OntoSenticNet 2: Enhancing Reasoning within Sentiment Analysis

Mauro Dragoni
Fondazione Bruno Kessler

Ivan Donadello
University of Bolzano

Erik Cambria
Nanyang Technological University

***Abstract*—Sentiment analysis is a trending topic that has not yet exhausted its attractiveness, despite the huge research effort carried out in the last fifteen years. One of the most promising directions to investigate is the integration of knowledge-based representations within sentiment analysis systems in order to enhance their expressiveness and, at the same time, to enable reasoning over the relevant information detected within opinion-based sources. In this paper, we present an improved version of OntoSenticNet providing (i) an updated definition of concepts, properties, and individuals together with an improved hierarchical organization of such entities; (ii) the modeling of the sentic algebra elements for supporting the execution of semantic sentiment operations at reasoning time; and (iii) the conceptual model of sentiment dependencies and discovery paths. The process of building OntoSenticNet 2 is discussed and some examples are proposed in order to illustrate the conceptual model.**

In recent years, sentiment analysis has become increasingly popular for processing social media data on online communities [1], social networks [2], and microblogging platforms [3]. Most of the literature is on English language but recently an increasing number of works are tackling the multilinguality issue [4], especially in booming online languages such as Chinese [5] and Arabic [6].

Besides traditional domains like business intelligence [7] and recommendation systems [8], sentiment analysis applications also include many other areas like financial forecasting [9], healthcare [10], cyber-harassment prevention [11], political forecasting [12], and dialogue systems [13]. However, mining opinions and sentiments from multimodal resources (texts, images, videos, audio recordings, etc.) is an extremely difficult

 This article is part of the IEEE IS ACSA series

task because it requires a deep understanding of the explicit and implicit, regular and irregular, features (linguistic, visual, or audio) of a resource. Existing approaches to multimodal sentiment analysis mainly rely on mapping multimodal information to parts of text in which opinions are explicitly described, such as polarity terms, affect words, and their co-occurrence frequencies [14]. However, opinions and sentiments associated with these parts of text are often conveyed implicitly through latent semantics, which make purely syntactic approaches ineffective.

The task of associating polarities to these features suffers from the limitation of not being able to perform inference operations on the concepts extracted (or mapped in case of multimodal resources) from the resource to analyze. An example is the following: let us assume to have a vocabulary of opinion concepts containing the primitive DECREASE_PAIN associated with a positive polarity. Sentiment analysis tools commonly available are based on detecting the exact match between chunks of text and concept labels. Thus, if a text contains chunks of the form `reduce_agony` or `diminish_affliction`, the system is not able to exploit sentiment information associated with the primitive DECREASE_PAIN due to the undefined relationship between the textual expression and the vocabulary's concept.

In this paper, we present OntoSenticNet 2[1], the second version of a commonsense ontology for sentiment analysis based on SenticNet [15], a semantic network of 200,000 concepts based on conceptual primitives (Fig. 1). The characteristics that distinguish OntoSenticNet 2 (available for download on SenticNet website[2]) from the first version of OntoSenticNet described in [16] are: (i) the update of the conceptual hierarchy and properties associating concepts and sentiment values together with the conceptualization of sentiment similarity; (ii) the conceptualization of sentic algebra enabling the support for modeling semantic sentiment operations; and (iii) the knowledge about sentiment dependencies and discovery path.

OntoSenticNet 2 does not blindly use keywords and word co-occurrence counts, but instead relies on the implicit meaning associated with commonsense concepts. Unlike purely syntactic techniques, OntoSenticNet 2 can detect subtly expressed sentiments by enabling the analysis of multiword expressions that do not explicitly convey emotion but are instead related to concepts that do. Moreover, the provided representation supports the integration of reasoning engines able to infer implicit sentiment information.

SENTIMENT ONTOLOGIES

Despite the rise of affective computing and related disciplines [17], there is a lack of sentiment ontologies. In particular, there are only two general models that are very limited in terms of functionalities and possibility of being integrated into real-world applications. The Emotion Markup Language (EML) was created for supporting the task of annotating documents with tags extracted from customized vocabularies[3]. On the one hand, this language is useful for creating emotional dictionaries for specific domain. On the other hand, the effort necessary for creating a new resource is significant and, at the same time, the promotion of a markup language fosters the proliferation of resources that often have a high linguistic and semantic overlap. This way, reusability is strongly penalized.

The other model is the Emotion Ontology (MFOEM), which was developed for supporting a structured representation of mental functioning, including mental processes such as cognition and traits such as intelligence[4]. This ontology can complement SenticNet in a sense that, while OntoSenticNet 2 is specifically thought for describing the emotional domain, MFOEM can be considered an upper level ontology that can be aligned with the top-level concepts of SenticNet. This way, OntoSenticNet 2 would benefit from categorizations and properties describing human brain from a more general perspective and, at the same time, the MFOEM ontology can exploit the granularity of OntoSenticNet 2 for accessing to real-world emotional items, e.g., documents, images, and videos. We leave this alignment task for future work.

[1]The ontology is available at http://w3id.org/ontosenticnet
[2]http://sentic.net/downloads
[3]http://w3.org/TR/emotionml
[4]http://bioportal.bioontology.org/ontologies/MFOEM

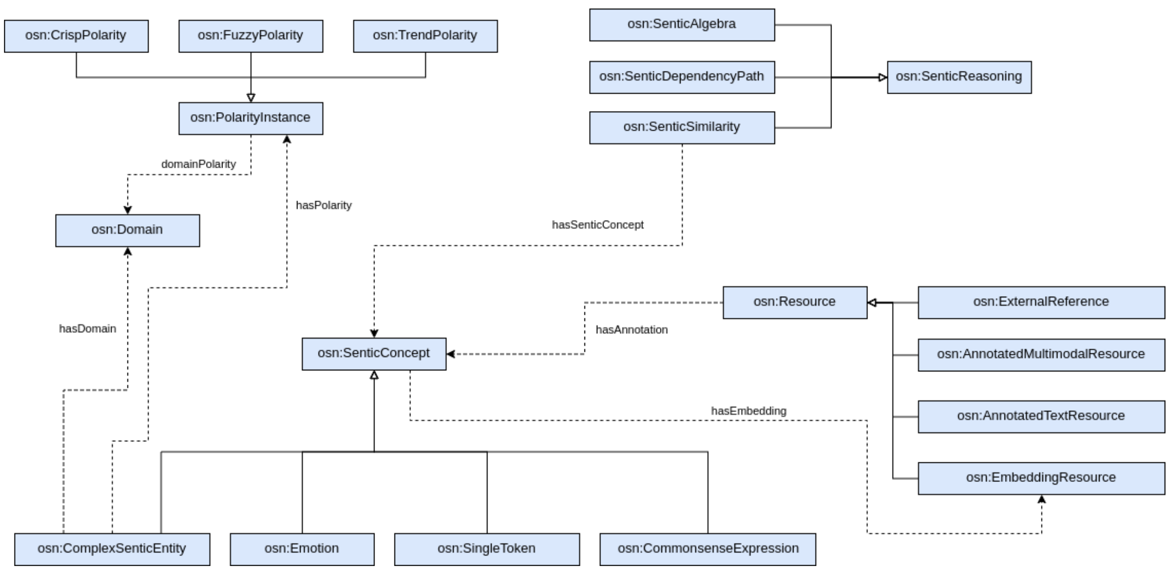

Figure 1. Overview of OntoSenticNet 2. Commonsense knowledge is organized at three levels: entities link down to concepts, which link down to conceptual primitives where meaning is encoded in terms of polarity and emotions.

BUILDING ONTOSENTICNET 2

The process executed for building OntoSenticNet 2 follows the same steps we performed for constructing the first version as described in [16]. The whole building process is summarized below. The process of building OntoSenticNet 2 follows the METHONTOLOGY methodology [18]. This methodology proposes a general method for building any kind of ontology or meta-ontology and it is based on the experience acquired in developing ontologies in the domain of chemicals. METHONTOLOGY provides a set of guidelines of how the activities identified in the ontology development process should be carried out, what kinds of techniques are the most appropriate in each activity, and what products each one produces. The methodology is split in seven phases summarized below.

Specification

OntoSenticNet 2 has been thought for filling the gap between fundamental emotion ontologies (like EML and MFOEM) that cannot be easily integrated into real-world applications and sentiment words dictionaries that, instead, do not support a semantic representation of the emotion domain. Moreover, OntoSenticNet 2 aims to bridge concepts and resources in order to enable its integration into complex annotation and reasoning frameworks. OntoSenticNet 2 is represented by using a natural-language semi-formal format due to the necessity of adopting concept names expressing specific meanings through their labels. The granularity level is classified as high thanks to the rich set of terminologies and commonsense expressions represented into the ontology.

Knowledge Acquisition

OntoSenticNet 2 is automatically constructed from three resources of affective commonsense knowledge: WordNet-Affect [19], Open Mind Common Sense [20] and GECKA [21]. Semantic and affective information are extracted from such knowledge through the ensemble application of spreading activation [22] and sentic neurons [23]. Concepts are organized in a hierarchical structure (Fig. 2) where emotion-laden concepts (words and multiword expressions that do not refer to emotions directly but instead express or elicit emotions from the interlocutors) are linked to emotion-related ones like `smile` and `frown`. These, in turn, are linked to emotion concepts like `happiness` and `sorrow` which, finally, are categorized as one of the 24 key concepts of the Hourglass of Emotions [24], a biologically inspired and psychologically motivated emotion categorization model for sentiment analysis.

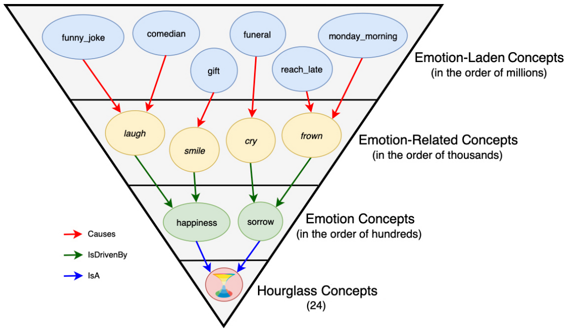

Figure 2. OntoSenticNet 2 dependency graph.

Conceptualization

The conceptualization of OntoSenticNet 2 was split in two steps. The first steps was covered by the knowledge acquisition phase, where most of the terminology is collected and directly modeled into the ontology. The second step, instead, consisted in defining the concept and the properties (both object and data) used for providing a detailed, as well as complex, representation of concepts polarities and for supporting annotation tasks. Details about the modeled concepts are provided in the next section, where the rationale of each concept is described.

Integration

No specific integrations have been performed during the developing process of OntoSenticNet 2. As mentioned earlier, OntoSenticNet 2 is not based upon existing meta-ontologies.

Implementation

The implementation of OntoSenticNet 2 is provided in two programming languages. RDF/XML provides a formal representation enabling the check of inconsistencies, the visualization of the ontology structure, and the ease of maintenance. Python, instead, provides an easier support for the integration of OntoSenticNet 2 into real-world applications. The two versions are always synchronized.

Evaluation

OntoSenticNet 2 has been verified by using the verification framework proposed in METHONTOLOGY. Based on the criteria proposed in the framework, OntoSenticNet 2 has been assessed as correct, complete, consistent, and not redundant.

Documentation

During the knowledge acquisition phase, we collected documentation about the information sources used for modeling OntoSenticNet 2. As for the conceptual documentation, we produced the set of guidelines we followed, for each phase, to model all concepts, objects properties, and data properties. The rationale behind the modeling of each concept is presented in the next section.

ENTITIES OF ONTOSENTICNET 2

OntoSenticNet 2 extends the four main branches defined within the first version of OntoSenticNet (i.e., `SenticConcept`, `Domain`, `PolarityInstance`, and `Resource`) with the `SenticReasoning` one. The `SenticReasoning` branch is originally presented in OntoSenticNet 2, while both the `SenticConcept` and the `PolarityInstance` branches have been updated with some new concepts. Then, new object properties and datatype properties have been added in order to increase the overall expressiveness of OntoSenticNet and to make it exploitable in more complex real-world scenarios.

The `SenticConcept` entity represents the conceptualization of what in sentiment analysis is termed *opinion word*. This entity represents the basic concept grouping all concepts that can be used for representing linguistic elements which can be associated with a sentiment expression. The `SenticConcept` entity subsumes four further concepts representing the four most well-known kind of elements that can be found in natural language texts and which can be exploited for annotating multimodal resources with sentiment information: `SingleToken`, `CommonsenseExpression`, `Emotion`, and `ComplexSenticEntity`.

The `SingleToken` concept embodies the *opinion word* element used to compute the orientation of text (i.e., positive or negative, but also neutral [25]) and its intensity value [26]. With the `CommonsenseExpression` concept, we provide the conceptualization of lexical expressions used for representing complex sentiment statuses. Through the modeling of these sentiment concepts, it is possible to support multimodal annotation activities where the single opinion is not expressive enough for providing a complete conceptual description of the sentiment status. The third concept is `Emotion`. It is used for representing Hourglass concepts like `joy` or `sadness` and it can be exploited as enabler for clustering instances of `SingleToken` or `CommonsenseExpression` concepts. The last concept of this branch is `ComplexSenticEntity`. This concept enables the representation of aggregations of emotions that are relevant with respect to a specific event, scenario, or domain. A sample instance of this concept might be a set of emotions composing the mental status of depression linked with the relevant domains and with the specific polarity values. Here, such a mental status can be represented by an aggregation of knowledge supporting its complete representation.

The second branch only contains the concept `Domain`. However, it represents an important knowledge aspect concerning the modeling of emotional status within real-world applications. The most valuable reason for considering the notion of `Domain` within sentiment analysis research is that there are plenty of adjectives or complex lexical expressions that assume opposite sentiment values in different contexts. Nevertheless, most of the literature concerning sentiment analysis and opinion mining does not handle emotional differences that the same lexical expression may elicit within different contexts or domains. Let us consider the adjective `small` as an example. If we talk about the ability of holding objects, it assumes a negative connotation. If we talk about portability, instead, it conveys a positive polarity. Likewise, `predictable` is bad for domains like books or movies but good in domains like finance. With this branch, we support the instantiation of domains and contexts that are of interest for the application where OntoSenticNet 2 is deployed into.

The third branch has the `PolarityInstance` as root concept. Concepts modeled within this branch represent different types of polarity that are supported by the OntoSenticNet 2 conceptual model. In the current version of the ontology, we modeled three representations for polarity values: `CrispPolarity` and `FuzzyPolarity`, which are inherited from the first version of OntoSenticNet, and the new `TrendPolarity` entity. Instances of the `CrispPolarity` concept provide the well-known single-value representation of polarity values and they are represented by double value types that are in turn linked with instances of `SenticConcept` entities. The rationale behind the `FuzzyPolarity` concept, instead, is to support the integration of uncertainty aspects within the representation of polarity values. This way, the conceptual model is able to express the possibility that a given polarity value is valid for a specific `SenticConcept`. This aspect is very interesting as well as important since the assignment of a specific value to a `SenticConcept` instance is a subjective task that may result in a collection of several polarity values. The use of fuzzy logic for representing these values allows real-world applications to properly interpret these polarity values during the inference task [27], [28]. Finally, with the new `TrendPolarity` concept, we equipped the OntoSenticNet 2 conceptual model with the possibility of describing polarities with trending information. The strength of this type of knowledge is to represent the evolution of sentiment information across a specific dimension. A funny, but realistic, example is related to the `CommonsenseExpression` instance `buy_christmas_gifts`: the sentiment linked with this expression can evolve from positive to light negative values in conjunction with the reduction of the time available for buying gifts.

The fourth branch is rooted within the `Resource` concept. The use of specific concepts for describing identifiers of distributed resources allows us to link OntoSenticNet 2 concepts with external artifacts (identified by persistent URIs). This way, we are able to create collections of annotated entities that can be exploited for further reasoning activities. Within OntoSenticNet 2, we identified four categories of resources represented by: `AnnotatedTextResource`, `AnnotatedMultimodalResource`, `EmbeddingResource`, and `ExternalReference`. Instances of `AnnotatedTextResource` refer to textual documents annotated with one or more instances of `SenticConcept`. Similarly, the `AnnotatedMultimodalResource` concept is instantiated when the annotated resources are images or videos. The support of this kind of conceptualization enables the possibility of refining, or learning, how `SenticConcept` instances are used for annotating content and, at the same time, to trigger machine learning activities on annotated resources for improving inference capabilities. Another resource category is identified by the `EmbeddingResource` concept.

Due to the massive use of features embeddings within the machine learning community, we wanted to provide a way for associating such embedding with OntoSenticNet 2. In the current version of the ontology, instances of this concept are represented by arrays of double values that can be directly linked with instances of the `SenticConcept`. Such links can be exploited for inference purposes and for supporting internal representation of documents within data repositories. The last concept of this branch is `ExternalReference`. The purpose of this concept is to work as a bridge between terminologies defined in external linguistic resources (for instance WordNet) with instances of type `SenticConcept`. Once such mappings are defined, OntoSenticNet 2 can be used as entry point for acquiring further information for the linked external resource.

The last branch concerns the root concept `SenticReasoning`. This set of concepts are originally included within OntoSenticNet 2 and they support the execution of further reasoning activities across the sentiment domain. `SenticReasoning` subsumes three entities: `SenticAlgebra`, `SenticDependencyPath`, and `SenticSimilarity`. `SenticAlgebra` allows for building sentiment expressions represented as algebraic operations that can be executed by a reasoner for inferring the overall sentiment value of a given situation. Moreover, through the concepts subsumed by the `SenticAlgebra` one, we are able to deconstruct multiword expressions into single tokens which polarity values can be aggregated (Fig. 3).

Figure 3. Examples of sentic algebra expressions.

A `SenticAlgebra` expression is composed by several elements all defined within OntoSenticNet 2: `SenticAlgebraExpression`, representing the actual whole sentiment expression; `SenticAlgebraFactor`, representing a single token of a sentiment expression; and `SenticAlgebraConnector`, representing the mathematical operators linking two instances of `SenticAlgebraFactor`.

The second reasoning-enabler concept is `SenticDependencyPath`. Through the instances of this concept, it is possible to model emotional paths defined by the conceptual connection between sentiment concepts. The usefulness of modeling emotional path is to enable the reasoner to better understand the positive and negative levels of sentiment information through chains of linguistic terms providing the connections between emotion primitives and more human-like terms like named entities. The `SenticDependencyPath` concepts subsume the different types of `SenticDependencyConcept` (namely, `Concept`, `NamedEntity`, `Primitive`, and `Superprimitive`) and `SenticDependencyDiscovery`, used for materializing the actual sentiment dependency discovery path found within SenticNet (Fig. 2).

Finally, the `SenticSimilarity` concept is used for instantiating the semantic similarity between two sentic elements defined within OntoSenticNet 2. Such a similarity can be computed between all types of elements and can be exploited at runtime for inference purposes. Besides concepts, OntoSenticNet 2 defines a list of ObjectProperty and DataProperty defining relationships between entities. We report them within Tables 1 and 2, respectively. Finally, we included two annotations for supporting unique identification of entities: `id` and `resourceIRI`.

Object Property	Domain	Range
complexPolarity	SenticConcept	PolarityInstance
discoveryPathFrom	SenticDependencyDiscovery	SenticDependencyConcept
discoveryPathTo	SenticDependencyDiscovery	SenticDependencyConcept
domainPolarity	PolarityInstance	Domain
hasAnnotation	AnnotatedTextResource AnnotatedMultimodalResource ExternalReference	SenticConcept
hasConnector	SenticAlgebraFactor	SenticAlgebraConnector
hasDomain	ComplexSenticEntity	Domain
hasEmbedding	SenticConcept	EmbeddingResource
hasFactor	SenticAlgebraExpression	SenticAlgebraFactor
hasPolarity	ComplexSenticEntity	PolarityInstance
hasPrimitive	CommonsenseExpression SingleToken	Emotion
hasSenticConcept	ComplexSenticEntity	SenticConcept
semanticTerm (antonym, hypernym, hyponym, synonym)	SenticConcept	SenticConcept ExternalReference

Table 1. List of the Object Properties included in OntoSenticNet 2.

Datatype Property	Domain	Range
crispPolarity	CrispPolarity	decimal
discoveryPathConfidence	SenticDependencyDiscovery	double
embeddingSize	EmbeddingResource	double
embeddingValues	EmbeddingResource	string
fuzzyShape	FuzzyPolarity	string
fuzzyValues	FuzzyPolarity	string
isLastFactor	SenticAlgebraFactor	boolean
polarityLabel	SenticConcept	string
positionInExpression	SenticAlgebraFactor	int
senticValue (introspection, temper, attitude, sensitivity, polarity)	SenticConcept	double
similarityValue	SenticSimilarity	double
trendPolarity	TrendPolarity	string

Table 2. List of the Datatype Properties included in OntoSenticNet 2.

CONCLUSION

Gauging public opinions has raised increasing interest within both the scientific and business communities because of the remarkable benefits offered by marketing and financial prediction, which have led to many exciting open challenges. While there are many lexicons and knowledge bases available for sentiment analysis, however, there is a lack of sentiment ontologies.

In this paper, we proposed OntoSenticNet 2, a conceptual model supporting the structuring analysis of emotions from multimodal resources based on SenticNet, a commonsense knowledge base for sentiment analysis. We discussed the methodology implemented for creating the resource and the rationale behind the main classes and properties modeled into the ontology. OntoSenticNet 2 is freely available for download and it can be easily integrated into business platforms and real-world applications.

Future work will focus on the development of an ecosystem of services and data that will be directly integrated into OntoSenticNet 2 in order to support the construction of smart emotion-sensitive applications.

REFERENCES

1. S. Cavallari et al., "Embedding both finite and infinite communities on graph," *IEEE Computational Intelligence Magazine*, vol. 14, no. 3, 2019, pp. 39–50.
2. D. Camacho et al., "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, 2020, pp. 88–120.
3. E. Cambria, H. Wang, and B. White, "Guest Editorial: Big Social Data Analysis," *Knowledge-Based Systems*, vol. 69, 2014, pp. 1–2.
4. S. L. Lo et al., "Multilingual sentiment analysis: From formal to informal and scarce resource languages," *Artificial Intelligence Review*, vol. 48, no. 4, 2017, pp. 499–527.
5. H. Peng, E. Cambria, and A. Hussain, "A Review of Sentiment Analysis Research in Chinese Language,"

Cognitive Computation, vol. 9, no. 4, 2017, pp. 423–435.
6. O. Oueslati et al., "A review of sentiment analysis research in Arabic language," *Future Generation Computer Systems*, vol. 112, 2020, pp. 408–430.
7. Y. Li et al., "Popularity Prediction on Vacation Rental Websites," *Neurocomputing*, vol. 412, 2020, pp. 372–380.
8. Y. Li et al., "Learning word representations for sentiment analysis," *Cognitive Computation*, vol. 9, no. 6, 2017, pp. 843–851.
9. A. Picasso et al., "Technical Analysis and Sentiment Embeddings for Market Trend Prediction," *Expert Systems with Applications*, vol. 135, 2019, pp. 60–70.
10. E. Cambria et al., "Sentic Computing for Patient Centered Applications," *IEEE ICSP*, 2010, pp. 1279–1282.
11. J. A. Kumar et al., "Comment Toxicity Detection via a Multichannel Convolutional Bidirectional Gated Recurrent Unit," *Neurocomputing*, vol. 441, 2021, pp. 272–278.
12. A. Khatua, A. Khatua, and E. Cambria, "Predicting Political Sentiments of Voters from Twitter in Multi-party Contexts," *Applied Soft Computing*, vol. 97, no. 106743, 2020.
13. Y. Ma et al., "A Survey on Empathetic Dialogue Systems," *Information Fusion*, vol. 64, 2020, pp. 50–70.
14. L. Stappen et al., "Sentiment Analysis and Topic Recognition in Video Transcriptions," *IEEE Intelligent Systems*, vol. 36, no. 2, 2021, pp. 88–95.
15. E. Cambria et al., "SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis," *CIKM*, 2020, pp. 105–114.
16. M. Dragoni, S. Poria, and E. Cambria, "OntoSenticNet: A commonsense ontology for sentiment analysis," *IEEE Intelligent Systems*, vol. 33, no. 3, 2018, pp. 77–85.
17. Y. Susanto et al., "Ten Years of Sentic Computing," *Cognitive Computation*, vol. 13, 2021.
18. M. Fernández-López, A. Gómez-Pérez, and N. Juristo, "Methontology: from ontological art towards ontological engineering," *Proc. Symposium on Ontological Engineering of AAAI*, 1997, pp. 33–40.
19. C. Strapparava and A. Valitutti, "WordNet-Affect: An Affective Extension of WordNet," *LREC*, 2004, pp. 1083–1086.
20. P. Singh et al., "Open Mind Common Sense: Knowledge Acquisition from the General Public." LNCS, vol. 2519, Springer, 2002, pp. 1223–1237.
21. E. Cambria et al., "GECKA: Game Engine for Commonsense Knowledge Acquisition," *AAAI FLAIRS*, 2015, pp. 282–287.
22. E. Cambria, D. Olsher, and K. Kwok, "Sentic Activation: A Two-Level Affective Common Sense Reasoning Framework," *AAAI*, 2012, pp. 186–192.
23. N. Howard and E. Cambria, "Sentic Neurons: Expanding Intention Awareness," US Patent 10,846,601, 2020.
24. Y. Susanto et al., "The Hourglass Model Revisited," *IEEE Intelligent Systems*, vol. 35, no. 5, 2020, pp. 96–102.
25. A. Valdivia et al., "Consensus vote models for detecting and filtering neutrality in sentiment analysis," *Information Fusion*, vol. 44, 2018, pp. 126–135.
26. M. S. Akhtar, A. Ekbal, and E. Cambria, "How Intense Are You? Predicting Intensities of Emotions and Sentiments using Stacked Ensemble," *IEEE Computational Intelligence Magazine*, vol. 15, no. 1, 2020, pp. 64–75.
27. M. Dragoni, A. Tettamanzi, and C. da Costa Pereira, "Propagating and Aggregating Fuzzy Polarities for Concept-Level Sentiment Analysis," *Cognitive Computation*, vol. 7, no. 2, 2015, pp. 186–97.
28. I. Chaturvedi et al., "Fuzzy Commonsense Reasoning for Multimodal Sentiment Analysis," *Pattern Recognition Letters*, vol. 125, no. 264-270, 2019.

Mauro Dragoni is currently a Research Scientist at Fondazione Bruno Kessler, Trento, Italy. His main research interests are AI and Digital Health with a focus on knowledge management, information retrieval, and machine learning. He received the Ph.D. degree in Computer Science from the University of Milan, Italy. Contact him at dragoni@fbk.eu.

Ivan Donadello is currently an Assistant Researcher at Free University of Bozen-Bolzano, Bolzano, Italy. His research focuses on neural-symbolic algorithms and pure machine-learning approaches for healthcare, explainable AI and predictive construction processes. He received the Ph.D. degree in Computer Science from University of Trento and Fondazione Bruno Kessler, Italy. Contact him at ivan.donadello@unibz.it.

Erik Cambria is the corresponding author and is currently an Associate Professor with Nanyang Technological University, Singapore. His main research interests are AI and affective computing. He received the Ph.D. degree in computing science and mathematics through a joint programme between the University of Stirling, UK, and MIT Media Lab, USA. Contact him at cambria@ntu.edu.sg.