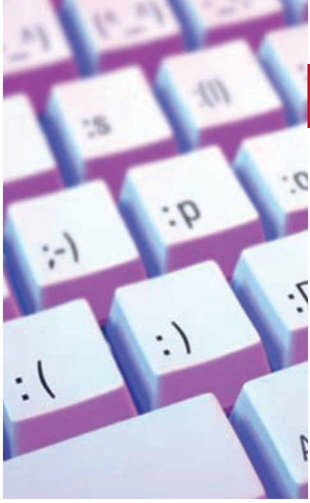

Editor: **Erik Cambria,** Nanyang Technological University, Singapore, cambria@ntu.edu.sg

Aspect-Based Extraction and Analysis of Affective Knowledge from Social Media Streams

Albert Weichselbraun, *University of Applied Sciences Chur and webLyzard*
Stefan Gindl, *MODUL University Vienna*
Fabian Fischer and Svitlana Vakulenko, *Vienna University of Economics and Business*
Arno Scharl, *MODUL University Vienna and webLyzard*

This article introduces an approach to analyze emotional values associated with brands and companies. Online media coverage about products and services typically refers to a wide range of aspects to which such emotional values apply. These aspects can include product features (such as a digital camera's maximum resolution), common applications (such as a smartphone used as a car navigation system), or perceptions in conjunction with a specific event (for example, as part of a sponsorship agreement). Our approach integrates affective and factual knowledge extraction to capture opinions related to specific aspects along multiple emotional dimensions. We use the automotive industry as a sample domain to demonstrate the proposed approach, given the large number of aspects that characterize its complex technical products.

Affective knowledge includes sentiment and other emotions expressed in a document, which are captured and evaluated by opinion-mining algorithms. Typically, such algorithms are based on machine learning, lexical methods, or a combination of both.[1] To identify entities and aspects, the presented system also extracts *factual knowledge* using a knowledge base built on data from linked data sources such as DBpedia and ConceptNet. This knowledge base holds information about products, including not only product characteristics but also corporate decision makers such as Martin Winterkorn, the former CEO of Volkswagen AG (www.dbpedia.org/page/Martin_Winterkorn).

The real-time social media streams used for the analysis originate from the Media Watch on Climate Change (www.ecoresearch.net/climate), a continuously updated knowledge repository on climate change and related environmental issues.[2] The system is based on the webLyzard Web intelligence platform (www.weblyzard.com), which extracts and visualizes knowledge from digital content streams to measure the impact of events and communication campaigns, independent of a specific domain. Adapted to the specific requirements of the Media Watch on Climate Change, the system collects, filters, and annotates documents from news media, social networking platforms, and the websites of Fortune 1000 companies and environmental organizations.[3]

Figure 1 shows the results of a sample query for the term "Volkswagen" in English-language news media published between July and December 2016. The screenshot reflects the significant media impact of the "Dieselgate" scandal (that is, manipulations to cheat official pollution tests), with most of the articles about Volks-wagen still focusing on this story. The event's dominance highlights the importance of aspect-centered approaches to opinion mining. Although the overall sentiment is negative, specific features such as *seat quality* or the *gearbox* receive positive feedback. Only a granular analysis that considers all relevant aspects can reveal such hidden knowledge, which is highly relevant for planning and evaluating corporate communication campaigns.

We tackle this challenge using the four emotional categories of SenticNet[4] in addition to the standard sentiment polarity, which helps to distinguish different aspects of the target's emotional load, and computing per-aspect sentiment values that account for different properties relevant to users. The major challenge lies in identifying these relevant aspects. Most aspect-oriented sentiment analysis approaches

1541-1672/17/$33.00 © 2017 IEEE
Published by the IEEE Computer Society

Figure 1. Screenshot of the Media Watch on Climate Change, a public Web intelligence application available at www.ecoresearch.net/climate (query: "Volkswagen"; English-language news media sites, 07–12/2016).

use frequency- and syntax-based methods without linking to background knowledge for further reasoning.

For evaluation purposes, we asked domain experts to assess the retrieved common and commonsense knowledge (for example, aspect is relevant, not relevant, or unsure whether it is relevant), and provided a radar chart visualization to demonstrate how the entities and the corresponding aspects perform according to the SenticNet emotional categories.

Methodology

Our approach pursues a flexible and automated strategy by linking input entities to DBpedia to obtain background knowledge on relevant properties of these entities (aspects). The dependency graph enrichment adds background information on emotional categories and trigger terms, as well as sentiment targets and aspects obtained from the knowledge acquisition components, yielding the opinion graph used in the knowledge extraction process. A text document is represented by several of these opinion graphs. Sentiment-target linking uses a machine learning classifier to connect sentiment targets and aspects to trigger terms. The component relies on sentence dependency graphs as input, which represent tokens as nodes and their dependencies as directed edges. The sentiment-parsing component finally extracts affective knowledge from the opinion graph. It refines this knowledge with factual knowledge on relevant sentiment targets and aspects obtained from graph mining, and stores it in the affective knowledge repository. Figure 2 summarizes the affective knowledge extraction process for identifying beliefs, opinions, and arguments in text documents.

Knowledge Acquisition

The knowledge acquisition component provides information for enriching dependency graphs, as outlined in the next section, with information on a term's polarity obtained from a polarity lexicon, its SenticNet emotional categories, common knowledge acquired from DBpedia, and commonsense knowledge from ConceptNet.

Since our work focuses on car brands and models, we only mine sentiment targets and sentiment aspects relevant to this domain. Algorithm 1 (Figure 3) captures information on companies, products, and aspects from DBpedia and ConceptNet. It obtains relations that lead from the entity (for example, "Volkswagen") to an associated aspect. For instance, the relation "manufacturer" yields "Lupo" or "Golf" from DBpedia. This association reveals that Volkswagen manufactured the car models

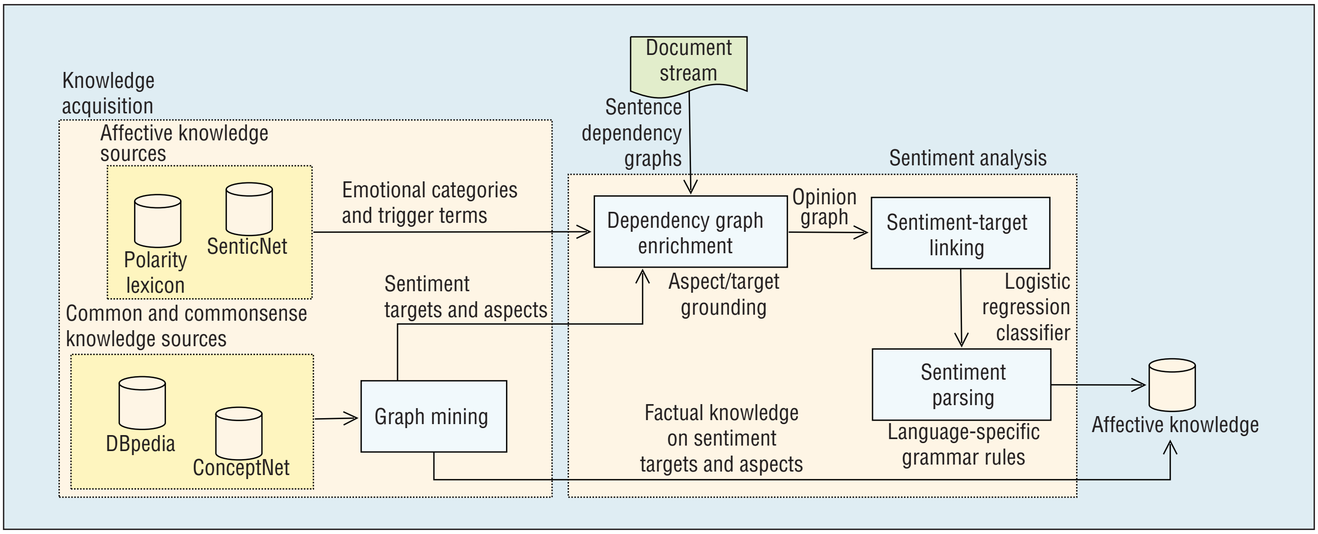

Figure 2. Main components of the affective knowledge extraction process. Preprocessing transforms documents into dependency graphs that are then enriched with external knowledge obtained from the knowledge acquisition component to create opinion graphs. Sentiment analysis extracts affective knowledge from these graphs that is then combined and extended with common and commonsense knowledge.

```
Require: sets of target_industries, target_predicates and product_predicates
// Lists for storing the results of the graph mining process
companies ←{}, products ←{}, entity_graph ←{},
// graph mining
for all triple from query (?s <rdf:type> <dbo:Company>) do
 if (?s <dbp:industry> ?o) and ?o in target_industries then
  companies.add(triple.s)
  entity_graph.add_triple(triple)
 end if
end for
for all triple from (query (?s ?p ?o ∈companies)
          ∪query(?s ∈companies ?p ?o)
          ∪query(?s ?p ∈target_predicates ?o))
 do
 if triple.p ∈product_predicates then
  products.add(triple.o)
 end if
 entity_graph.add_triple(triple)
end for
for all triple from query (?s <dbp:aka> ?o) do
 entity_graph.add_triple(triple)
end for
return companies, products, entity_graph
```

Figure 3. Algorithm 1 extracts sentiment targets and aspects as well as the corresponding context information from DBpedia. At first, the algorithm mines companies and products relevant to the target industry and then obtains subgraphs with context information on these two entity types.

Lupo and Golf. The DBpedia relation "keyPerson" yields "Martin Winterkorn" and "Ferdinand Piëch," both former chairs of Volkswagen, as important persons related to the company.

A set of predefined relations helps restrict the aspects to those most relevant for the investigation:

- DBpedia: dbo:manufacturer, dbo:key Person, dbo:product, dbp:team
- ConceptNet: PartOf, HasA, UsedFor, MadeOf

The algorithm obtains not only the label of the DBpedia resources but also all linked aliases. Additionally, it automatically creates aliases by removing tokens that are shared between the manufacturer and the product. This means it can automatically create the alias "Golf" from the car entity "Volks-wagen Golf" and the company entity "Volkswagen," thereby increasing the achievable recall.

To increase coverage, the graph-mining component queries ConceptNet for automobile properties and adds commonsense knowledge, such as that a car is a means of transport and has a steering wheel and a trunk (aspects). This component later uses the obtained relations (such as products produced by a company or key people working for that company) in conjunction with commonsense knowledge (such as major parts of such an entity or its typical applications) to enrich the dependency graph.

This knowledge-rich approach has two advantages over frequency-based methods that rely on syntactic features. First, the created affective knowledge base captures not only related entities but also the corresponding relation types. Second, grounding targets to DBpedia helps to obtain additional information such as abstracts, further relations, and car type.

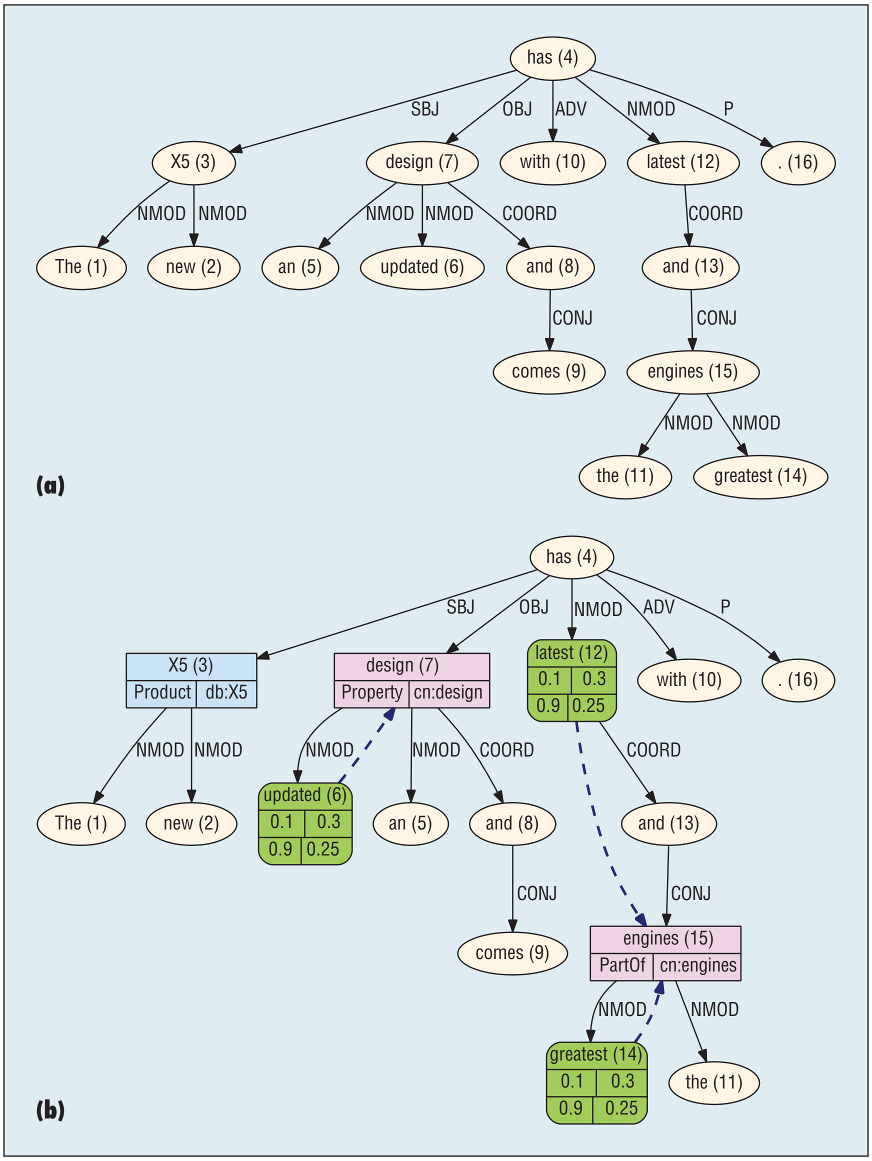

Figure 4. Dependency graph enrichment: (a) dependency tree of the sentence, "The new X5 has an updated design and comes with the latest and greatest engines"; and (b) enriched with opinions (blue: targets with type and DBpedia concept; violet: aspects with type and ConceptNet grounding; green: positive sentiment terms and sentic values; and dashed lines connect sentiment terms with their targets).

Sentiment Analysis

Using the affective and factual resources provided by the knowledge acquisition component, sentiment analysis follows a three-step process: dependency graph enrichment, sentiment-target linking, and sentiment parsing.

Dependency graph enrichment. Enriching the sentence dependency graphs with emotional categories, trigger terms that indicate negations or modify sentiment values, sentiment targets, and sentiment aspects obtained from the knowledge acquisition component yields the *opinion graph*, which we use in the subsequent sentiment-target linking and sentiment parsing steps.

After creating the dependency parse tree (see Figure 4a), the system draws upon the knowledge acquisition component to ground target concepts

Related Work in Sentiment Analysis

A better understanding of sentiment is crucial for building next-generation artificial intelligence systems and increasing the value of business intelligence applications.[1] This requires the integration of multiple approaches into a unified system, including the three research areas outlined in the following.

Emotion Analysis

Emotion analysis draws upon psychology research. For instance, SenticNet[2] is based on Plutchik's wheel of emotions.[3] It contains 50,000 concepts and maps them to the four dimensions proposed in the Hourglass of Emotions[4]: "aptitude" (confident in interaction benefits), "attention" (interested in interaction contents), "pleasantness" (amused by interaction modalities), and "sensitivity" (comfortable with interaction dynamics). WordNet-Affect[5] has affective labels such as "emotion," "mood," and "cognitive state" to approximately 2,800 WordNet synsets. The General Inquirer provides emotional categories such as "virtual," "pleasure," and "pain."[6] EmoLex contains approximately 10,000 terms,[7] and Affective Norms for English Words knows the three categories "valence" (from unpleasant to pleasant), "arousal" (from calm to excited), and "dominance."[8]

Sentiment-Target Linking

This research field identifies the target of an opinionated statement. For instance, "VW Golf" is the target of "reliable" in the statement, "The VW Golf is reliable." Rule-based approaches to sentiment-target linking use manually designed heuristics to find valid sentiment-target pairs—for example, sentiment-target proximity (distance-based approaches),[9] semantic frames,[10] or syntax-based approaches relying on a handful of patterns.[11,12] Supervised machine learning methods collect patterns from annotated corpora automatically. For example, Lei Zhuang and his colleagues[13] and Liheng Xu[14] automatically extract dependency patterns between sentiments and their targets.

Corpora such as J.D. Power and Associates (JDPA) support the evaluation of such tools.[15] We used a similar approach to build our classifier and further optimized its performance by evaluating and selecting features and including additional patterns learned from the multiperspective question answering (MPQA) corpus.[16]

Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis extends target-dependent sentiment analysis and identifies opinions on aspects of that entity. For example, given an entity "car," its design and engine characteristics are different aspects of the same entity. Most research focuses on product reviews and links mentioned aspects to opinions.[17] State-of-the-art approaches use term or *n*-gram frequencies[18,19] and frequently employ machine learning—for example, conditional random fields (CRF),[20] deep learning,[21] and latent Dirichlet allocation (LDA).[22] Other approaches combine syntactic rules and lexical resources.[23,24]

Our approach uses a knowledge base to identify aspects. This approach is similar to work by Caroline Brun and her colleagues, who bootstrap an aspect lexicon using a training corpus by combining WordNet and Wikipedia,[25] or Basant Agarwal and his colleagues, who access ConceptNet and WordNet to create a product-review-specific ontology.[26]

Proper opinion analysis is a combination of all these methods. After identifying an emotion, it is necessary to connect it to its target to allow reasoning such as, "who thinks what about whom?" Finally, identifying additional aspects related to the target gives higher granularity and further insight into the true meaning of the expressed opinion.

(that is, cars) to DBpedia. Afterward, it uses this information together with the context retrieved from DBpedia to query the knowledge acquisition component for aspects relevant to the targets from ConceptNet and to link these aspects to the corresponding ConceptNet nodes (see Figure 4b).

The affective knowledge extraction uses lexical lookups to identify tokens carrying affective knowledge and assigns them a value in the range [–1, 1]. The component supports multiple emotional categories. Grounding emotion triggers is not limited to string matching; rather, it is also aware of parts-of-speech (POS) tags. In the case of "like," for example, it differentiates between the use as a positive verb and as a neutral comparison term.

The system ignores product aliases unless the entity (obtained from DBpedia), its manufacturer, or the company's aliases occur in the text. This avoids problems with generic names (such as numbers, frequent domain-agnostic terms, or short character sequences) and allows it to correctly identify "BMW" and "X5" (for example, in "Yesterday BMW showed its newest SUV for the first time. The new X5 has an updated design and comes with the latest and greatest engines") without creating links if "BMW" is not mentioned.

The discovery of an aspect requires its subsequent linking to an entity (for example, "steering wheel" and "car"). A colocation heuristic helps find the closest candidate by scanning the current sentence first and, if unsuccessful, the entire document. The sentiment-target linking classifier then links the common and commonsense knowledge to the affective knowledge targeted at it.

Sentiment-target linking. Sentiment-target linking uses a set of sentiment terms (that is, terms indicating a certain emotion or sentiment) $S_m = \{t_{s_i}\}$ and target terms (that is, sentiment targets or aspects) $T_m = \{t_{t_j}\}$ extracted from sentence m, and returns a set of valid sentiment-target pairs: $\{(t_{s_i}, t_{t_j})\}$,

References

1. E. Cambria, "Affective Computing and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, 2016, pp. 102–107.
2. E. Cambria and A. Hussain, *Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis*, Springer, 2015.
3. R. Plutchik, "The Nature of Emotions," *Am. Scientist*, vol. 89, no. 4, 2001, pp. 344–350.
4. E. Cambria, A. Livingstone, and A. Hussain, "The Hourglass of Emotions," *Cognitive Behavioural Systems*, LNCS 7403, Springer, 2012, pp. 144–157.
5. C. Strapparava and A. Valitutti, "WordNet-Affect: An Affective Extension of WordNet," *Proc. 4th Int'l Conf. Language Resources and Evaluation* (LREC 04), 2004, pp. 1083–1086.
6. P.J. Stone, *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press, 1966.
7. S. Mohammad and P.D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," *Computational Intelligence*, vol. 29, no. 3, 2013, pp. 436–465.
8. M.M. Bradley and P.J. Lang, *Affective Norms for English Words (ANEW): Stimuli, Instruction Manual, and Affective Ratings*, tech. report, Center for Research in Psychophysiology, Univ. of Florida, Gainesville, 1999.
9. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2004, pp. 168–177.
10. S.-M. Kim and E. Hovy, "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," *Proc. Workshop Sentiment and Subjectivity in Text*, 2006, pp. 1–8.
11. S. Gindl, A. Weichselbraun, and A. Scharl, "Rule-Based Opinion Target and Aspect Extraction to Acquire Affective Knowledge," *Proc. WWW Workshop Multidisciplinary Approaches to Big Social Data Analysis* (MABSDA 13), 2013, pp. 557–564.
12. S. Poria et al., "Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns," *IEEE Computational Intelligence*, vol. 10, no. 4, 2015, pp. 26–36.
13. L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie Review Mining and Summarization," *Proc. 15th ACM Int'l Conf. Information and Knowledge Management*, 2006, pp. 43–50.
14. L. Xu et al., "Mining Opinion Words and Opinion Targets in a Two-Stage Framework," *Proc. 51st Ann. Meeting Assoc. for Computational Linguistics* (ACL 13), vol. 1, 2013, pp. 1764–1773.
15. J.S. Kessler and N. Nicolov, "Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations," *Proc. 3rd Int'l AAAI Conf. Weblogs and Social Media* (ICWSM 09), 2009, pp. 90-97.
16. T. Wilson, "Fine-Grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States," PhD diss., Univ. of Pittsburgh, 2008.
17. M. Pontiki et al., "Semeval-2014 Task 4: Aspect-Based Sentiment Analysis," *Proc. 8th Int'l Workshop Semantic Evaluation* (SemEval 14), 2014, pp. 27–35.
18. E. Marrese-Taylor, J.D. Velásquez, and F. Bravo-Marquez, "A Novel Deterministic Approach for Aspect-Based Opinion Mining in Tourism Products Reviews," *Expert Systems with Applications: An Int'l J.*, vol. 41, no. 17, 2014, pp. 7764–7775.
19. Y. Wu and M. Ester, "FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering," *Proc. 8th ACM Int'l Conf. Web Search and Data Mining* (WSDM 15), 2015, pp. 199–208.
20. M. Chernyshevich, "Cross-Domain Extraction of Product Features Using Conditional Random Fields," *Proc. 8th Int'l Workshop on Semantic Evaluation* (SemEval 14), 2014, p. 309–313.
21. S. Poria, E. Cambria, and A. Gelbukh, "Aspect Extraction for Opinion Mining with a Deep Convolutional Neural Network," *Knowledge-Based Systems*, vol. 108, Special Issue on New Avenues in Knowledge Bases for Natural Language Processing, 2016, pp. 42–49.
22. S. Poria et al., "Sentic LDA: Improving on LDA with Semantic Similarity for Aspect-Based Sentiment Analysis," *Proc. Int'l Joint Conf. Neural Networks* (IJCNN 16), 2016, pp. 4465–4473.
23. K. Veselovská and A. Tamchyna, "UFAL: Using Hand-crafted Rules in Aspect Based Sentiment Analysis on Parsed Data," *Proc. 8th Int'l Workshop Semantic Evaluation* (SemEval 14), 2014, pp. 694–698.
24. A.G. Pablos et al., "Unsupervised Acquisition of Domain Aspect Terms for Aspect Based Opinion Mining," *Procesamiento del Lenguaje Natural*, vol. 53, Sept. 2014, pp. 121–128.
25. C. Brun, D.N. Popa, and C. Roux, "XRCE: Hybrid Classification for Aspect-Based Sentiment Analysis," *Proc. 8th Int'l Workshop Semantic Evaluation* (SemEval 14), 2014, pp. 838–842.
26. B. Agarwal et al., "Sentiment Analysis Using Common-Sense and Context Information," *J. Computational Intelligence and Neuroscience*, vol. 2015, Jan. 2015, article 30.

where $y\left(t_{s_i}, t_{t_j}\right) = True$. Hereby, we formulate the sentiment-target linking task as a binary classification problem. The classification function y reflects whether sentiment t_{s_i} and target tokens t_{t_j} constitute a valid sentiment-target pair.

The component starts by generating all possible edges between the set of targets and the set of sentiments as candidates for valid sentiment-target pairs and further evaluates each of them independently.

The component extracts features for every observation of a sentiment-target pair and uses them as input for the classification model previously trained on a corpus annotated with correct sentiment-target pairs. To train the classifier, it uses observations from a corpus annotated with words and phrases expressing sentiments $\left\{t_{s_i}\right\}$, targets $\left\{t_{t_j}\right\}$, and relations between them $\left\{\left(t_{s_k}, t_{t_l}\right)\right\}$. An observation $x\left(t_{s_i}, t_{t_j}\right)$ is a set of features that captures syntactic relations between the sentiment token t_{s_i} and the target token t_{t_j}. A recursive feature elimination (RFE) procedure yields an optimal feature set to be extracted from the opinion graph for each observation of a sentiment-target pair $x\left(t_{s_i}, t_{t_j}\right)$, which comprises features such as POS tags and dependencies between the sentiments and target nodes in the graph.

The sentiment-target linking uses a logistic regression classifier trained on the J.D. Power and Associates (JDPA, http://verbs.colorado.edu/jdpacorpus) sentiment corpus and the Multiperspective Question Answering (MPQA, http://mpqa.cs.pitt.edu/corpora/mpqa_corpus) opinion corpus, version 2.0 (also see the sidebar). An evaluation of the sentiment-target linking performance achieved an F-measure of 0.90 when evaluated on the gold-standard annotations for about 12,000 sentiment-target pairs with stratified tenfold cross validation.

Sentiment parsing. Grammar rules and heuristics help identify and extract

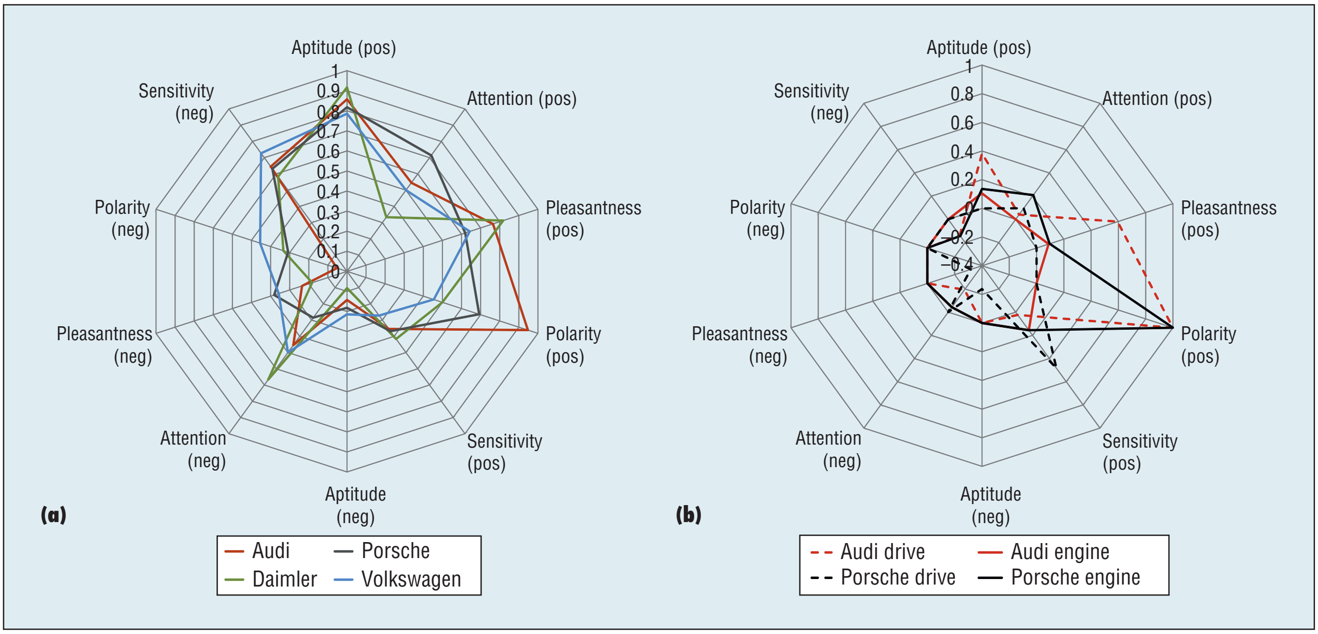

Figure 5. SenticNet emotional categories and polarities for (a) selected car brands, and (b) the aspects (in this case, product features) "drive" and "engine."

Table 1. Statistics of the acquired background knowledge.

Description	No.
Companies active in the automotive industries	881
Key people in these companies	349
Car entities	4,898
Car aliases	7,111
Car aspects	30

affective knowledge—for example, negation detection to invert the polarity of a negated term. It uses nodes marked as triggers and stoppers to determine the start and end of the negation scope within the opinion graph, and supports multiple negation.

Aggregating the opinion triggers that have been linked to a particular sentiment target yields the target's value for the corresponding emotional category. By considering different sentiment aspects in this aggregation process, the system can analyze the emotions contributed by each aspect, yielding visualizations such as the one presented in Figure 5.

Data Analytics

An RDF triple store serves to store affective and factual knowledge. A proof-of-concept data analytics application queries the affective knowledge base to compare the emotions associated with four automobile brands having high media coverage (Audi, Daimler, Porsche, and Volkswagen). It contrasts this analysis with an evaluation of two different aspects (drive and engine) relevant to products of two of these brands (Audi and Porsche).

The affective knowledge repository facilitates polarity classification and emotional analysis aligned with the "Hourglass of Emotions" (see the sidebar). For instance, the "engine" of "VW" receives a sensitivity of −0.07, whereas "Golf" has a sensitivity of 0.014. After determining the emotional strength associated with each company and aspect, we aggregate over all aspects and calculate a total value using the following formula:

$$strength_{emotion} = \frac{k}{n}, \qquad (1)$$

where k is the number of positive occurrences (negative occurrences for negative strength) of the emotional dimension, while n is the total number of occurrences of this emotion. A summary of the obtained results is presented later.

Experiments

Using a subset of the archive of the Media Watch on Climate Change (social media messages published between 28 September and 28 November 2015), the evaluation corpus consists of 1,000 Twitter and Google+ postings containing the word "car," and 4,000 referring to one of the car brands Audi, Daimler, Porsche, and Volkswagen. The former helped extract sentiment aspects and targets contained in the knowledge base, the latter supported the evaluation of aspect-based emotion analysis.

Graph Mining Results

The approach introduced earlier yields a considerable amount of background knowledge from DBpedia and ConceptNet that has been used for the sentiment analysis. Table 1 lists the number of entities, aliases, and aspects acquired from the common and commonsense knowledge sources.

Table 2 shows the obtained entities and properties for the company Tesla Motors and the car Tesla Model S, demonstrating the level of detail achieved with the presented approach.

Table 2. Extracted entities and aspects connected to the car company Tesla Motors and the car Tesla Model S.

Entity	Relation	Aspect
Tesla Motors	Type	Company
	Industry	Car
	Manufacturer	Tesla Model S, Tesla Roadster
	Product	Luxury vehicle
	Key person	JB Straubel, Elon Musk, chief executive officer, chief technology officer, chair
Tesla Model S	Aka	WhiteStar, Model S
	HasA	trunk, radio, headlight, four wheel, seat, wheel, engine, window, four tires
	MadeOf	steel, metal
	PartOf	trunk, engine, transmission, radiator, body, hood, tire, fender, door, tire, engine, steer wheel, drive train, wheel
	UsedFor	drive, transportation, travel

Evaluation of the Extracted Knowledge

The following evaluation draws upon the 50 most frequently occurring sentiment aspects and targets in the evaluation corpus to assess the usefulness and impact of the knowledge extracted by the graph mining. Five independent domain experts classified the usefulness of each extracted concept for describing aspects relevant to the perception (polarity, emotions) of car companies, brands, and products in one of three categories: *useful* (the aspect is related to the domain), *not useful* (the aspect has no connection to the domain), and *neutral* (the term is too generic to be clearly associated with the domain). On average, 81.2 percent of the extracted concepts have been considered useful. The Krippendorff alpha for inter-rater agreement between experts is 0.504, reflecting only a moderate agreement among domain experts.

The evaluation illustrates two shortcomings of the current approach. First, the assumption that automotive companies only manufacture cars does not hold true. Among the 50 most frequent entities/aspects in the "car" corpus, the system identified "knife" because the company American Expedition Vehicles also produces knives. We investigated narrowing the products based on their `rdfs:class` property but encountered a diverse set of assigned classes that have no single common superclass or shared property.

Second, two ambiguous car brands showed up in the evaluation: the short-lived WiLL and SEAT, which was often confused with car seat. WiLL could be tackled by allowing only certain aspects to be matched with verbs (for example, aspects connected with the "UsedFor" predicate to a car). SEAT, however, is difficult to ground: in social media, capitalization cannot reliably be used for disambiguation (since often the text is all lowercase), and the domain car fits both car seat and the car brand SEAT.

Aspect-Based Analysis of Brand Perceptions

Using the data analytics approach presented earlier and building on previous work to visualize emotions along multiple dimensions,[5] we show how the affective knowledge extracted from social media messages can be associated with the investigated car brands (Figure 5a). Applying the Hourglass of Emotions to the emotional dimensions "aptitude," "attention," "pleasantness," and "sensitivity" lets us map numerical chart values to their emotional equivalents. The car brand Audi, for example, shows a strong association with positive aptitude (0.86), which maps to the emotion "admiration" on the Hourglass of Emotions. The brand is also associated with a moderate negative sensitivity (0.65), which is equivalent to "fear." Negative attention (0.45) reveals that "surprise" is also associated with "Audi."

"Volkswagen" has the most significant peaks in the negative direction—for example, a negative attention of 0.5, a negative pleasantness of 0.36, and a negative sensitivity of 0.72. These values map to "surprise," "sadness," and "terror" on the Hourglass of Emotions. This result is in line with the negative media coverage about the exhaust scandal.

Emotion analysis provides detailed feedback on the public perception of a company. A brand might outperform another in one aspect, such as product quality, but might have to catch up on another aspect, such as service. The radar chart in Figure 5b, for example, shows that the Porsche engine has a considerably higher attention and sentiment than its competitor, but Audi excels in pleasantness and sentiment when focusing on actually driving the car.

Among the main challenges of deploying aspect-based opinion mining algorithms for Web intelligence applications are the required scalability of the computational methods, and appropriate visual representations that convey the aspect structure and associated emotions in an intuitive manner. The European research project Adaptive Scalable Analytics Platform (ASAP, www.asap-fp7.eu) is currently tackling both challenges. ASAP will enable us to perform the required complex computations on high-volume content streams from social networking platforms, and to provide real-time visualizations of the evolving aspect structure as part of an interactive dashboard—going beyond standard representations such as trend lines and radar charts.[6]

Future research will also apply the presented methods in different domains and demonstrate their applicability beyond specific products and services. Measuring the impact of international marketing and public outreach campaigns, for example, would significantly benefit from an aspect-oriented approach. Simple bipolar metrics such as sentiment cannot adequately reflect the underlying complexities when millions of stakeholders use digital channels to participate in public debates about complex, multi-faceted topics. ■

Acknowledgments

The work presented here was conducted as part of the IMAGINE project, funded by the Swiss Commision for Technology and Innovation (CTI); the uComp project, which receives funding from EPSRC EP/K017896/1, FWF 1097-N23, and ANR-12-CHRI-0003-03, in the framework of CHIST-ERA ERA-NET; and the DecarboNet and ASAP projects, which receive funding from the EU 7th Framework Program for Research, Technology Development and Demonstration under grant agreements 610829 and 619706, respectively.

References

1. A. Weichselbraun, S. Gindl, and A. Scharl, "Extracting and Grounding Contextualized Sentiment Lexicons," *IEEE Intelligent Systems*, vol. 28, no. 2, 2013, pp. 39–46.
2. A. Scharl et al., "Semantic Systems and Visual Tools to Support Environmental Communication," *IEEE Systems J.*, preprint, 8 Sept. 2015; doi:10.1109/JSYST.2015.2466439.
3. A. Scharl et al., "Scalable Knowledge Extraction and Visualization for Web Intelligence," *Proc. 49th Hawaii Int'l Conf. System Sciences* (HICSS 16), 2016, pp. 3749–3757.
4. E. Cambria et al., "SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives," *Proc. 26th Int'l Conf. Computational Linguistics: Technical Papers* (COLING 16), 2016, pp. 2666–2677.
5. A. Scharl et al., "Analyzing the Public Discourse on Works of Fiction—Automatic Emotion Detection in Online Media Coverage about HBO's Game of Thrones," *Information Processing and Management*, vol. 52, no. 1, 2016, pp. 129–138.
6. A.M. Brasoveanu et al., "Visualizing Statistical Linked Knowledge for Decision Support," *Semantic Web J.*, vol. 8, no. 1, 2017, pp. 113–137.

Albert Weichselbraun is a professor of information science in the Swiss Institute for Information Research at the University of Applied Sciences Chur, and technical director of webLyzard technology. His research interests include big data, natural language processing, and text mining. Weichselbraun has a habilitation and doctorate in information systems from the Vienna University of Economics and Business. Contact him at albert.weichselbraun@htwchur.ch.

Stefan Gindl is an assistant professor in the Department of New Media Technology at MODUL University Vienna. His research interests include text mining, natural language processing, machine learning, and artificial intelligence. Gindl has a PhD in computer sciences from the Vienna University of Technology. Contact him at stefan.gindl@modul.ac.at.

Fabian Fischer is a researcher in the Institute for Information Business at the Vienna University of Economics and Business. His research interests include natural language processing, algorithmics, big data, and privacy. Fischer has an MS in computational intelligence from the Vienna University of Technology. Contact him at fabian.fischer@wu.ac.at.

Svitlana Vakulenko is a researcher in the Institute for Information Business at the Vienna University of Economics and Business. Her research interests are natural language processing, including information extraction, machine learning, and social media. Vakulenko has an MSc in software engineering from the University of Tartu, Estonia. Contact her at svitlana.vakulenko@wu.ac.at.

Arno Scharl is a professor of information systems in the Department of New Media Technology at MODUL University Vienna, and the managing director of webLyzard technology. His research interests include media intelligence, big data analytics, information visualization, and the integration of semantic and geospatial Web technology. Scharl has a habilitation and doctorate in management information systems from the Vienna University of Economics and Business. Contact him at scharl@weblyzard.com.