow.pl

ANALIZA WYDŹWIĘKU POLSKOJĘZYCZNYCH OPINII KONSUMENCKICH UKIERUNKOWANYCH NA CECHY PRODUKTU[1]

FEATURE-BASED SENTIMENT ANALYSIS OF OPINIONS IN POLISH

DOI: 10.15611/pn.2016.427.16

Streszczenie: Opinie konsumenckie są szczególnym rodzajem dokumentów ze względu na swoją zawartość – ich głównym celem nie jest przekazanie obiektywnej informacji, ale subiektywnego nastawienia autora do produktu lub jego elementów. Przedstawione w opinii nastawienie opiniodawcy nazywane jest wydźwiękiem. Opinia może odnosić się do produktu w całości, lub też do jego elementów. Celem pracy jest zaprezentowanie autorskiej metody automatycznej oceny wydźwięku opinii konsumenckich ukierunkowanych na cechy produktu. Zadanie to jest realizowane poprzez analizę słów występujących w bezpośrednim sąsiedztwie miejsca w tekście zawierającego odniesienie do zidentyfikowanych cech produktu. Wyznaczone na podstawie tekstu oceny wyróżnionych elementów produktu mogą zostać przypisane do właściwych elementów drzewa opisu produktu, a następnie przetwarzane w sposób właściwy dla przyjętego celu analizy.

Słowa kluczowe: *text-mining*, *opinion mining*, *sentiment analysis*, ontologia, odległość semantyczna, analiza ukierunkowana na cechy produktu.

Summary: Consumer reviews are a special type of text documents due to their content – their main purpose is not to provide objective information, but to show a subjective attitude of its authors to the product or its components. The attitude presented in the opinion is called overtones. Opinion may refer to a product as a whole or its components. The aim of the paper is to present the authors' method for automatic evaluation of features-concentrated opinions overtones. This task is realized by analyzing the words in the direct neighbourhood of the product's characteristics found in the text. Sentiments of distinguished product's components identified on the basis of opinion can be assigned to the appropriate parts of the product description tree and then processed according to the purpose of analysis.

Keywords: text-mining, opinion mining, sentiment analysis, ontology, semantic similarity, feature-based analysis.

[1] Publikacja dofinansowana ze środków przyznanych Wydziałowi Zarządzania Uniwersytetu Ekonomicznego w Krakowie w ramach dotacji przyznanej na utrzymanie potencjału badawczego oraz w ramach dotacji na finansowanie zadań służących rozwojowi młodych naukowców.

1. Wstęp

Opinie konsumenckie są szczególnym rodzajem dokumentów ze względu na swoją zawartość – ich głównym celem jest przekazanie nie obiektywnej informacji, ale subiektywnego nastawienia autora do produktu lub jego elementów. Przedstawione w opinii nastawienie opiniodawcy nazywane jest wydźwiękiem. W najprostszej sytuacji wyróżnia się opinie o wydźwięku pozytywnym lub negatywnym. W części badań uwzględnia się również opinie neutralne. Opinia może odnosić się do produktu w całości lub też do jego elementów. W tym drugim przypadku ogólne nastawienie autora do produktu powstaje poprzez zagregowanie ocen cząstkowych, dotyczących poszczególnych elementów. Dokonując agregacji, należy uwzględnić znaczenie poszczególnych składowych w całości produktu.

Celem artykułu jest prezentacja autorskiej metody automatycznej oceny wydźwięku opinii konsumenckich ukierunkowanych na cechy produktu. Zakłada się, że struktura produktu opisana jest w postaci ontologii przyjmującej postać drzewa. W trakcie pierwszego kroku analizy znajdujące się w tekście opinii wyrazy odnoszące się do ocenianego produktu wiązane są z właściwymi elementami ontologii, a następnie analizowany jest wydźwięk (nacechowanie) słów znajdujących się w bezpośrednim sąsiedztwie zidentyfikowanych terminów. Na tej podstawie dokonywana jest ocena wymienionej w tekście cechy produktu. Drzewiasta struktura ontologii pozwala na agregowanie tak uzyskanych informacji i uzyskanie ocen o ustalonym przez badacza stopniu szczegółowości.

2. Automatyczna analiza opinii konsumenckich

Automatyczna analiza opinii konsumenckich (*sentiment analysis*, *opinion mining*) to ogół działań mających na celu zautomatyzowanie procesu wyszukiwania, ekstrakcji i analizy danych pochodzących ze specyficznych tekstów, jakimi są opinie użytkowników. Są to działania z pogranicza przetwarzania języka naturalnego (*Natural Language Processing* – NLP), lingwistyki komputerowej (*computational linguistics*) oraz eksploracyjnej analizy tekstu (*text mining*). Jej celem jest określenie nastawienia autora wypowiedzi do jej przedmiotu.

W literaturze światowej zagadnienie *analizy* lub *eksploracji opinii* (określane jako *opinion mining*) pojawiło się w 2003 r. [Dave, Lawrence, Pennock 2003]. Stosowane jest również pojęcie *analiza wydźwięku* (*sentiment analysis*), które pojawiło się w pracach [Das, Chen 2001; Turney 2002; Pang, Lee, Vaithyanathan 2002; Nasukawa, Yi 2003]. W wielu przypadkach wspomniane terminy stosowane są zamiennie. Wydaje się jednak, że analiza wydźwięku jest zagadnieniem znacznie węższym i stanowi jeden z problemów rozpatrywanych w ramach automatycznej analizy opinii konsumenckich. Badanie wydźwięku porównać można do zada-

nia klasyfikacji wzorcowej polegającego na przypisaniu opinii do jednej z grup obejmujących wypowiedzi pozytywne, negatywne lub neutralne. Zadanie to określane jest również jako badanie polaryzacji (*polarity*) opinii.

Opinie można podzielić na grupy według ich formatu [Liu 2007]:

1) zalety i wady, oraz podsumowanie,
2) zalety i wady,
3) dowolny.

Analiza opinii konsumenckich jest jednym z najbardziej wymagających problemów rozpatrywanych na gruncie automatycznej analizy języka naturalnego. Do najważniejszych czynników utrudniających analizę należy zaliczyć [Cambria i in. 2013; Cambria, Hussain 2012; Pang, Lee 2008]):

- cel badań – jakim jest analiza emocji lub nastawienia – nie zawsze jest w sposób bezpośredni przedstawiony w tekście,
- sarkastyczny lub ironiczny charakter wypowiedzi,
- błędy ortograficzne i stylistyczne w opiniach umieszczanych w serwisach społecznościowych,
- potrzeba przeprowadzania analiza nawiązań (wielokrotne odwołań do obiektów bezpośrednio nieprzywoływanych w tekście określanych jako odwołania koreferencyjne),
- badanie znaczenia powtórzeń zwrotów lub słów (anafor),
- konieczność właściwej interpretacji wyrażeń negujących,
- problem identyfikacji nazw własnych,
- stosowanie porównań,
- wieloznaczność słów i dłuższych wypowiedzi (np. stwierdzenie „idź na spacer" samo w sobie nie ma negatywnego zabarwienia, ale jeśli pojawi się w recenzji filmu to nabiera ujemnego wydźwięku).

2.1. Podejścia do automatycznej analizy opinii konsumentów

W ramach automatycznej analizy opinii konsumenckich wyróżnić można trzy rodzaje działań, takie jak [Liu 2007]:

- identyfikacja ogólnego charakteru i klasyfikacja opinii,
- analiza ukierunkowana na cechy produktu,
- analiza porównawcza produktów.

Pierwsze z wymienionych zadań ma na celu określenie ogólnego nastawienia autora do opiniowanego produktu i zaklasyfikowanie opinii jako pozytywnej, negatywnej lub neutralnej. Tak opisany problem określany jest również jako identyfikacja polaryzacji, nacechowania lub wydźwięku opinii lub analiza sentymentu.

Celem analizy ukierunkowanej na cechy produktu jest wydobycie z tekstu opinii fragmentów dotyczących poszczególnych cech produktu i określenie stosunku autora tekstu do każdej z nich. W niektórych pracach analiza ukierunkowana na

cechy produktu opisywana jest jako głębszy poziom analizy nacechowania opinii [Liu 2010]. Zebrane w ten sposób informacje cząstkowe mogą zostać w trakcie kolejnych etapów analizy zagregowane w celu wyznaczenia ogólnego nastawienia opiniującego do produktu. Wykonanie takiej operacji wymaga wiedzy dotyczącej struktury produktu będącego przedmiotem opinii determinującej sposób wykonania agregacji.

Szczególnym przypadkiem analizy opinii konsumenckich jest analiza porównawcza mająca na celu dokonanie oceny dwóch produktów lub ich cech w kategoriach *lepszy* lub *gorszy*.

Wyróżnić można cztery text miningowe podejścia do automatycznej analizy opinii konsumentów [Lula, Wójcik 2011].

1. Podejście oparte na słowach (*word-based approach*). Stosując podejście oparte na słowach, dokonuje się podziału tekstu na poszczególne słowa. W kolejnym kroku określa się nacechowanie każdego ze słów. Do wykonania tej operacji wykorzystywane są słowniki sentymentu, które mogą dla znajdujących w nich słów zawierać informację o nacechowaniu w postaci etykiety pozytywny lub negatywny lub też zawierać informację o nacechowaniu w postaci wartości numerycznej. Słowom niewystępującym w słowniku przypisuje się zwykle nacechowanie neutralne. Słowniki sentymentu tworzone są głównie dla języka angielskiego (za przykład może służyć SentiWordNet [Esuli, Sebastiani 2006]). Ocena podejścia opartego na słowach wykorzystującego chmury tagów zbudowane z wad i zalet będących częścią opinii przedstawiona została w pracy [Wójcik, Tuchowski 2013].

2. Podejście oparte na wzorcach (*pattern-based approach*). Jedną z najistotniejszych wad podejścia opartego na słowach jest to, że w trakcie analizy nie uwzględnia się informacji o kolejności słów, lecz każdy wyraz rozpatruje się niezależnie od kontekstu, w jakim wystąpił. Podejście oparte na wzorcach ma na celu przynajmniej w sposób częściowy rozwiązać ten problem. Zakłada ono, że algorytm analizujący tekst opinii potrafi rozpoznać frazy lub specyficzne struktury gramatyczne i uwzględnić te informacje przy określaniu nacechowania. Przykładem tego typu podejścia może być zastosowanie reguł aplikacji Spejd opartych na mechanizmie wyrażeń regularnych w sposób ułatwiający tworzenie opisów sekwencji słów. Omówienie i ocenę podejścia opartego na wzorcach zamieszczono w pracy [Wójcik, Tuchowski 2015].

3. Podejście oparte na ontologii (*ontology-based approach*). Podejście oparte na ontologii zakłada, że w trakcie analizy opinii dostępna jest wiedza na temat opisywanego produktu (elementów składowych i sposobu ich powiązania, cech opisujących najistotniejsze charakterystyki czy też realizowanych funkcji). Sformalizowany opis wiedzy dotyczącej produktu przybiera postać ontologii. W filozofii pojęcie ontologii odnosi się do nauki o bycie. Za prekursorów rozważań z tego zakresu uznaje się Parmenidesa z Etei, Platona i Arystotelesa. Natomiast samo

pojęcie *ontologia* zaistniało się w nauce w XVII w. Po raz pierwszy pojawiło się w pracy Jacobusa Lorhardusa w 1606 r. [Lorhardus 1606]. W niniejszej pracy termin ten używany jest w znaczeniu podanym przez T.R. Grubera, który określa ontologię jako *jednoznacznie zdefiniowaną specyfikację konceptualizacji pewnego obszaru wiedzy* [Gruber 1993]. Ontologia traktowana jest jako jedna z podstawowych metod reprezentacji wiedzy o charakterze dziedzinowym. Jej stosowanie wymaga zdefiniowania klas będących opisami typy obiektów występujących w rozpatrywanym wycinku rzeczywistości, powiązań pomiędzy klasami oraz obiektów (o strukturze opisanej przez zdefiniowane wcześniej klasy) reprezentujących elementy składające się na rozpatrywany fragment rzeczywistości. Nie ulega wątpliwości, że najistotniejszym pojęciem związanym z ontologiczną reprezentacją wiedzy jest klasa. Klasa reprezentuje dany rodzaj (typ) obiektów. Definiując klasę należy wskazać jej atrybuty (czyli cechy charakterystyczne) oraz charakterystyczne dla niej zachowania (czynności charakterystyczne dla danego typu obiektów lub stany, w których może się on znaleźć). Klasy są zwykle ze sobą powiązane, a istniejące związki najczęściej dotyczą relacji hierarchicznych i wskazują na klasę bazową i klasy potomne, będące szczególnymi przypadkami (uszczegółowieniem) klasy podstawowej. Obszerniejsze omówienie metod analizy opinii wykorzystujących ontologie zawarte jest w pracy [Wójcik, Tuchowski 2014].

4. Podejście wykorzystujące uczenie maszynowe (*machine learning approach*). Podejście wykorzystujące uczenie maszynowe zakłada, że wiedza na temat badanego zjawiska pozyskiwana jest w wyniku analizy i uogólniania informacji opisujących kolejne jego realizacje. Najistotniejszą zaletą takiego podejścia jest możliwość zastąpienia wiedzy eksperckiej pozyskiwanej od człowieka wiedzą pozyskaną w wyniku eksploracji danych. Należy jednak również pamiętać o słabych stronach uczenia maszynowego: konieczności zgromadzenia dużej liczby przypadków uczących, złożoności obliczeniowej i konieczności doboru właściwego algorytmu pozyskiwania wiedzy.

W pracy [Cambria i in. 2013] znaleźć można podobną klasyfikację podejść do automatycznej analizy opinii konsumentów. Koncentrując się na analizie ukierunkowanej na cechy produktu, można w niej zastosować trzy pierwsze podejścia.

3. Propozycja autorskiej metody analizy nacechowania opinii ukierunkowanych na cechy produktu

Uprzednio przeprowadzone badania pokazały zalety oraz wady różnych podejść do automatycznej analizy wydźwięku polskojęzycznych opinii konsumentów ukierunkowanych na cechy produktu. Wyniki tych analiz skłoniły autorów do opracowania modelu, który będzie łączył w sobie różne podejścia, wykorzystując ich zalety dla większej efektywności podejścia mieszanego. W badaniach wykorzystano materiał

badawczy zebrany na potrzeby wcześniejszych eksperymentów, aby wyniki były porównywalne.

Celami prac były:

1) zaproponowanie algorytmu pozwalającego w sposób automatyczny określić nastawienie konsumentów do poszczególnych cech ocenianego produktu oraz do produktu jako całości. Przyjęto, że algorytm powinien wykorzystywać wiedzę dziedzinową określoną w postaci ontologii, ale jednocześnie nie powinien wymagać manualnego znakowania fragmentów tekstu odpowiadającym poszczególnym pojęciom zdefiniowanym w ontologii;

2) przeprowadzenie przykładowych badań przy wykorzystaniu zaproponowanej metody.

3.1. Zbiór opinii

W badaniach empirycznych wykorzystano 737 opinii w formie pierwszej (wady, zalety, posumowanie). Opinie pochodziły z serwisu Ceneo.pl[2] i dotyczyły smartfonów Samsung Galaxy S II, S III, S4 oraz S5. Do każdej opinii dołączona była ocena punktowa w postaci gwiazdek w przedziale [0,5; 5] z krokiem 0,5.

Opinie z serwisu internetowego zostały pobrane do bazy. Następnie z bazy danych zostały one wyeksportowane do plików tekstowych. Każda opinia została zapisana w osobnym pliku tekstowym.

3.2. Słowniki

W badaniach empirycznych wykorzystano słowniki wyrazów pozytywnych i negatywnych utworzone w trakcie wcześniejszych badań [Wójcik, Tuchowski 2015]. Każdy ze słowników zawiera ok. 200 słów. Wyrazy w słownikach występują w wersji podstawowej (po redukcji do rdzenia). Słowniki przygotowano w dwóch wersjach:

- sentyment o wartości 1 dla słów pozytywnych i –1 dla negatywnych,
- sentyment dodatni dla słów pozytywnych i ujemny dla negatywnych, wartość zależy od siły nacechowania, wartości całkowite od –10 do 10 bez 0.

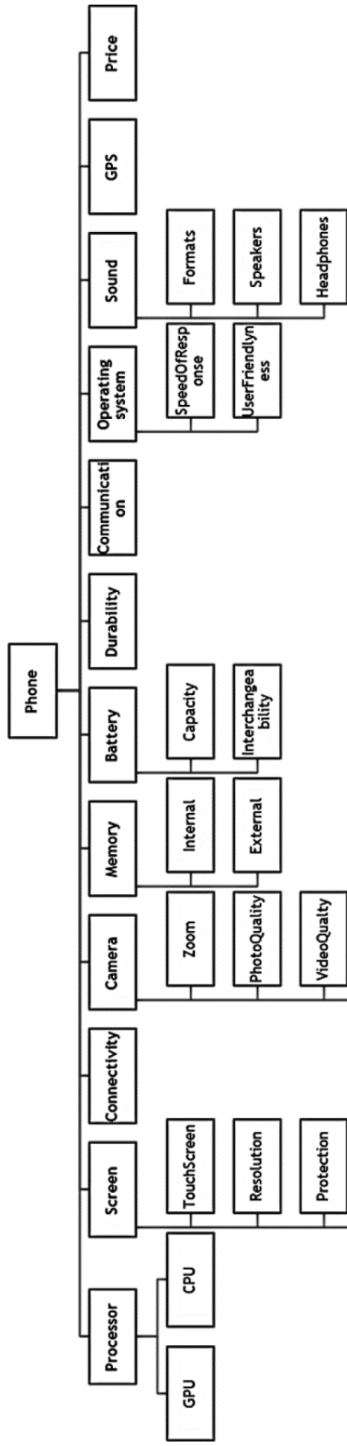
W słownikach pominięto problematyczne słowa, takie jak: wysoki/niski, szybko/wolno, długo/krótko. Słowa te w zależności od kontekstu będą miały przeciwne nacechowanie.

3.3. Ontologia

Na potrzeby badań utworzona została ontologia smartfonu.

Przyjęto, że struktura telefonu opisana jest przez drzewo przedstawione na rys. 1.

[2] Dostęp 3.09.2014 r.



Rys. 1. Schemat ontologii wykorzystanej w badaniach

Źródło: opracowanie własne.

3.4. Algorytm

Proponowany algorytm wykorzystuje:
- wiedzę dziedzinową opisującą strukturę opisywanego produktu w postaci ontologii,
- słownik nacechowania.

Danymi przetwarzanymi przez algorytm są teksty zawierające opinie konsumenckie.

Proponowany algorytm składa się z następujących kroków.

Krok 1. Scalenie wszystkich analizowanych opinii w jeden dokument.

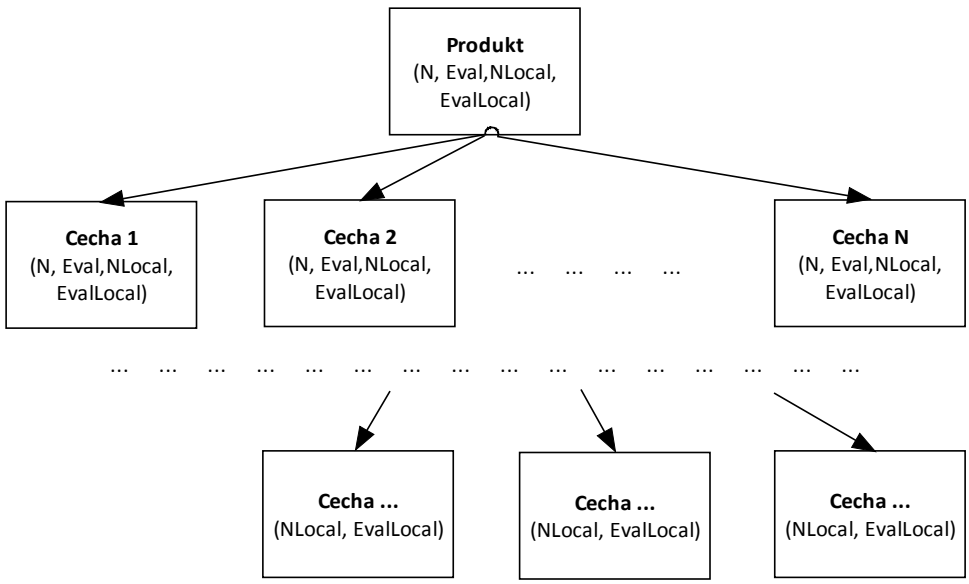
Krok 2. Podział dokumentu zawierającego opinie na fragmenty. Podział dokonywany jest w miejscu wystąpienia separatora. W realizowanym eksperymencie przyjęto, że rolę separatora spełnia znak przejście do nowego akapitu oraz kropka kończąca zdania.

Krok 3. Wstępne przetworzenie tekstu zawierającego opinie. Operacja ta obejmuje:
- przekształcenie wyrazów do formy podstawowej,
- zamianę wszystkich liter na małe,
- usunięcie znaków interpunkcyjnych i wyrazów o długości nie przekraczającej dwóch znaków,
- usunięcie wyrazów nieistotnych (na podstawie stoplisty),
- powiązanie pojęć pochodzących z ontologii z wyrazami wchodzącymi w skład badanego tekstu i odnoszącymi się do elementów telefonu (np. takie określenia, jak „głośniki", „głośnik" czy „głośniczki" powiązano z elementem „Speakers").

Krok 4. Utworzenie w obrębie każdego fragmentu tekstu (utworzonego w drugim kroku algorytmu) podłańcuchów o określonej długości, wykorzystując podejście oparte na przesuwnym oknie. W trakcie realizacji obliczeń zastosowano okno o szerokości trzech wyrazów, co doprowadziło do utworzenia trigramów.

Krok 5. Na podstawie podłańcuchów utworzonych w poprzednim kroku tworzona jest macierz współwystępowania. Jest to macierz, której kolumny odpowiadają cechom produktu uwzględnionym w przyjętej ontologii opisującej opiniowany produkt, wiersze zaś odnoszą się do poszczególnych określeń występujących w słowniku nacechowania. Wartości macierzy wskazują ile razy w wyodrębnionych podłańcuchach występują łącznie odpowiadająca kolumnie cecha produktu i odpowiadające wierszowi określenie nacechowania.

Krok 6. Realizacja obliczeń.



Rys. 2. Drzewo prezentujące schemat obliczeń

Źródło: opracowanie własne.

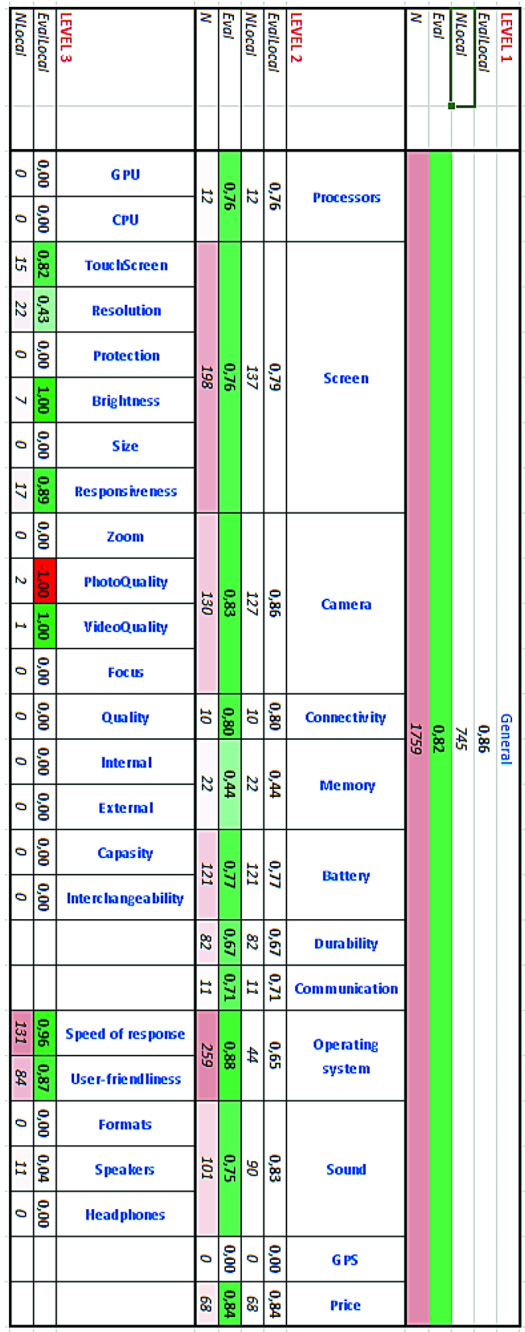
Obliczenia wykonywane są w kolejności określonej przez strukturę drzewa opisującego oceniany produkt (rys. 2).

W pierwszej kolejności następuje przejście po liściach drzewa. Dla każdego z nich wyznaczane są dwie wartości:

- *NLocal* – zarejestrowana w macierzy współwystępowania liczba odwołań do cechy produktu reprezentowanej przez dany element ontologii,
- *EvalLocal* – wyznaczona na podstawie opinii ocena *i*-tej cechy produktu będąca wartością z przedziału [–1, +1]. Wartość ta obliczana jest według wzoru:

$$EvalLocal_i = \frac{\sum_{k=1}^{N_{\text{pos}}} w_k^{\text{pos}} l_k^{pos} - \sum_{k=1}^{N_{\text{neg}}} w_k^{\text{neg}} l_k^{\text{neg}}}{\max\left(\sum_{k=1}^{N_{\text{pos}}} w_k^{\text{pos}} l_k^{\text{pos}}, \sum_{k=1}^{N_{\text{neg}}} w_k^{\text{neg}} l_k^{\text{neg}} \right)}, \tag{1}$$

gdzie: N_{pos}, N_{neg} – liczby terminów występujących w słowniku terminów o nacechowaniu pozytywnym i słowniku wyrazów o nacechowaniu negatywnym; w_k^{pos}, w_k^{neg} – to wartości nacechowania odpowiadające *k*-temu terminowi ze słownika nacechowania; l_k^{pos}, l_k^{neg} – to liczba odwołań do *k*-tego terminu ze słownika nacechowania (pozytywnego lub negatywnego) występującego łącznie w podłańcuchach (krok 4) z odwołaniem do *i*-tej cechy produktu.



Rys. 3. Wyniki obliczeń wydźwięku opinii o smartfonach Samsung Galaxy S ukierunkowanych na cechy produktu

Źródło: opracowanie własne.

Następnie obliczenia prowadzone są dla wszystkich pozostałych węzłów. Oprócz wartości NLocal i EvalLocal obliczane są:

- *N* – zarejestrowana w macierzy współwystępowania liczba odwołań do cechy produktu reprezentowanej przez rozpatrywany element ontologii i wszystkie elementy podrzędne (będące potomkami bezpośrednimi lub pośrednimi danego elementu),
- *Eval* – należąca do przedziału [–1, +1] ocena cechy produktu reprezentowanej przez rozpatrywany element ontologii liczona jako średnia ważona z wartości *EvalLocal* obliczonej dla bieżącego elementu i wartości *EvalLocal* wyznaczonych dla wszystkich bezpośrednich potomków elementu bieżącego. W charakterze wag wykorzystywane są wartości *NLocal*.

Przedstawiony schemat postępowania powtarzany jest dla wszystkich węzłów wewnętrznych, w tym również dla elementu będącego korzeniem drzewa. Obliczone miary określają nastawienie autorów opinii do rozpatrywanej cechy produktu (wszystkie węzły z wyjątkiem korzenia) lub do produktu jako całości (korzeń drzewa). Wyznaczone mierniki należą do przedziału [–1, +1]. Wartość ujemna wskazuje na ocenę negatywną, dodatnia zaś na pozytywną.

3.5. Wyniki analizy opinii dotyczących telefonów komórkowych Samsung Galaxy

Obliczenia prowadzono według przedstawionego powyżej schematu. Wyniki zostały zaprezentowane na rys. 3.

W celu ułatwienia interpretacji uzyskanych wyników oceny pozytywne zaznaczono kolorem zielonym, negatywne zaś – kolorem czerwonym. Stopień wypełnienia komórek zawierających informację o liczebności jest proporcjonalny do liczby opinii, a tym samym określa zaufanie do wartości określającej ocenę danego aspektu produktu.

4. Zakończenie

Na podstawie przeprowadzonych badań wydaje się, że podejście oparte na współwystępowaniu cech produktu i ich oceny pozwala zidentyfikować nastawienie użytkownika zarówno do produktu, jak i do jego składowych.

Mimo uproszczenia modelu poprzez nieuwzględnienie wielu aspektów lingwistycznych opinii, uzyskane wyniki mogą mieć wartościowy charakter. Mogą również posłużyć do modyfikacji ontologii poprzez wskazanie cech niekomentowanych przez użytkowników oraz tych komentowanych najczęściej.

Podejście wymaga dalszych badań związanych z właściwym doborem długości analizowanego podłańcucha tekstu i elementów rozdzielających fragmenty wypowiedzi.

Literatura

Cambria E., Hussain A. 2012, *Sentic Computing: Techniques, Tools, and Applications*, http://books.google.pl/books?id=8DPLZ8klJrkC (10.10.2015).

Cambria E., Schuller B., Xia Y., Havasi C. 2013 *New avenues in opinion mining and sentiment analysis*, IEEE Intelligent Systems, vol. 28, no. 2, s. 15–21, doi:10.1109/MIS.2013.30.

Das S., Chen M., 2001, *Yahoo! for {Amazon}: Extracting market sentiment from stock message boards*, [w:] *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA*), vol. 33, Bangkok.

Dave K., Lawrence S., Pennock D.M., 2003, *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*, [w:] *Proceedings of the 12th International Conference on World Wide Web*, ACM, New York, s. 519–528, doi:10.1145/775152.775226.

Esuli A., Sebastiani F., 2006, *SENTIWORDNET: A Publicly Available Lexical Resource*, [w:] *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, s. 417–422, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.7217.

Gruber T.R., 1993, *A translation approach to portable ontology specifications*, Knowledge Acquisition, vol. 5, no. 2, s. 199–220.

Liu B., 2007, *Web DataMining. Exploring Hyperlinks, Contents, and Usage Data*, Springer-Verlag, Heidelberg–Berlin .

Liu B., 2010, *Sentiment analysis and subjectivity*, [w:] N. Indurkhya, F. Damerau (red.), *Handbook of Natural Language Processing*, Chapman & Hall/CRC Machine Learning & Pattern Recognition Series, ,Chapman & Hall/CRC, t. 2, s. 627–666.

Lorhardus J., 1606, *Ogdoas Scholastica, continens Diagraphen Typicam artium: Grammatices (Latinae, Graecae), Logices, Rhetorices, Astronomices, Ethices, Physices, Metaphysices, seu Ontologia*, Apud Georgium Straub, Sangalii.

Lula P. Wójcik K., 2011, *Sentiment analysis of consumer opinions written in Polish*, Economics and Management, nr 16, s. 1286–1291.

Nasukawa T., Yi J., 2003, *Sentiment analysis: Capturing favorability using natural language processing*, [w:] *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP)*, ACM, New York, s. 70–77.

Pang B., Lee L., 2008, *Opinion mining and sentiment analysis*, Foundations and Trends in Information Retrieval, vol. 2, no ½, s. 1–135.

Pang B., Lee L., Vaithyanathan S., 2002, *Thumbs Up? Sentiment Classification Using Machine Learning Techniques*, *Proceedings of EMNLP*, s. 79–86.

Turney P., 2002, *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*, *Proceedings of the Association for Computational Linguistics (ACL)*, s. 417–424.

Wójcik K., Tuchowski J., 2013, *Sentiment analysis of opinions about hotels extracted from the Internet*, [w:] P. Lula, B. Mikuła, A. Jaki (red.), *Knowledge – Economy – Society. Global and Regional challenges of the 21st Century Economy*, Foundation of the Cracow University of Economics, Kraków, s. 755–771.

Wójcik K., Tuchowski J., 2014, *Ontology based approach to sentiment analysis*, [w:] P. Lula, T. Rojek (red.), *Knowledge – Economy – Society. Contemporary Tools of Organisational Resources Management*, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków, s 268–279.

Wójcik K., Tuchowski J., 2015, *Wykorzystanie metody opartej na wzorcach w automatycznej analizie opinii konsumenckich*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 385, Taksonomia 25: *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 314–324.