Distante, D.; Faralli, S.; Rittinghaus, S.; Rosso, P.; Samsami, N. (2022). DomainSenticNet: An Ontology and a Methodology Enabling Domain-aware Sentic Computing. Cognitive Computation. 14(1):62-77. https://doi.org/10.1007/s12559-021-09825-w

Noname manuscript No.
(will be inserted by the editor)

DomainSenticNet: An Ontology and a Methodology Enabling Domain-aware Sentic computing

Damiano Distante · Stefano Faralli · Steve Rittinghaus · Paolo Rosso · Nima Samsami

Received: date / Accepted: date

Abstract Background: In recent years, *SenticNet* and *OntoSenticNet* have represented important developments in the novel interdisciplinary field of research known as *Sentic Computing*, enabling the development of a variety of *Sentic* applications. In the present paper, we propose an extension of the *OntoSenticNet* ontology named DomainSenticNet, and contribute an unsupervised methodology aimed to support the development of domain-aware *Sentic* applications.
Methods: We developed an unsupervised methodology that, for each concept in *OntoSenticNet*, mines semantically related concepts from *WordNet* and *Probase* knowledge bases, and computes domain distributional information from the entire collection of *Kickstarter* domain-specific crowdfunding campaigns. Subsequently, we applied DomainSenticNet to a prototype tool for Kickstarter campaign authoring and success prediction and demonstrated improvement in the interpretability of sentiment intensities.
Results and Conclusions: DomainSenticNet is an extension of the *OntoSenticNet* ontology that integrates each of the 100,000 concepts included in *OntoSenticNet* with a set of semantically related concepts and domain distributional information. The defined unsupervised methodology is highly repli-

Damiano Distante
University of Rome Unitelma Sapienza E-mail: damiano.distante@unitelmasapienza.it

Stefano Faralli (corresponding author)
University of Rome Unitelma Sapienza E-mail: stefano.faralli@unitelmasapienza.it

Steve Rittinghaus
Independent researcher, Freelancer Digital Transformation, Baden-Württemberg, Germany E-mail: steverittinghaus@gmail.com

Paolo Rosso
Universitat Politècnica de València E-mail: prosso@dsic.upv.es

Nima Samsami
Independent researcher, Software Architect, Baden-Württemberg, Germany E-mail: nima.samsami@outlook.com

cable and can be easily adapted to build similar domain-aware resources from different domain corpora and external knowledge bases. Used in combination with *OntoSenticNet*, DomainSenticNet may favor the development of novel hybrid aspect-based sentiment analysis systems and support further research on *Sentic Computing* in domain-aware applications.

Keywords Sentic Computing, SenticNet, OntoSenticNet, Kickstarter, interpretability, opinion mining, marketing.

1 Introduction

In the last decades, the Internet has become the preferred communication channel for novel forms of everyday human activities. As recently highlighted by the unfortunate global situation caused by the Covid-19 pandemic, people are now able to perform new activities online to replace or complement traditional behaviors. Popular examples of new forms of activity domains include: e-learning, e-commerce, telehealth, telemedicine, social media, and e-government. Within this context, the majority of the above mentioned sectors and fields of research are benefiting from analyses of popular opinions and sentiments that are massively and extensively conveyed over the Internet, via user generated contents. To support this, researchers are investigating and developing methodologies of Aspect-based Sentiment Analysis (ABSA). As reported by recent surveys [10, 13, 12], the literature on ABSA has identified many open challenges to be solved. The authors of [14] hold that state-of-the-art ABSA approaches can be broadly categorized into *symbolic* and *sub-symbolic* approaches. *Symbolic* approaches "consist of machine learning techniques that perform sentiment classification based on word co-occurrence frequencies". *Sub-symbolic* approaches, on the other hand, "include the use of lexicons, ontologies, and semantic networks to encode the polarity associated with words and multiword expressions". In both cases, ABSA "is a suitcase research problem" [10] that requires many natural language processing (NLP) challenges to be overcome.

In this paper, we introduce DomainSenticNet, an extension of the *OntoSenticNet* ontology [14] aimed at favoring the development of hybrid ABSA systems by leveraging the advantages of both *symbolic* and *sub-symbolic* approaches. It is a resource written in OWL - the W3C Web Ontology Language standard - that, for each of the 100,000 *OntoSenticNet* concepts, provides a set of semantically related concepts and domain distributional information. Specifically, to build DomainSenticNet, for each of the concepts in *OntoSenticNet*, we mined semantically related concepts from the knowledge bases *WordNet* [18] and *Probase* [33], and obtained domain distributional information by computing the distribution of occurrences and co-occurrences of the concept across domain specific texts extracted from textual descriptions of the entire collection of *Kickstarter*[1] crowdfunding campaigns.

[1] `https://www.kickstarter.com/`.

The present paper describes the unsupervised methodology we designed to build our resource, which can be replicated to generate similar resources from different domain corpora and external knowledge bases. Therefore, DomainSenticNet, used in combination with *OntoSenticNet*, can support future investigations of *Sentic Computing* [7] for domain-aware research and applications. Moreover, in this paper, we discuss the practical usage of our resource and present an example of a real application that is able to provide a high level of interpretability of sentiment intensities expressed for domain aspects.

The remainder of the paper is organized as follows: Section 2 states our research objectives; Section 3 describes DomainSenticNet and the unsupervised methodology we designed to construct it from the external knowledge bases *WordNet* [18] and *Probase* [33], and the textual description of *Kickstarter* crowdfunding campaigns; Section 4 describes an example of a real application that, drawing on DomainSenticNet, demonstrates improved interpretability of aspect-based sentiment analysis outcomes; Section 5 summarizes the existing literature, related to our work and, finally, Section 6 provides concluding remarks.

The DomainSenticNet project page is available at `https://github.com/needindex/domainsenticnet`. The related resources are publicly available under *Attribution 4.0 International (CC BY 4.0)*.[2]

2 Research objectives

OntoSenticNet [14] is a commonsense ontology for sentiment analysis based on *SenticNet*, a semantic network of 100,000 concepts. In the present paper, our main research objective was to provide an extension (not a substitution) of *OntoSenticNet* in order to:

- RO1: provide a wider coverage of domain specific concepts (not yet included in *SenticNet*) to support the development of novel hybrid (symbolic and sub-symbolic) domain-specific *SenticNet*-based ABSA systems;
- RO2: include, for each concept, effective and human readable information on the domain pertinence; and
- RO3: use standard knowledge representation language to ease the adoption and reuse of our *OntoSenticNet* extension.

Additionally, with respect to the methodology, we had one further research objective:

- RO4: to define a replicable (and generalized) methodology that could be adapted with minimal efforts to cover additional concepts and domains.

In the following section (Section 3), we describe the resource and the methodology.

[2] `https://creativecommons.org/licenses/by/4.0/deed.en`.

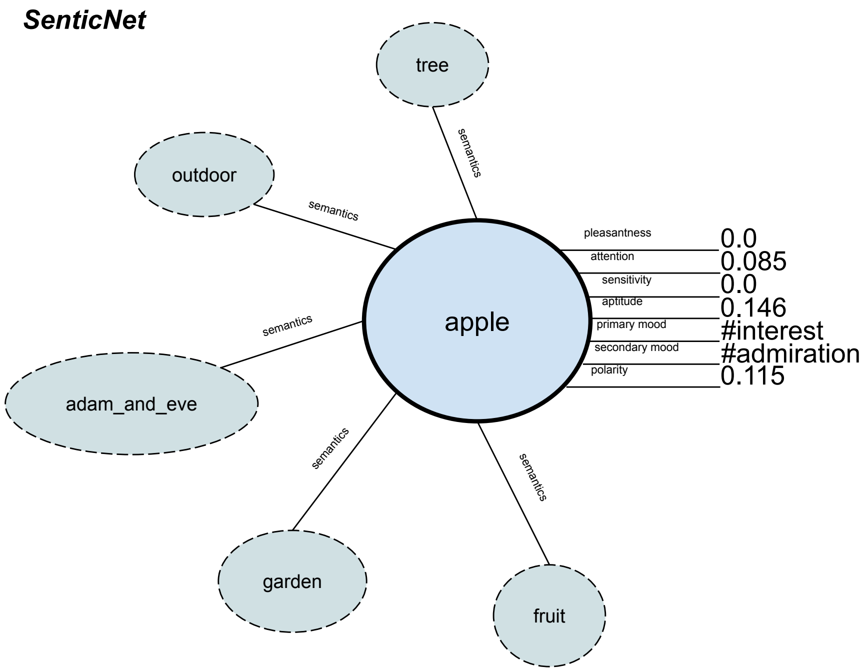

Fig. 1 Representation of the *SenticNet* concept "apple".

3 DomainSenticNet resource and methodology

In this section, we introduce DomainSenticNet and describe the unsupervised methodology we defined to create the resource.

DomainSenticNet is a resource that extends *OntoSenticNet* with:

1. additional related concepts harvested from external knowledge bases;
2. distributional information, i.e., occurrences and co-occurrences of each *SenticNet* concept and related concepts, in domain-related texts.

To illustrate the characteristics of our resource, in Figure 1 we visually represent the original *SenticNet* concept "apple" as a graph. In this graph, nodes represent *SenticNet* concepts and edges represent semantic relatedness between pairs of concepts. Figure 2 shows a visual representation of the corresponding "apple" concept in DomainSenticNet. In this figure, additional nodes represent semantically related concepts mined from external knowledge bases and edges are complemented by domain distributional information about occurrences and co-occurrences in domain texts.

Figure 3 depicts the methodology workflow we designed and performed to generate the DomainSenticNet resource. The methodology included four main steps:

– Step 1: expansion (see Section 3.1);
– Step 2: mining of domain corpora (see Section 3.2);

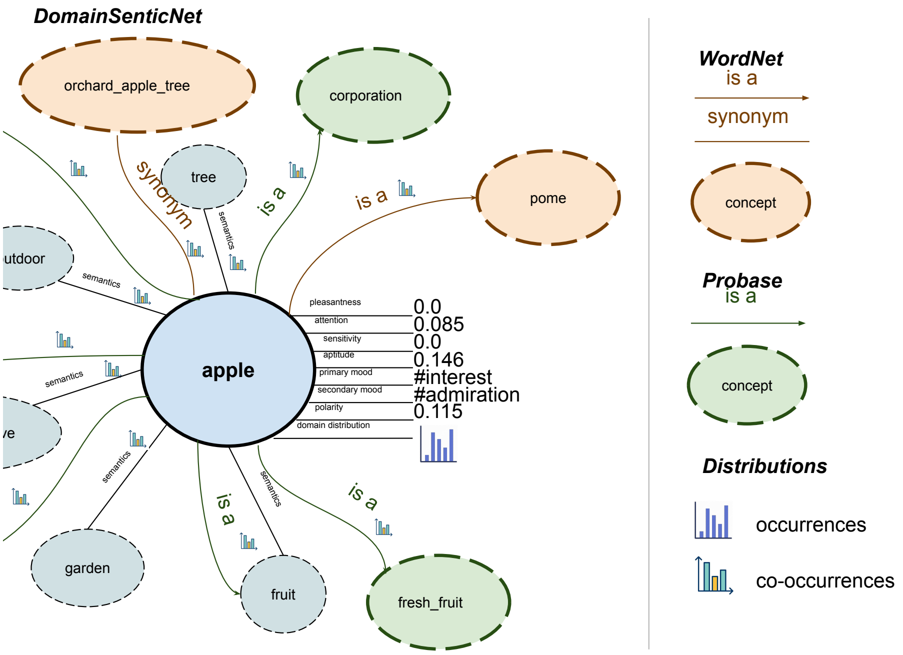

Fig. 2 Excerpt of the DomainSenticNet concept "apple".

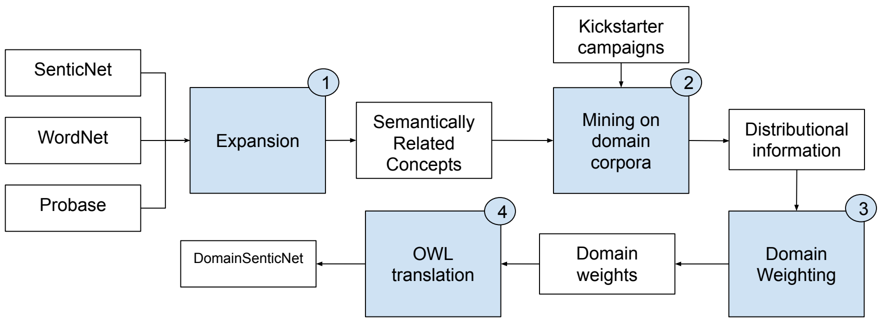

Fig. 3 DomainSenticNet construction workflow.

– Step 3: domain weighting (see Section 3.3);
– Step 4: OWL translation (see Section 3.4).

In the following sections, we describe each of the four steps and, without loss of generality, make explicit reference to the external knowledge bases and corpora used to generate DomainSenticNet.

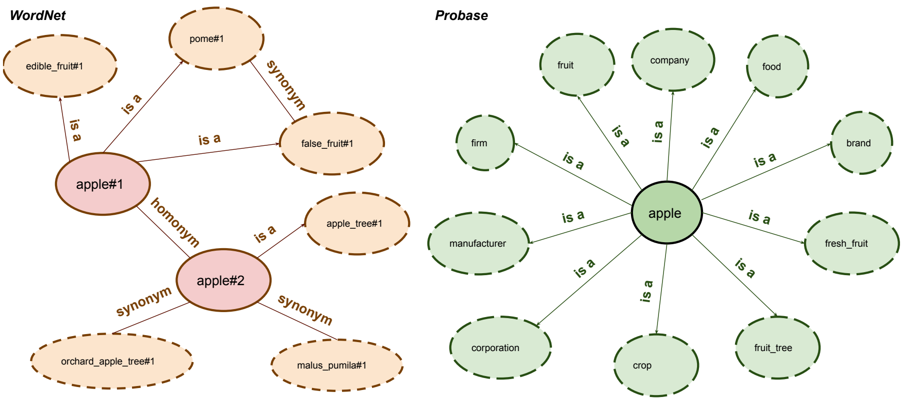

Fig. 4 Excerpt example of the semantically related concepts and relations considered during the expansion step (step 1) for the *SenticNet* concept "apple".

3.1 Expansion

In order to address our first research objective (see Section 2, RO1), in the first step of our workflow, for each concept $\in$ *SenticNet*, we searched for semantically related concepts in the external external knowledge bases *WordNet* [18] and *Probase* [35]. In both knowledge bases, we first identified all concepts corresponding to those in *SenticNet*. Then, in order to collect all neighborhood concepts, for each identified concept, we performed a 1-hop visit on the corresponding knowledge graphs, following the hypernymy ("is a") and synonymy relationships. Figure 4 shows an excerpt of the semantically related concepts we found for the "apple" *SenticNet* concept. For this concept, we first identified the concepts "apple#1" and "apple#2" in *WordNet*, and "apple" in *Probase*. Subsequently, we collected two synonyms (i.e., "malus pumila" and "orchard apple tree") and four hypernyms (i.e., "apple tree," "edible fruit," "false fruit" and "pome") from *WordNet*, and ~4.6K hypernyms (e.g., "brand," "corporation", "company," "crop," "firm," "food," "fresh fruit," "fruit," "fruit tree," "manufacturer," ...) from *Probase*.

3.2 Mining of domain corpora

Distributional information was at the base of our second research objective (see Section 2, RO2). To tackle this objective, we applied standard text mining techniques on domain-specific corpora, to compute: i) the number of occurrences of concepts belonging to *SenticNet*; and ii) the number of co-occurrences of each concept in *SenticNet* and the semantically related external concepts we previously harvested in Step 1 (see Section 3.1). As a medium size collection of domain-specific texts, *Kickstarter* was chosen as a data source.[3]

[3] Monthly updated dataset of Kickstarter campaign URLs available at: `https://webrobots.io/kickstarter-datasets/`.

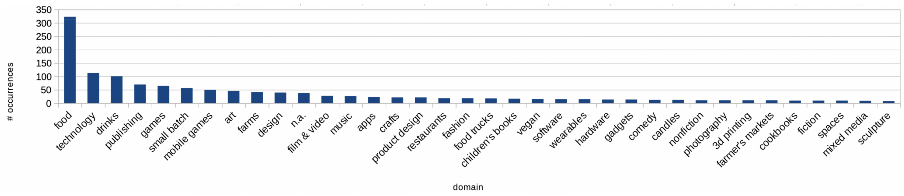

Fig. 5 Occurrence distribution (top 36 domains) of the word "apple" in domain corpora extracted from Kickstarter campaigns.

Kickstarter, a popular source for data scientists, includes approximately 480K campaign descriptions[4] in the form of hypertexts including text, images, videos, and hyperlinks.[5] To identify the domains of interest of each campaign, we leveraged the labels available on the *Kickstarter* platform to categorize each campaign description. In Table 1, we present an excerpt of the 15 main domain categories of Kickstarter, with related sub-categories.[6] The number of occurrences and co-occurrences was computed in four sub-steps:

– Step 2.1: starting from the campaign URLs, we retrieved campaign textual descriptions by means of a custom-made crawler;
– Step 2.2: for each word w corresponding to one of the concepts generated in Step 1 (see Section 3.1) and for each textual campaign description t, we computed the number of occurrences $occ(w, t)$ of the word w in t;
– Step 2.3: for each campaign description t and for each pair of word $\{w_1, w_2\}$ s.t. $occ(w_1, t) > 0$ and $occ(w_2, t) > 0$, we computed the number of co-occurrences $co\_occ(w_1, w_2, t)$ of words w_1 and w_2 in the description t, as $co\_occ(w_1, w_2, t) = occ(w_1, t) * occ(w_2, t)$;
– Step 2.4: since *Kickstarter* campaigns are labeled with two domain categories (i.e., a main category and an optional sub-category), we leveraged this labeling to compute the distributions of occurrences and co-occurrences of concepts across domains.

Returning to the "apple" concept example, Figure 5 depicts the distribution of occurrences of the word "apple" over each resulting domain corpus; Figure 6 presents the co-occurrences distribution for the pair of words "apple" and "brand".

3.3 Domain weighting

Since most distributional methodologies perform better using normalized weights, to complete our second research objective (see Section 2, RO2), we defined a

[4] Real-time statistics accessible at: `https://www.kickstarter.com/help/stats`.
[5] We were able to crawl a total of ∼230K Kickstarter descriptions from the original ∼480K campaigns.
[6] An overview of the respective domains and related statistics is available at: https://www.kickstarter.com/help/stats.

Table 1 Excerpt of *Kickstarter* campaign domains of interest (categories) and sub-domains (sub-categories) (February 2020).

Main category	**Sub-categories**
Art	Ceramics, Conceptual Art, Digital Art, Illustration, Installations, Mixed Media, Painting, Performance Art, Public Art, Sculpture, Textiles, Video Art
Comics	Anthologies, Comic Books, Events, Graphic Novels, Webcomics
Crafts	Candles, Crochet, DIY, Embroidery, Glass, Knitting, Pottery, Printing, Quilts, Stationery, Taxidermy, Weaving, Woodworking
Dance	All Dance Projects, Performances, Residencies, Spaces, Workshops
Design	Architecture, Civic Design, Graphic Design, Interactive Design, Product Design, Typography
Fashion	Accessories, Apparel, Childrenswear, Couture, Footwear, Jewelry, Pet Fashion, Ready-to-wear
Film & Video	Projects, Action, Animation, Comedy, Documentary, Drama, Experimental, Family, Fantasy, Festivals, Horror, Movie Theaters, Music Videos, Narrative Film, Romance, Science Fiction, Shorts, Television, Thrillers, Webseries
Food	Bacon, Community Gardens, Cookbooks, Drinks, Events, Farmer's Markets, Farms, Food Trucks, Restaurants, Small Batch, Spaces, Vegan
Games	Gaming Hardware, Live Games, Mobile Games, Playing Cards, Puzzles, Tabletop Games,Video Games
Journalism	Audio,Photo,Print,Video,Web
Music	blacks, Chiptune, Classical Music, Comedy, Country & Folk, Electronic Music, Faith, Hip-Hop, Indie Rock, Jazz, Kids, Latin, Metal, Pop, Punk, R&B, Rock, World Music
Photography	Animals, Fine Art, Nature, People, Photobooks, Places
Publishing	Academic, Anthologies, Art Books, Calendars, Children's Books, Comedy, Fiction, Letterpress, Literary Journals, Literary Spaces, Nonfiction, Periodicals, Poetry, Radio $ Podcasts, Translations, Young Adult, Zines
Technology	3D Printing, Apps, Camera Equipment, DIY Electronics, Fabrication Tools, Flight, Gadgets, Hardware, Makerspaces, Robots, Software, Sound, Space Exploration, Wearables, Web
Theater	Comedy, Experimental, Festivals, Immersive, Musical, Plays, Spaces

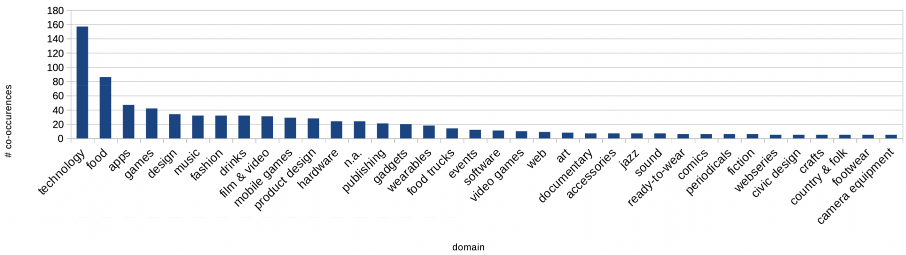

Fig. 6 Co-occurrence distribution (top-36 domains) of the words "apple" and "brand" in domain corpora extracted from Kickstarter campaigns.

proper transformation to obtain correct domain distributional information in the third step of our workflow. To this end, we defined a domain relevance function that assigned each *SenticNet* concept w a domain relevance with respect to a corpus C_d. The function was defined as follows:

$$domainOccScore(w, C_d) = \frac{\sum_{t \in C_d} occ(w, t)}{|C_d|} \tag{1}$$

where C_d included all textual descriptions of the *Kickstarter* campaigns labeled with a specific domain category d.

Additionally, in order to represent the domain relevance of a pair of related concepts $\{w_1, w_2\}$ we defined:

$$domainCooccScore(w_1, w_2, C_d) = \frac{\sum_{t \in C_d} co\_occ(w_1, w_2, t)}{|C_d|} \tag{2}$$

Continuing the "apple" concept example, Figure 7 shows the domain distribution of the *domainOccScore* for the concept "apple", and Figure 8 presents

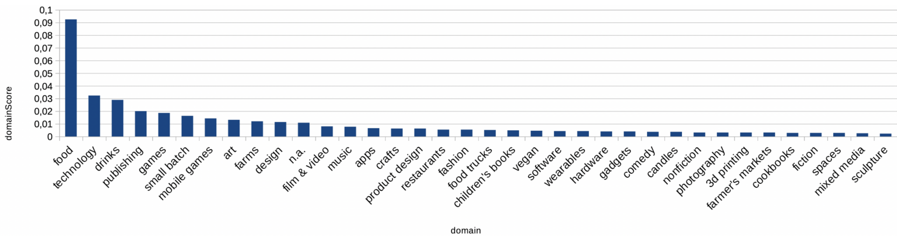

Fig. 7 $domainOccScore(w, C_d)$ distribution ($d \in$ set of top 36 domains) for the DomainSenticNet concept $w =$ "apple".

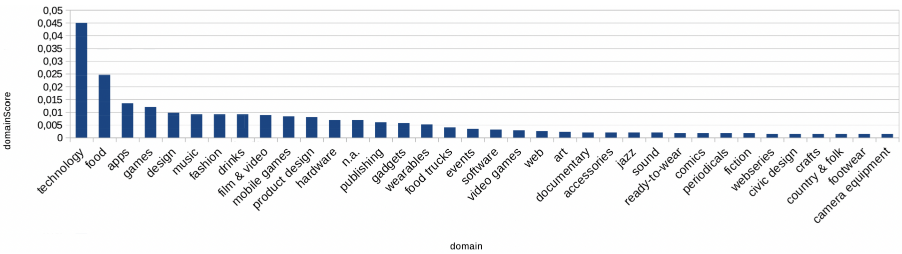

Fig. 8 $domainCooccScore(w_1, w_2, C_d)$ distribution ($d \in$ set of top 36 domains) for the DomainSenticNet concepts $w_1 =$ "apple" and $w_2 =$ "brand".

Table 2 Top 40 most co-occurring concepts across domains (DCS = *domainCooccScore*)

Rank	Concept	Domain	DCS	Rank	Concept	Domain	DCS
1	product	technology	0.46	21	product	hardware	0.11
2	work	technology	0.34	22	choice	technology	0.10
3	device	technology	0.34	23	device	hardware	0.09
4	food	food	0.28	24	being	games	0.09
5	being	technology	0.19	25	work	mobile games	0.09
6	system	technology	0.17	26	name	games	0.09
7	work	games	0.15	27	product	food	0.09
8	case	technology	0.15	28	store	technology	0.09
9	name	technology	0.14	29	work	publishing	0.09
10	good	technology	0.14	30	device	gadgets	0.08
11	platform	technology	0.13	31	service	technology	0.08
12	website	technology	0.12	32	work	food	0.08
13	player	games	0.12	33	product	wearables	0.08
14	product	gadgets	0.12	34	good	food	0.08
15	idea	technology	0.12	35	player	mobile games	0.08
16	business	food	0.12	36	work	hardware	0.08
17	business	technology	0.11	37	fruit	food	0.07
18	company	technology	0.11	38	food	food trucks	0.07
19	computer	technology	0.11	39	name	food	0.07
20	work	film and video	0.11	40	product	design	0.07

the domain distribution of *domainCooccScore* for the two semantically related concepts "apple" and "brand". Finally, in Table 2, we provide the top 40 most co-occurring concepts with "apple," across domains.

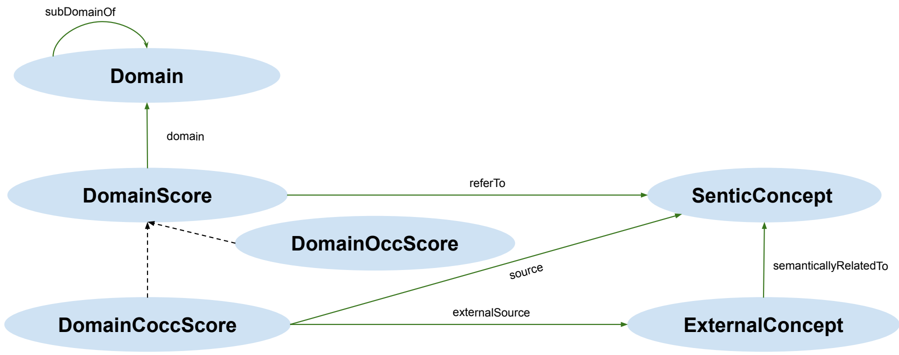

Fig. 9 Overview of the DomainSenticNet scheme.

3.4 OWL translation

In order to address the third research objective (see Section 2, RO3), in the fourth step of our workflow (see Figure 3, block 4), we translated all collected domain distributional information into an OWL representation. As shown in the ontology schema depicted in Figure 9, DomainSenticNet refers to the original definition of *SenticConcept*, thus enabling reference to all original *OntoSenticNet* facts.

As an example, in *OntoSenticNet* [14], the concept "apple" is defined as follows:

```
<owl:NamedIndividual rdf:about="urn:absolute:ontosenticnet#apple">
    <rdf:type rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
    <aptitude rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.146</aptitude>
    <attention rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.085</attention>
    <pleasantness rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0</pleasantness>
    <polarity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0.077</polarity>
    <polarityText>positive</polarityText>
    <primitiveURI rdf:resource="urn:absolute:ontosenticnet#admiration"/>
    <primitiveURI rdf:resource="urn:absolute:ontosenticnet#interest"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#adam_and_eve"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#fruit"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#garden"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#outdoor"/>
    <semantics rdf:resource="urn:absolute:ontosenticnet#tree"/>
    <sensitivity rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">0</sensitivity>
    <text>apple</text>
</owl:NamedIndividual>
```

where: i) *aptitude*, *attention*, *pleasantness*, and *sensitivity* are defined as *SenticValues* for the corresponding *Hourglass of Emotions* model dimensions; ii) *polarity* is the overall sentiment polarity; iii) *semantics* are properties representing five semantically related concepts (e.g., *adam_and_eve*, *fruit*, *garden*, *outdoor* and *tree*); and iv) *primitiveURI* refers to two primitive moods (e.g., *admiration* and *interest*).

In order to represent all of the concepts mined from the external knowledge bases in the first step (see Figure 3, block 1), we defined the "ExternalConcept" class as follows:

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#ExternalConcept"/>
```

The above class enables the model to reference concepts such as the "malus pumila", which *WordNet* presents as a synonym of the *SenticNet* concept "apple". Instances of the "ExternalConcept" class have two annotation properties, namely *provenance* and *text*, which represent the source knowledge base and the lexeme, respectively:

```
<owl:AnnotationProperty rdf:about="urn:absolute:ontosenticnet#provenance"/>
<owl:AnnotationProperty rdf:about="urn:absolute:ontosenticnet#text"/>
```

As an example, the external concept "malus pumila" is defined as:

```
<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#malus_pumila">
    <rdf:type rdf:resource="urn:absolute:domainsenticnet#ExternalConcept"/>
    <provenance>WordNet</provenance>
    <text>malus pumila</text>
    <semanticallyRelatedTo  rdf:resource="urn:absolute:ontosenticnet#apple"/>
    ...
</owl:NamedIndividual>
```

where *semanticallyRelatedTo* is an *ObjectProperty* defined as:

```
<owl:ObjectProperty rdf:ID="semanticallyRelatedTo">
  <rdfs:domain rdf:resource="#ExternalConcept"/>
  <rdfs:range  rdf:resource="#SenticConcept"/>
</owl:ObjectProperty>
```

In order to represent each of the 176 considered domains, we defined the following *Domain* class:

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#Domain"/>
```

The 15 main categories and 161 sub-categories were then defined as subclasses of the *Domain* class.

```
<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#subDomainOf">
    <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#Domain"/>
    <rdfs:range rdf:resource="urn:absolute:domainsenticnet#Domain"/>
</owl:ObjectProperty>
```

As an example, the resulting definition for the domain "Ceramics" includes the annotation property *subDomainOf*, representing the fact that "Ceramics" is a sub-domain of "Art".

```
<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#domain_ceramics">
    <rdf:type rdf:resource="urn:absolute:domainsenticnet#Domain"/>
    <subDomainOf rdf:resource="urn:absolute:domainsenticnet#domain_art"/>
    <text>Ceramics</text>
</owl:NamedIndividual>
```

To represent the domain weights described in Section 3.3, we provided the definitions for *DomainScore*, *DomainOccScore*, and *DomainCooccScore* classes, as follows:

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainScore"/>

<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainOccScore">
    <rdfs:subClassOf rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
</owl:Class>
```

```
<owl:Class rdf:about="urn:absolute:domainsenticnet#DomainCooccScore">
    <rdfs:subClassOf rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
</owl:Class>
```

The datatype property *score* represents a numeric weight:

```
<owl:DatatypeProperty rdf:about="urn:absolute:domainsenticnet#score">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#double"/>
</owl:DatatypeProperty>
```

and the following object property *domain* represents the domain related to a score:

```
<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#domain">
    <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
    <rdfs:range rdf:resource="urn:absolute:domainsenticnet#Domain"/>
</owl:ObjectProperty>
```

Finally, the object properties *referTo*, *source*, and *externalSource* bind a *DomainScore* to one or more *SenticConcept*s or *ExternalConcept*s:

```
<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#referTo">
    <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainScore"/>
    <rdfs:range rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#source">
    <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
    <rdfs:range rdf:resource="urn:absolute:ontosenticnet#SenticConcept"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:about="urn:absolute:domainsenticnet#externalSource">
    <rdfs:domain rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
    <rdfs:range rdf:resource="urn:absolute:domainsenticnet#ExternalConcept"/>
</owl:ObjectProperty>
```

As an example, the $domainOccScore$("apple", D_{food}), defined in Section 3.3, is represented as follows:

```
<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#dos_apple_51">
    <rdf:type rdf:resource="urn:absolute:domainsenticnet#DomainOccScore"/>
    <referTo rdf:resource="urn:absolute:ontosenticnet#apple"/>
    <domain rdf:resource="urn:absolute:domainsenticnet#domain_food"/>
    <score>0.0924971363115693</score>
</owl:NamedIndividual>
```

and the $domainCooccScore$("apple", "company", $D_{technology}$), defined in Section 3.3, is represented as follows:

```
<owl:NamedIndividual rdf:about="urn:absolute:domainsenticnet#dcs_apple_2047">
    <rdf:type rdf:resource="urn:absolute:domainsenticnet#DomainCooccScore"/>
    <referTo rdf:resource="urn:absolute:ontosenticnet#apple"/>
    <externalSource rdf:resource="urn:absolute:domainsenticnet#company"/>
    <domain rdf:resource="urn:absolute:domainsenticnet#domain_technology"/>
    <score>0.1145475372279496</score>
</owl:NamedIndividual>
```

3.5 Results

DomainSenticNet was the result of our investigations aimed at achieving research objectives RO1, RO2, and RO3 (see Section 2).

The proposed approach was the result of RO4 (see Section 2), which primarely aimed at defining a generalized methodology that could be easily adapted to cover additional concepts and domains. In fact, the methodology is able to generate similar resources by simply using different domain corpora and external knowledge bases as input (see Figure 3). Moreover, the methodology can be used to provide both domain distributional information and OWL representations for semantic networks other than *OntoSenticNet*, such as *DBpedia* and *WebIsADB* [17].

DomainSenticNet can be enhanced as a dynamic resource[7] in two ways:

1. by integrating significant variations in concept collections and domain distribution of occurrences and co-occurrences linked to future releases of domain corpora and external knowledge bases; and
2. by including timestamps (e.g., campaign start times) of domain corpora (e.g., dumps of Kickstarter campaign URLs[8]) or other references to specific time in a temporal dimension in domain distributional information.

To address the above-mentioned dynamicity, we created a project Web page[9] and established a maintenance schedule for the generation of time based update releases.

4 Domain-aware Kickstarter campaign success prediction with DomainSenticNet

In this section, we present an example application of DomainSenticNet.

GameOn [16] is a prototype application designed to support the authoring of successful crowdfunding campaigns in Kickstarter.

The main characteristics of *GameOn*[10] are:

- it automatically induces (by means of clustering) a partition of semantically related domain aspects mined from user-generated product and service reviews, with each cluster representing an “influencing factor” for the campaign success;[11]
- it employs *SenticNet* to perform ABSA and to identify emotional intensities expressed in textual campaign descriptions for the above mentioned domain aspects;

[7] Real-time data are widely recognized as the life blood of a variety of applications (e.g., [10]).

[8] `https://webrobots.io/kickstarter-datasets/`.

[9] `https://github.com/needindex/domainsenticnet`.

[10] `https://github.com/needindex/gameon`.

[11] It is worth noticing that the tool can also process human crafted partitions of domain aspects.

– it aggregates the above-mentioned emotional intensities into a statistical index (*NeedIndex*), which: i) identifies the most influencing factors in campaign success; ii) calibrates an Objective and Key Result scale (OKR)[12] to interpret *NeedIndex*es, through the identification of low and high emotional intensity bounds, delimiting *low*, *medium* and *high* emotional intensity states;
– it leverages DomainSenticNet to further tune (for a given domain of interest) the OKR scale of interpretations for emotional intensities.

Figure 10 presents a screenshot of the application's graphical user interface. Each time a user inputs a campaign description, composed of both a textual description and a funding goal, success prediction is performed. The figure refers to an example campaign in the domain of "mobile games". In order to clearly explain the prediction outcome, the application additionally reports, for each influencing factor, an overview of the emotional intensities expressed in the textual description (see Figure 10, part B). The above mentioned overview includes the *NeedIndex* values computed for each influencing factor (Figure 10, part C).

Finally, the application compares the computed *NeedIndex*es against the average of the corresponding indexes of successful "mobile games" campaigns in the past 3 seasons (see Figure 10, parts B and C). Therefore, in this application, *NeedIndex*es are used to both train the model for campaign success forecasting and to provide highly interpretable explanations of the prediction outcomes. *NeedIndex*es are thus effective indicators used by the application to suggest actions to be performed on the textual descriptions to refine the emotional intensities expressed with respect to influencing factors (i.e., clusters).

Using DomainSenticNet, the application is also able to provide a domain adaption (at a cluster level) of *NeedIndex* OKR scales of interpretation, whereby the resulting states of emotional intensities are calibrated with respect to the *domainOccScore* (defined in Section 3) for the "mobile games" domain.

To convey the previously mentioned calibration of OKR scales, Figure 11 presents two OKR scales related to the interpretation of emotional intensities. The top part of the figure shows the original OKR scale (not adapted to the domain of interest), wherein two threshold values (i.e., 0.3 and 0.5) represent the lower and upper bounds used to identify the range of *NeedIndex*es values corresponding to a *medium* emotional intensity level. In contrast, the bottom part of the figure depicts the domain-adapted scale with the corresponding bounds for the cluster labeled "education," wherein the adapted medium level is bounded by the thresholds 0.22 and 0.43. For each cluster, the relevant bounds are obtained by computing the average *domainOccScore*s of concepts (in the cluster) occurring in unsuccessful and successful campaign descriptions, respectively.

[12] OKR models are commonly used by very successful companies such as Amazon, Facebook and Google. `https://www.whatmatters.com/faqs/how-to-grade-okrs` `https://conceptboard.com/blog/okr-google-goal-setting-success/`.

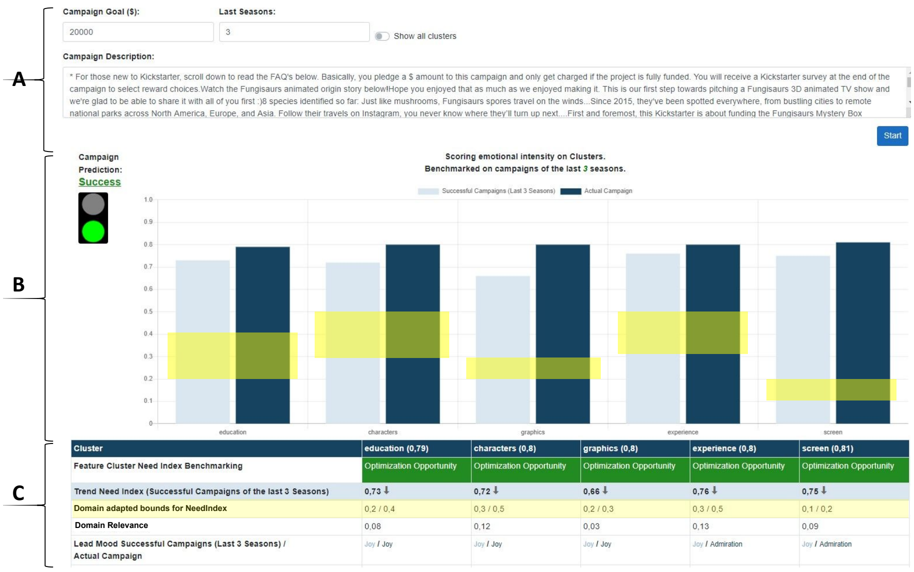

Fig. 10 A screenshot of the *GameOn* user interface.

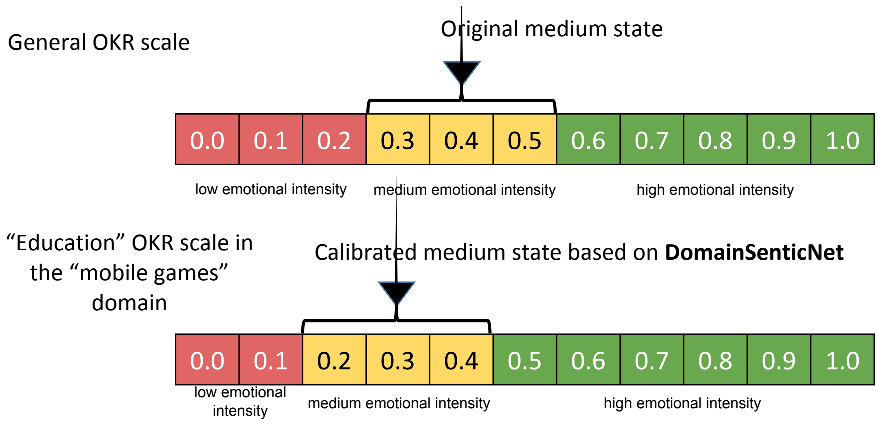

Fig. 11 The original (top) and domain-adapted (bottom) version of the OKR scale for the influencing factor “education” in the “mobile games” domain.

Figure 10 part C, shows both the “Domain adapted NeedIndex Bounds” and “Domain Relevance”. Domain adapted emotional intensity states reflect both the average emotional intensity and the domain relevance for successful and unsuccessful campaigns, respectively.

In the case of the "education" cluster[13], the *medium* emotional intensity state produced lower values for two main reasons: i) in the considered Kickstarter dataset, the emotional intensities provided for the corresponding influential factors in the "mobile games" domain were lower than the average observed over the previous three seasons with respect to other aspects; and ii) the average *domainOccScore* of the corresponding aspects indicated a lower domain pertinence.

5 Related Works

In this paper, we have presented DomainSeticNet as a resource to extend *OntoSenticNet*, a state-of-the-art commonsense ontology [14].

OntoSenticNet is an ontological representation of *SenticNet* [11], which is a resource resulting from the combined application of symbolic and sub-symbolic artificial intelligence methodologies to automatically discover conceptual primitives from text and link them to commonsense concepts and named entities. *SenticNet* includes the definition of 100K concepts (called *SenticConcept*).[14] Each *SenticConcept* (see Figure 1 for a visual representation of the concept "apple") is defined by: i) a multiword expression; ii) the weights for the four dimensions of the *Hourglass of Emotions* model [28] (i.e., *pleasantness*, *attention*, *sensitivity*, and *aptitude*); iii) primary and secondary mood labels (e.g., "#interest","#admiration"); iv) a polarity score; and v) a collection of five semantically related *SenticConcept*.

OntoSenticNet is an ontological definition of the semantic network induced by the 100K *SenticConcept*s. Its main characteristic is its ability to provide a precise conceptual hierarchy, including associated concepts and sentiment values. Hence, *OntoSenticNet* is a preferential resource for developing state-of-the-art applications of sentiment analysis based on *SenticNet*.

In recent years, *SenticNet* and *OntoSenticNet* have represented important developments for research. In particular, the findings from Cambria's research group have enabled a novel interdisciplinary field of research known as *Sentic Computing* [7]. Within *Sentic Computing*, many successful investigations have generated novel insights in the domains of knowledge representation [2], deep learning-based ABSA [24], business intelligence [19], social media marketing [6], recommender systems [3] and financial forecasting [37], to name only a few.

In the remainder of this section, we summarize the relevant literature pertaining to key aspects of the definition and construction of DomainSenticNet resource.

[13] The education cluster groups the following aspects: "education," "student," "school," "college," "instruction," "classroom," "brain," "growth," "level,", "course," "knowledge," "career," "tutorial," "education," "lecture," "tutor," "teacher," "learning," "teaching," and "skill".

[14] SenticNet 6 has recently been released. This updated resource now contains 200K concepts [8].

In constructing the proposed resource, with the aim of collecting neighborhood semantically related concepts from external knowledge graphs, we applied basic graph mining techniques (as described in Section 3.1). In general, the task of collecting semantically related concepts from affordable or noisy automatically acquired external knowledge graphs can be performed by sophisticated approaches (see [26] for a recent survey). As an example, the authors of [29] experimented with similarity expansion-based techniques and obtained high levels of efficiency and precision in the task of extending new concepts in a given knowledge base.

As already mentioned, the backbone of DomainSenticNet is the *OntoSenticNet* ontological description of *SenticNet*. One of the key characteristics of *SenticNet* is that, all concepts are defined with valued attributes derived from the *Hourglass of Emotions* model [9].[15] Therefore, *SenticNet* is considered an appropriate knowledge base for the development of human interpretable sentiment analysis approaches.

The availability of the above mentioned resources is beneficial to all Ontology-Driven Sentiment Analysis (ODSA)-based applications. Specifically, the authors of [4] recently surveyed works applying ODSA to customer reviews. Furthermore, as an example of an ODSA-based approach, the authors of [25] presented a hybrid solution for sentence-level ABSA, using a lexicalized domain ontology in combination with neural attention networks.

Researchers in this field are also exploring the creation of new resources to be leveraged in ODSA-based applications. As an example, in [23], the authors presented a methodology to extend ontologies in the "Materials Science" domain. The presented approach leveraged the titles and abstracts of 600 domain publications and complemented a given ontology with additional concepts and axioms, by means of a phrase-based topic model approach. In a similar direction, the authors of [38] proposed the addition of SOBA—a semi-automated methodology to generate ontologies—to ODSA applications.

In contrast to works mentioned above, our methodology (see Section 3) is unsupervised and can be easily adapted to include other external knowledge bases and multiple domain corpora. In this way, it automatically generates a high coverage of domain relevant concepts (not included in *OntoSenticNet*) and related distributional information for an arbitrarily defined set of domains of interest. Additionally, the present paper discussed a real application that benefited from the availability of DomainSenticNet, in terms of both sentiment analysis performance and ease of interpretation (see Section 4).

As discussed in the Introduction (see Section 1), DomainSenticNet is suitable for use in domain-aware sentiment analysis applications. Such applications have recently been improved, due to advancements in semi-supervised learning [15] and more specifically in semi-supervised learning for social data analysis [20, 5]. Researchers are experimenting with semi-supervised learning a potentially more robust solution to problems such as, word polarity disambiguation [36] and extraction of actionable information from unstructured text

[15] A recent model revision is described in [34].

[21]. As an example, in [22], the authors presented a deep learning approach named $ConvNet - SVM_{BoVW}$ for fine-grained sentiment analysis. The model combined textual and visual features built on a convolution neural network (ConvNet) enhanced with the contextual scoring mechanism of SentiCircle [30]. The proposed model was able to perform sentiment polarity classification with 91% accuracy. Moreover, in [1], the authors recently provided a Stacked Ensemble-based methodology to assess the emotional intensities in texts related to a general domain, and performed sentiment analysis in the financial domain. With respect to the two above mentioned studies, and in line with the findings of [32], the distributional information of DomainSenticNet may be coupled with contextual semantic features to address the problem of word polarity disambiguation. Finally, our resource may also be leveraged to improve the interpretability and explainability of sentiment analysis outcomes (see Section 4, in which we discuss these two properties through a real application).

6 Conclusions

This paper has presented DomainSenticNet—a resource that extends the *OntoSenticNet* commonsense ontology with: i) additional related concepts harvested from external knowledge bases; and ii) distributional information on the occurrences and co-occurrences of each *OntoSenticNet* concept and related concepts in domain corpora. The paper has also described the methodology we adopted to generate DomainSenticNet. This methodology can be easily adapted to process different domain corpora and external knowledge bases, in order to generate domain-aware resources similar to ours and to extend semantic networks other than *OntoSenticNet*. Therefore, it can also enable the computation of domain-adapted scales of interpretation to benchmark domain ABSA application outcomes (as shown in Section 4).

To provide a concrete example of the benefit of DomainSenticNet to a variety of applications, we described a prototype tool for successful Kickstarter campaign authoring and campaign success prediction. Specifically, we discussed the high human interpretability level of both the prediction outcomes and the changes suggested for campaign descriptions, in order to improve the likelihood of success. Moreover, the domain distributional information provided by DomainSenticNet enables it to produce domain-adapted scales of interpretation for predictive features at the level of influencing factors.

Regarding resource dynamicity (discussed in Section 3.5), we identify two opportunities: i) integrating updated releases (including new portions of the domain corpus), and ii) extending the current DomainSenticNet ontology schema with the inclusion of a time dimension. Additional dynamicity can be further leveraged by means of applying the proposed methodology (see Section 3) to other application-specific corpora. For instance, in the e-commerce domain, product and service reviews can be leveraged to capture the dynamics and trends of emotional intensities within customer opinion statements. Therefore, DomainSenticNet provides a basis for further interdisciplinary research within *behavioral economics*, *applied data sciences* and *applied mathematics*, with the aim of increasing the resource *"dynamicity"* to apply to an unlimited range of applications.

Additionally, in order to address the above mentioned interdisciplinary investigations, we aim at studying the effectiveness of *causal inference* approaches such as the *DoWhy* [31] framework. The *DoWhy* framework can be leveraged to gain insight into cause and effect relationships when domain adaption is applied. Such insights can then support the development and the interpretation of calculated domain-aware emotional intensity weights. Specifically, we are interested in the ability of the *DoWhy* approach in identifying the correlation magnitude of unexploited features in classification models [27], thus enabling, for example, the magnitude of missing domain concepts to be determined.[16]

The current version of DomainSenticNet does not include sentiment polarities for *ExternalConcepts*; instead, it references *OntoSenticNet* for *SenticConcept* sentiment polarities. Therefore, another possible future research might aim at "propagating" the *Hourglass of Emotions* dimension weights and polarities to a collection of added external concepts. In addition, similar to [11, 8], our resource opens an avenue for further research on the generation of contextual domain embeddings in deep neural network-based applications. Finally, as discussed in Section 5, approaches such as [1, 21, 22] can leverage DomainSenticNet as an effective resource to improve interpretability and explainability in domain-aware sentic applications.

Compliance with Ethical Standards:

The present work did not involve any research with human participants or animals performed by any of the authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Akhtar, M.S., Ekbal, A., Cambria, E.: How intense are you? predicting intensities of emotions and sentiments using stacked ensemble [application notes]. Computer Intelligence Magazine **15**(1), 64–75 (2020). DOI 10.1109/MCI.2019.2954667. URL `https://doi.org/10.1109/MCI.2019.2954667`
2. Alhussien, I., Cambria, E., NengSheng, Z.: Semantically enhanced models for commonsense knowledge acquisition. In: 2018 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1014–1021. November 17-20, Singapore (2018). DOI 10.1109/ICDMW.2018.00146. URL `https://doi.org/10.1109/ICDMW.2018.00146`

[16] `https://microsoft.github.io/dowhy/dowhy_confounder_example.html`.

3. Angulo, C., Falomir, I.Z., Anguita, D., Agell, N., Cambria, E.: Bridging cognitive models and recommender systems. Cognitive Computation **12**(2), 426–427 (2020). DOI 10.1007/s12559-020-09719-3. URL `https://doi.org/10.1007/s12559-020-09719-3`
4. Bandari, S., Bulusu, V.V.: Survey on ontology-based sentiment analysis of customer reviews for products and services. In: K.S. Raju, R. Senkerik, S.P. Lanka, V. Rajagopal (eds.) Data Engineering and Communication Technology, vol. 1079, pp. 91–101. Springer Singapore, Singapore (2020). DOI 10.1007/978-981-15-1097-7_8. URL `https://doi.org/10.1007/978-981-15-1097-7_8`
5. Billal, B., Fonseca, A., Sadat, F., Lounis, H.: Semi-supervised learning and social media text analysis towards multi-labeling categorization. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 1907–1916. December 11-14, Boston, MA, USA (2017). DOI 10.1109/BigData.2017.8258136. URL `https://doi.org/10.1109/BigData.2017.8258136`
6. Cambria, E., Grassi, M., Hussain, A., Havasi, C.: Sentic computing for social media marketing. Multimedia Tools and Applications **59**(2), 557–577 (2012). DOI 10.1007/s11042-011-0815-0. URL `https://doi.org/10.1007/s11042-011-0815-0`
7. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: Sentic Computing: Exploitation of Common Sense for the Development of Emotion-Sensitive Systems, *Lecture Notes in Computer Science*, vol. 5967, pp. 148–156. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). DOI 10.1007/978-3-642-12397-9_12. URL `https://doi.org/10.1007/978-3-642-12397-9_12`
8. Cambria, E., Li, Y., Xing, F.Z., Poria, S., Kwok, K.: Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, CIKM '20, p. 105–114. Association for Computing Machinery, New York, NY, USA (2020). DOI 10.1145/3340531.3412003. URL `https://doi.org/10.1145/3340531.3412003`
9. Cambria, E., Livingstone, A., Hussain, A.: The hourglass of emotions. In: A. Esposito, A.M. Esposito, A. Vinciarelli, R. Hoffmann, V.C. Müller (eds.) Cognitive Behavioural Systems, COST 2012 International Training School, vol. 7403, pp. 144–157. Springer Berlin Heidelberg (2012). DOI 10.1007/978-3-642-34584-5_11. URL `https://doi.org/10.1007/978-3-642-34584-5_11`
10. Cambria, E., Poria, S., Gelbukh, A., Thelwall, M.: Sentiment analysis is a big suitcase. IEEE Intelligent Systems **32**(06), 74–80 (2017). DOI 10.1109/MIS.2017.4531228. URL `https://doi.ieeecomputersociety.org/10.1109/MIS.2017.4531228`
11. Cambria, E., Poria, S., Hazarika, D., Kwok, K.: Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In: S.A. McIlraith, K.Q. Weinberger (eds.) Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), pp. 1795–1802. AAAI Press, New Orleans, Louisiana, USA (2018). URL `https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16839`
12. Chakraborty, K., Bhattacharyya, S., Bag, R.: A survey of sentiment analysis from social media data. IEEE Transactions on Computational Social Systems **7**(2), 450–464 (2020). DOI 10.1109/TCSS.2019.2956957. URL `https://doi.org/10.1109/TCSS.2019.2956957`
13. Chauhan, G.S., Meena, Y.K.: Domsent: Domain-specific aspect term extraction in aspect-based sentiment analysis. In: A.K. Somani, R.S. Shekhawat, A. Mundra, S. Srivastava, V.K. Verma (eds.) Smart Systems and IoT: Innovations in Computing, vol. 141, pp. 103–109. Springer Singapore, Singapore (2020). DOI 10.1007/978-981-13-8406-6_11. URL `https://doi.org/10.1007/978-981-13-8406-6_11`
14. Dragoni, M., Poria, S., Cambria, E.: Ontosenticnet: A commonsense ontology for sentiment analysis. IEEE Intelligent Systems **33**, 77–85 (2018). DOI 10.1109/MIS.2018.033001419. URL `https://doi.org/10.1109/MIS.2018.033001419`
15. van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine Learning **109**(2), 373–440 (2020). DOI 10.1007/s10994-019-05855-6. URL `https://doi.org/10.1007/s10994-019-05855-6`
16. Faralli, S., Rittinghaus, S., Samsami, N., Distante, D., Rocha, E.: Emotional intensity-based success prediction model for crowdfunded campaigns. Information Processing & Management **58**(1), article ID 102394 (2021). DOI doi.org/10.1016/j.ipm.2020.102394. URL `https://doi.org/10.1016/j.ipm.2020.102394`

17. Faralli, S., Velardi, P., Yusifli, F.: Multiple knowledge GraphDB (MKGDB). In: Proceedings of The 12th Language Resources and Evaluation Conference, pp. 2325–2331. European Language Resources Association, Marseille, France (2020). URL https://www.aclweb.org/anthology/2020.lrec-1.283
18. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. Language, Speech, and Communication. MIT Press, Cambridge, MA (1998)
19. Fernandez-Breis, J.T., Qazi, A., Raj, R.G., Tahir, M., Cambria, E., Syed, K.B.S.: Enhancing business intelligence by means of suggestive reviews. The Scientific World Journal **vol. 2014**, article ID 879323 (2014). DOI 10.1155/2014/879323. URL https://doi.org/10.1155/2014/879323
20. Hussain, A., Cambria, E.: Semi-supervised learning for big social data analysis. Neurocomputing **275**, 1662 – 1673 (2018). DOI 10.1016/j.neucom.2017.10.010. URL https://doi.org/10.1016/j.neucom.2017.10.010
21. Khatua, A., Khatua, A., Cambria, E.: A tale of two epidemics: Contextual word2vec for classifying twitter streams during outbreaks. Information Processing & Management **56**(1), 247 – 257 (2019). DOI 10.1016/j.ipm.2018.10.010. URL https://doi.org/10.1016/j.ipm.2018.10.010
22. Kumar, A., Srinivasan, K., Cheng, W.H., Zomaya, A.Y.: Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. Information Processing & Management **57**(1), article ID 102141 (2020). DOI 10.1016/j.ipm.2019.102141. URL https://doi.org/10.1016/j.ipm.2019.102141
23. Li, H., Armiento, R., Lambrix, P.: A method for extending ontologies with application to the materials science domain. Data Science Journal **18**, 1–21 (2019). DOI 10.5334/dsj-2019-050. URL https://doi.org/10.5334/dsj-2019-050
24. Ma, Y., Peng, H., Cambria, E.: Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In: AAAI Conference on Artificial Intelligence, pp. 5876–5883 (2018). URL https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16541
25. Meškelė, D., Frasincar, F.: Aldonar: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalized domain ontology and a regularized neural attention model. Information Processing & Management **57**(3), article ID 102211 (2020). DOI 10.1016/j.ipm.2020.102211. URL https://doi.org/10.1016/j.ipm.2020.102211
26. Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web **8**(3), 489–508 (2017). DOI 10.3233/SW-160218. URL https://doi.org/10.3233/SW-160218
27. Pearl, J., Mackenzie, D.: The Book of Why. Basic Books, New York (2018). URL https://dl.acm.org/doi/book/10.5555/3238230
28. Plutchik, R.: The nature of emotions. American scientist **89**(4), 344–350 (2001). URL https://www.jstor.org/stable/27857503
29. Rajagopal, D., Cambria, E., Olsher, D., Kwok, K.: A graph-based approach to commonsense concept extraction and semantic similarity detection. In: Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion, p. 565–570. Association for Computing Machinery, New York, NY, USA (2013). DOI 10.1145/2487788.2487995. URL https://doi.org/10.1145/2487788.2487995
30. Saif, H., Fernandez, M., He, Y., Alani, H.: Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In: V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, A. Tordai (eds.) The Semantic Web: Trends and Challenges, pp. 83–98. Springer International Publishing, Cham (2014). DOI 10.1007/978-3-319-07443-6_7. URL https://doi.org/10.1007/978-3-319-07443-6_7
31. Sharma, A., Kiciman, E.: DoWhy: A Python package for causal inference (2019). URL https://github.com/microsoft/dowhy
32. Shiller, R.: Narrative economics. American Economic Review **107**, 967–1004 (2017). DOI 10.1257/aer.107.4.967. URL https://doi.org/10.1257/aer.107.4.967
33. Song, Y., Wang, H., Wang, Z., Li, H., Chen, W.: Short text conceptualization using a probabilistic knowledgebase. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Three, IJCAI'11, p. 2330–2336. AAAI Press, Barcelona, Catalonia, Spain (2011)

34. Susanto, Y., Livingstone, A.G., Ng, B.C., Cambria, E.: The hourglass model revisited. IEEE Intelligent Systems **35**(5), 96–102 (2020). DOI 10.1109/MIS.2020.2992799. URL `https://doi.org/10.1109/MIS.2020.2992799`
35. Wu, W., Li, H., Wang, H., Zhu, K.Q.: Probase: A probabilistic taxonomy for text understanding. In: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12, p. 481–492. Association for Computing Machinery, New York, NY, USA (2012). DOI 10.1145/2213836.2213891. URL `https://doi.org/10.1145/2213836.2213891`
36. Xia, Y., Cambria, E., Hussain, A., Zhao, H.: Word polarity disambiguation using bayesian model and opinion-level features. Cognitive Computation **7**(3), 369–380 (2015). DOI 10.1007/s12559-014-9298-4. URL `https://doi.org/10.1007/s12559-014-9298-4`
37. Xing, F.Z., Cambria, E., Welsch, R.E.: Natural language based financial forecasting: a survey. Artificial Intelligence Review **50**(1), 49–73 (2018). DOI 10.1007/s10462-017-9588-9. URL `https://doi.org/10.1007/s10462-017-9588-9`
38. Zhuang, L., Schouten, K., Frasincar, F.: Soba: Semi-automated ontology builder for aspect-based sentiment analysis. Journal of Web Semantics **60**, article ID 100544 (2019). DOI 10.1016/j.websem.2019.100544. URL `https://doi.org/10.1016/j.websem.2019.100544`