Semantic Web 0 (0) 1–12
IOS Press

Human Affective States Ontology for Sentiment Analysis

Rana Abaalkhail [a,b,*], Benjamin Guthier [c] and Abdulmotaleb El Saddik [a]
[a] *Multimedia Communications Research Laboratory, University of Ottawa, 800 King Edward Ave, K1N 6N5, Ottawa, ON, Canada*
E-mail: {rabaa006, elsaddik}@uottawa.ca
[b] *King Saud University, Information System, Riyadh, Kingdom of Saud Arabia*
[c] *University of Mannheim, Department of Computer Science IV, Mannheim, Germany*

Abstract. Social media provides a platform where users share an enormous amount of information about events, products, experiences and more. This information may contain user sentiments and feelings. Sentiment analysis helps monitor and analyze the opinions of users. An ontology has the ability to express the concepts shared, as well as their relationships, in a semantically rich representation. This strong feature enables an ontology to be applied in the area of sentiment analysis. In this paper, we propose the development of a Human Affective States Ontology, which we will refer to as HASO. We employ HASO to the problem of sentiment analysis. We argue that this ontology can compete with state of the art machine learning approaches to detect the sentiment contained in textual data. By using HASO, we classify the sentiment found in the SemEval-2017 dataset and compare our results with those obtained by the teams that participated in this task. The results of our work show the effectiveness of the proposed ontology (HASO) in capturing sentiment, especially when compared to machine learning approaches.

Keywords: Ontology, Machine learning, Sentiment Analysis, Emotion

1. Introduction

Micro-blogs such as Twitter are rich sources for sentiment analysis. Opinion mining, or sentiment analysis, is a branch of text mining and is used to determine the polarity of a textual sentence, for example positive, negative or neutral. Sentiment analysis allows us to observe public needs, moods and behaviors. It also helps in many domains such as politics, movie ticket sales, and general customer satisfaction [1]. The purpose of sentiment analysis is to expose the attitude that people have toward a subject or an entity. The analysis of structured and unstructured data plays a noteworthy role in decision making, ranging from movie selection to determine our daily satisfaction needs [2].

Even though lexicon-based and machine learning approaches have earned a strong reputation in the area of sentiment analysis, there still exists a gap in the semantic understanding of textual content. An ontology has the capability of capturing the semantic association between concepts and the relationships within content. With such ability, the manual annotation needed in the machine learning approaches can be eliminated. As a result, the sentiment analysis community is moving towards an ontological approach to represent a common-sense knowledge base [1].

An ontology can be defined as a "branch of philosophy that is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality" [3]. It can also be seen as a catalogue that shows specific field entities and the relationship between them. It can serve as an answer for questions such as "Which class does an entity belong to?". Ontologies help to form structural knowledge about a domain and define a common vocabulary

*Corresponding author. E-Mail: rabaa006@uottawa.ca.

to be shared within that domain. Developing an ontology enables knowledge and information sharing.

In this paper, we examine the capability of the ontological approach against the machine learning algorithms for sentiment analysis on social media by using our proposed Human Affective States Ontology (HASO). We argue that the ontological approach can compete with state of the art machine learning algorithms to capture a more comprehensive sentiment behavior from informal textual contents, specifically in social media.

The remainder of this paper is organized as follows. In Section 2 we present the related work regarding ontology-based sentiment analysis. Moreover, we introduce the emotion-related lexicons and language that we used in the development of HASO. The development and the modularization of HASO is introduced in Section 3. Section 4 describes the HASO sentiment analysis and its performance result. Our conclusions and an outlook are provided in Section 5.

2. Related Work

Ontologies have become more popular in many fields including web technologies and data integration [4]. They help form structural knowledge for many domains, and define a common vocabulary to be shared within a domain. Developing an ontology enables knowledge and information sharing. In addition, it defines a set of data and their structure from the applications in the same domain. Related work in ontology-based sentiment analysis is presented in Section 2.1. The emotion related lexicons and language that we used in the development of HASO is presented in Section 2.2. A more comprehensive survey of existing ontologies on affective states can be found in [5].

2.1. Related Work in Ontology-Based Sentiment

Ontologies have rich semantic representations since they capture the semantic association between concepts and relationships. Consequently, the sentiment analysis (SA) community is moving towards an ontological approach to represent a common-sense knowledge base [1]. An ontology can be designed as domain knowledge. Consequently, certain words can have different polarities depending on the domain and the context. In addition, some domains have special words to express sentiments [6]. An ontology can be developed to analyze sentiments in a specific domain. The Ontology-based Sentiment Analysis Process for Social Media content (OSAPS) is proposed in [7] to identify the problem areas based on customer feedback of postal service delivery issues and generate automated online replies for those issues. They build an ontology model from extracted data (Tweets) and use it to identify issues from the negative sentiments recognized, using SentiStrength. The process includes data cleaning, extracting only a combination of noun and verb tags for query building, and retrieving information from SPARQL Query from the ontology model. A domain ontology for smartphones was created to analyze tweets related to smartphones. The aim of the ontology is to accept tweets as input and provide sentiment analysis in the domain. The ontology consists of smartphone vocabulary. OpenDover's was used to assign a sentiment score to each tweet. OpenDover's is a web service that tags opinions and sentiments in a textual corpus and assigns a sentiment score [-10,10] [8].

An ontology for mobile product sentiment analysis was therefore created. The ontology was created in the OWL format by retrieving the vocabulary and the data features from mobile and online shopping sites. For example: camera is part of vocabulary and zoom capacity is a feature. The feature opinion score is obtained from the Stanford Natural Language Processor tool, which ranges from [-2 to 2]. They built the SPARQL query interface to accept and answer user queries about mobile products. For example, a possible query can be "mobile with good battery life" and from the stored features and opinion scores, the system answers the user's query [9].

An ontology for sentiment analysis of electronic products was also created. The ontology contains a sentiment class that has subclasses for emotion related words (happiness and sadness) and electronic products. The emotion words and the electronic products were extracted from an online customer reviews survey. How Net dictionary was used to calculate the words' semantic similarities. As a result, the ontology can analyze a user query such as "Which tablet PC is excellent? " [10]

On the other hand, an ontology can help analyze sentiments in a free domain. The Emotive Ontology [11]was built to detect and analyze emotions in informal texts obtained from social media. The approach consists of detecting a range of eight high-level emotions: anger, confusion, disgust, fear, happiness, sadness, shame and surprise. The Emotive Ontology is also capable of expressing the intensity of the emotions. During the creation of the Emotive Ontology, many dictionaries and word datasets such as WordNet were consulted. Natural Language Processing (NLP) and part of speech tagging were used as pre-processing steps for emotion detection. The ontology was tested and evaluated on a dataset taken from Twitter. Even though the Emotive Ontology [11]was created to capture emotions from textual content, its performance evaluation is limited for three reasons: 1) the dataset was small, containing only 150 tweets, 2) it was annotated by only two users, 3) the tweet collection was event related.

2.2. Emotion-Related Lexicons and Language

This section describes the emotion dictionaries (or "lexicons") that are used in our ontology, which is presented Section 3.
Emotion dictionaries group lists of words based on different principles such as synonymy or sentiment (positive, negative).

Emotion Markup language (EmotionML) [1] is a general-purpose emotion annotation and representation language that provides a standard emotion representation format. It consists of the emotion vocabularies and their features.

Since the data is annotated in a standard way, the interpretation of the message between systems is the same. EmotionML uses Ekman's, Cowie's, OCC categories, FSRE's, and Frijda's discrete emotions vocabularies. In addition, it illustrates the PAD dimensional model to represent emotions and their features. The language can be applied in different contexts such as data annotation and emotion recognition. The annotation can be applied to text, static images, speech recordings, and video.

WordNet [2] is an online lexicon for the English language that is distributed into five categories: nouns, verbs, adjectives, adverbs, and function words. It clusters words together based on their meanings and defines semantic relations between words. It also groups them into sets of synonyms called synsets. WordNet currently contains 155,287 words that are organized into 117,659 synsets.

To include concepts of affect, **WordNet-Affect** [3] was developed as an extension that labels synsets with emotions, moods and behaviors. WordNet-Affect creates an additional hierarchy in WordNet with emotion labeling. The hierarchy of WordNet-Affect categorizes emotion words into classes such as positive emotion, negative emotion, and neutral emotion . In addition, EmoSenticNet [4] is an expansion of WordNet Affect, broadening the emotion label vocabulary.

SenticNet 3 [5] is a concept-level opinion lexicon for sentiment analysis. It includes polarity for words and multi-word expressions. The polarity can be a number in the range between -1 and 1, or it can be a flag (positive or negative). SenticNet 3 contains 30,000 common and common-sense concepts. It is different from other sentiment analysis resources such as WordNet-Affect, because it associates semantics and sentics with common and common-sense knowledge. Common-sense knowledge can help determine the polarity of a concept in a multi-word expression sentence. This improves subsequent text-based sentiment analyses.

NRC Word-Emotion Association Lexicon [6] which is a list of English words and their association with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

The Harvard General Inquirer [7] is a lexicon attaching syntactic, semantic to part-of-speech tagged words.

MPQA Subjectivity lexicon [8] labels the words with information about their polarity (positive, neutral or negative), as well as the intensity of the polarity (weak or strong).

AFINN [9] is a dictionary that has a list of English words rated for valence with an integer between -5,

[1]http://www.w3.org/TR/emotionml/
[2]https://wordnet.princeton.edu/

[3]https://www.gsi.dit.upm.es/ontologies/wnaffect/
[4]https://www.gelbukh.com/emosenticnet/
[5]http://sentic.net/
[6]http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
[7]http://www.wjh.harvard.edu/ inquirer/
[8] http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
[9]http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

+5, and zero for neutral.

SentiStrength [10] SentiStrength is a dictionary that has a list of English words, especially for short text, rated with an integer between -5, +5, and zero for neutral.

In addition, we used the Oxford English Dictionary [11] to extend the list of synonyms.

Indeed, psychological theories include many vocabularies that represent emotions, moods, and sentiments. In order to make these lexicons usable in software, we describe them in a standardized format. We thus propose the Human Affective States Ontology (HASO) to represent human emotion, mood, and sentiment models. Building an ontology that covers many psychological theories creates opportunities to build semantic applications with different scenarios and purposes.

3. HASO Development

The Human Affective States Ontology **(HASO)** [12] has been developed in the OWL language. It provides knowledge and a common vocabulary regarding human affective states (emotion, mood, sentiment), in a machine-accessible or machine-readable format. Nowadays, humans and computer applications often need to communicate and share knowledge. However, everyone expresses themselves in his or her own language, with different terms and meanings. Ontologies aim to unify the terms and meanings in order to enable effective communication between people and computers. Ontologies capture the domain knowledge and provide an approved understanding of the domain. The study of human emotion, mood, and sentiment is significant as these concepts have an impact on human behavior. Building an ontology for this domain allows us to then build a semantic application [12].

In Section 3.1 we explain the HASO engineering process. In Section 3.2 we then illustrate the HASO modularization process .

[10] http://sentistrength.wlv.ac.uk/
[11] https://en.oxforddictionaries.com/
[12] http://www.mcrlab.net/datasets/

3.1. HASO Engineering

We followed Methontology [13], a methodology that assists users in the creation of ontologies, to build HASO. It contains the ontology's entire life cycle in the development process.

Ontology development requires the determination of the ontology's information resources, called Knowledge Acquisition. Since the proposed ontology represents human affective states, HASO uses psychological theories and existing ontologies from the same domain. Fortunately, the appearance of the affective computing paradigm allows us to use theories and findings from psychology in the development of human affective applications. Indeed, We argue that psychological theories are the primary point of an ontology design in the domain of human affective states. These theories form the basis for the proposed ontology. HASO also uses a lexicon and a thesaurus that cover human affective states as a source of knowledge, as stated in Section 2.

Subsequently, we need to model the ontology's conceptual model, which starts with a glossary of terms, and then groups the terms into classes and subclasses (concepts) and properties (verbs). Figure1 shows the main entities in HASO and the relationships between the entities.

The class *Affective State* represents human affective states. The ones considered here are Emotion, Mood, and Sentiment [5]. The *Affective State Model* represents the psychological models for each affective state [14]. *Affective State* thus has a relationship with the class *Affective State Model* through the "hasModel" relationship. Analogously, *Affective State Model* connects to the *Affective State* class through the "isModelFor" relationship. *Affective State Recognition* represents the ways or methods to detect each affective state. Thus, an *Affective State* "isDetectedFrom" an *Affective State Recognition* method.

After determining the ontology knowledge acquisition and the ontology conceptualization model, we need to look for existing related ontologies and take the advantage of one of the ontology's most valuable features, which is ontology reusing [15]. Indeed, the ability to reuse an ontology is considered to be a significant and valuable feature of ontology engineering. Ontology reuse can take many different forms, such as reusing the top level of an ontology, reusing a smaller

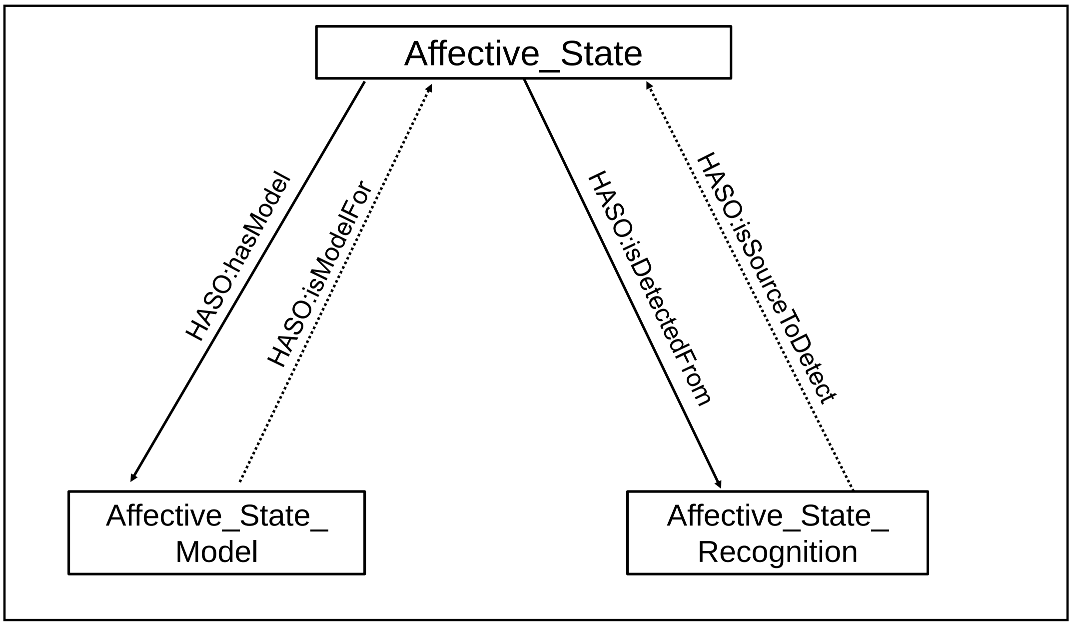

Fig. 1. Conceptual model of the HASO depicting the major entities. Rectangular shapes represent classes, arrows indicate object properties, and dotted arrows indicate the inverse of object properties.

part of an ontology, or extending the existing ontology [16].

Figure 2 shows the imported ontologies and illustrates the links between them and HASO. We reuse HEO because it represents the appraisal OCC model as well as the Frijda Action Tendency model. We also reuse the negative, positive, neutral vocabularies from the WordNet-Affect ontology. Since HASO represents human affective states we connect our ontology with Friend Of A Friend (FOAF). FOAF [13] is an ontology that is used to describe a person, their activities and their relations to other people. It consists of classes that represent a person (first name, family name), gender, age, education, organization, homepage, information about organizational project(s) they are involved in, culture, etc.

An ontology is defined as an "explicit specification of a conceptualization" [17]. It consists of classes, properties and individuals that define a particular domain [18].

Figure 3 shows the main classes of HASO. Humans have *Mental States* and *Physical States*. The *Mental States* that represent the *Affective states* can be divided into three sub-classes: *Emotion*, *Mood*, and *Sentiment*. To express the ways in which the affective states can be represented, we created the *Affective State Model* class. An *Emotion* can be described in a discrete way by using the property "hasCategory", in a dimensional way by using the property "hasDimension", and in a componential way by using the property "hasAppraisal" [14].

When defining an emotion in a discrete way, we introduce subclasses under the *Discrete Emotional Model*. One of the common subclasses is *Basic Emotion Category*. In addition, HASO contains a subclass to express the emotion classified by the model of Douglas-Cowie [19], and reused emotion vocabularies defined by EmotionML. These vocabularies were clustered in groups, but HASO expresses them in the subclasses of *Every Day Emotion Category*, *OCC Emotion Category*, *Frijda Category*, and *FSRE Category*. HASO also includes a *Social Emotion Category* subclass to represent social emotions. Moreover, HASO defines subclasses under the *Discrete Emotional Model* based on Drummond's emotion vocabulary. They are categorized under the *Drummond Category* subclass.

[13] http://www.foaf-project.org/

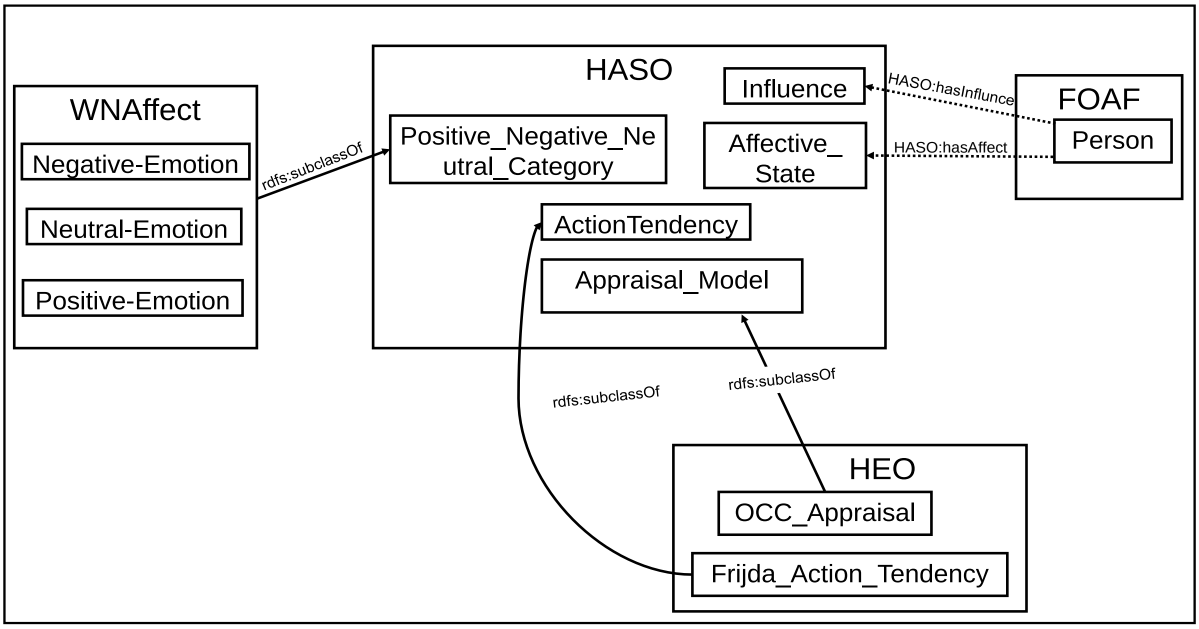

Fig. 2. Linking HASO to the imported ontologies. Rectangular shapes represent classes, solid arrows indicate a subclass relationship, and dotted arrows indicate object properties.

Based on the Drummond Emotion Vocabulary, we created a *LevelOfEmotion* subclass where emotion can be classified as *Light*, *Medium*, or *Strong*. The emotion vocabulary in HASO connects with the mentioned subclass through the property "hasLevel". The *Positive Negative Neutral Category* subclass was created to include vocabularies from WordNet , NRC Word-Emotion Association Lexicon , SenticNet , EmoSenticNet , a Harvard General Inquirer lexicon , MPQA Subjectivity lexicon , AFINN , SentiStrength. In addition, we used the Oxford English Dictionary [14] to extend the list of synonyms. We also represent the vocabularies from the psychology theories in the human affective states domain.

Emotion can be defined by existing dimensional models. HASO represents these models by including the subclasses named *Circumplex Model*, *Fontaine Model*, *PAD Model*, and *Watson and Tellegen Model* under the super-class *Dimensional Emotional Model*. Additionally, an emotion can be expressed by a componential model. HASO represents *OCC Appraisal* as a subclass of the class *Appraisal - Componential Model*. The *OCC Appraisal* subclass was reused from the HEO ontology.

Due to the similarities between mood and emotion, mood can also be expressed by a discrete or a dimensional model[20]. Hence, the *Mood* class connects to the *Discrete Emotional Model* and the *Dimensional Emotional Model* subclasses through the properties "hasCategory", and "hasDimension". Furthermore, *Sentiment* can be represented by a discrete model through the *Positive Negative Neutral Category* subclass.

An emotion can have an action tendency which defines the emotion action outcome. It is expressed through the property "hasActionTendency". HASO defines the *Frijda Action Tendency* as a subclass of the *Action Tendency* class.

To include the ways in which humans express emotion, the *Emotional Expression Cue* super-class was added to HASO. Its sub-classes are *Conductal Emotional Cues*, *Physiological Emotional Cues*, *Textual Emotional Cues*, and *Verbal Emotional Cues*, which is in line with the emotion ontology by Obrenovic et al.[21]. *Conductal Emotional Cues* are further structured into facial expressions, gestures and speech. The

[14]https://en.oxforddictionaries.com/

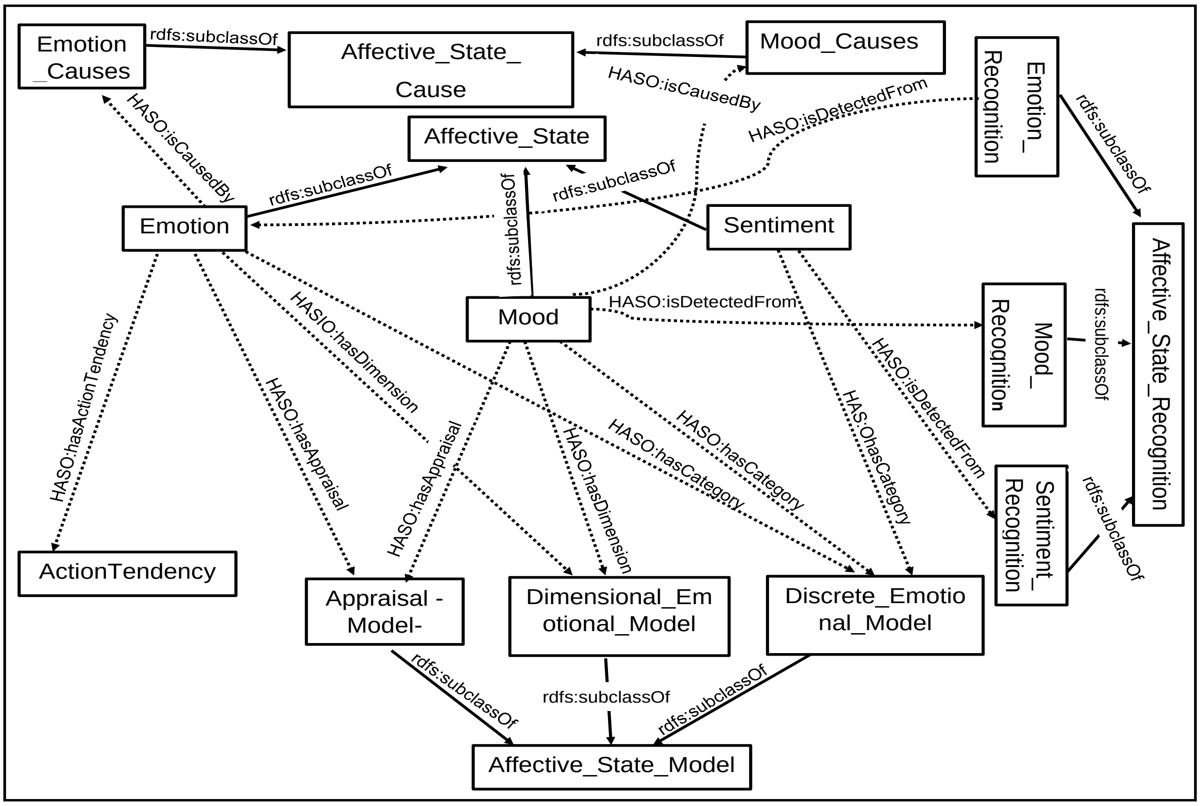

Fig. 3. Graphical representation of the main classes of HASO and the properties. Rectangular shapes represent classes, solid arrows indicate subclass relationships, and dotted arrows indicate object properties.

relationship between *Emotions* and their Emotional Expression Cues is modeled by the property "isExpressedThrough".

To include the causes for mood and emotion, we created the *Affective State Cause* class. *Emotion* and *Mood* are connected to *Emotion Causes* and *Mood Causes*, respectively, through the property "isCausedBy".

HASO proposes an *Affective State Recognition* class to model the possible ways of collecting information to identify human affective states. Hence, *Emotion*, *Mood*, and *Sentiment* connect to *Emotion Recognition*, *Mood Recognition*, and *Sentiment Recognition*, respectively, through the "isDetectedFrom" object property.

As we propose to use HASO for sentiment analysis, we give more details and a visualization of the "Discrete Model" class-subclass. Figure 4 shows the Discrete Model class-subclass Visualization. It shows the hierarchy, object properties, and data properties. The Positive Negative Neutral Category and Psychological Theories are subclasses of Discrete Model. The Psychological Theories are: Basic Emotion Category, Douglas-Cowie Category, Drummond Category, Every Day Emotion Category, Frijda Category, FSRE Category, OCC Emotion Category, and Social Emotion Category.

The "Affective State Annotations" class was created with three individuals: negative affective state, positive affective state, and neutral affective state. Moreover, we added individuals for each "Discrete Model" subclass, which are the words that were extracted from the dictionaries, the lexicons, and the psychological theories. Each individual can belong to more than one subclass, particularly one of them is the Positive Word, Negative Word, or Neutral Word subclass. In addition, we created a data property "strength score" with a range of `xsd:integer` $[-5, 5]$. We extracted all the individual strength scores from SentiStrength and AFINN. For example, the individual "happy" has a strength score of 2, the individual "sad" has a

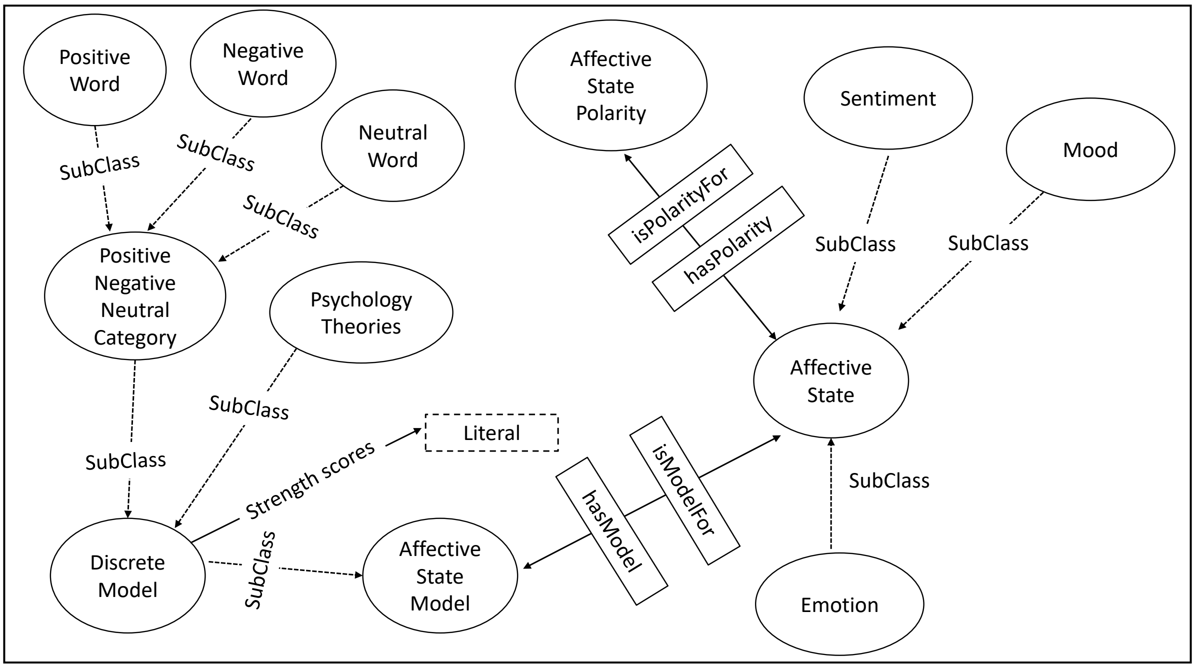

Fig. 4. Discrete Model class-subclass Visualization. Oval shape represents class , and subclass , rectangle shape represents object property, and round dot rectangle represents data property.

strength score of -4, and the individual "scholarship" has a strength score of 0. Because HASO covers a wide range of words, if we don't find the individual word's strength score in any of the lexicons, then we assign the strength score of one of its synonyms that is mentioned in one of the lexicons. HASO can be downloaded under the link: http://www.mcrlab.net/wp-content/uploads/2018/06/Proposed-Ontology-Human-Affective-States-HASO.zip

Table 1

HASO modules (sub-ontology) after the second iteration of modularization

Module	Module Name	Class Count	Properties Count
Original Ontology	HASO	228	182
M1	Affective States Recognition	33	2
M2	Emotion Expression Cues	53	22
M3	Affective States Causes Reaction	35	7
M4	Affective States Dimensional Model	11	8
M5	Affective States Appraisal Model	20	6
M6	Affective States Discrete Model	38	13
M7	Affectional (AFO)	13	6

3.2. HASO Modularization

In many domains, ontologies can have thousands of axioms to cover the domain's concepts and topics. Nevertheless, large ontologies can face problems of scalability, reusability, and validation. Dividing the ontology into self-contained modules that handle subtopics of the large ontology can solve many of these problems. This activity is called Ontology Modularization. According to the NeOn dictionary, Ontology Modularization is the activity that takes as an input an ontology and produces modules for this ontology to support maintenance and reuse [22]. HASO covers a wide range of human affective states and therefore many topics. Through modularization, we create modules that handle parts of the ontology. We follow the tasks that were determined in [23].

Task 1: Identify the Purpose of Modularization since recognizing the need for the modularization is vital and will guide the whole process. The reasons for the modularization of HASO are to facilitate the processes of understanding the whole ontology, of maintenance, of reusability, and of validation. This can be accomplished by producing modules of a manageable size, such that each module covers a subtopic of HASO.

Task 2: Select a Modularization Approach
The approach can be determined based on the modularization purpose. However, the modularization process can be performed in an iterative manner, based on the modularization criteria, which is the next task. We chose the partitioning approach and divided HASO into modules, where each module handles a part of HASO. This process produces modules of a controllable size.

Task 3: Define Modularization Criteria
Based on the purpose of the HASO modularization, we chose the criteria of **local correctness** and **size**[24] Local correctness means that nothing is added to the modules if it was not originally in the ontology. The size of a module refers to the number of classes, properties and individuals which they contain. A module's size can determine its future maintainability.

Task 4: Select a Base Modularization Technique
There are many techniques and tools for ontology modularization. We chose to use the "copy/move/delete axioms" functionality in protégé to achieve HASO modularization.

Task 5: Combine the Results
Combine the Results. In this step, modules are generated using the "copy/move/delete axioms" functionality as a first iteration.

Task 6: Evaluate the Modularization
Modules were evaluated against the HASO modularization criteria. The evaluation outcome determines whether or not a new iteration needs to be carried out. It can also indicate whether or not another modularization approach should be applied. After evaluating the HASO modules, we find that some modules contradict the purpose of the modularization. As a result, a final iteration was performed using the approaches of partitioning and extraction. At the end of the second iteration, the evaluation yielded satisfactory results.

Task 7: Finalize Modularization
When the outcome of the evaluation is acceptable, the output is all of the modules that were produced from the modularization process. Table 1 shows the modules of HASO after applying the second iteration of modularization. HASO modularization modules can be downloaded under the link: http://www.mcrlab.net/wp-content/uploads/2018/06/Proposed-Ontology-Human-Affective-States-HASO-Modularization.zip

We evaluate HASO and its modularization modules by using OOPS![15], a web-based tool that scans for major pitfalls. We also verified the consistency by running a Pellet reasoner.

4. Sentiment analysis based on HASO

The massive amount of content generated on a daily basis by users on social media such as Twitter provides a significant opportunity to monitor and analyze public emotional responses to events. We employ HASO to classify the sentiment polarity on the test data for SemEval-2017 Task 4, Subtask A[16] .

The mission in Subtask A was: "Message Polarity Classification: Given a message, classify whether the message is a positive, negative, or neutral sentiment". The dataset contained general data without a specific topic. As mentioned in Section 3.1, The HASO integrates emotion vocabularies from psychological theories as well as from a wide range of language lexicons and thesauruses in order to ensure that a large set of emotion related vocabulary was covered. Moreover, commonly encountered emotions within the collected data were added to the ontology. Indeed, understanding the collected data allows for the proper representation in the ontology. Our goal was to demonstrate the effectiveness of our ontology compared to machine leaning approaches.

In Section 4.1 we introduce the tweet polarity calculation algorithm. The results of our ontology-based sentiment analysis are compared with the machine learning method in Section 4.2

4.1. Tweet Polarity Calculation Algorithm

To employ HASO in the sentient analysis area we design and develop a tweet polarity calculation algorithm. As Shown in Figure 5, We divide the tweet polarity calculation algorithm into three blocks: tweets pre processing, tweet processing , and tweet post processing.

[15] http://oops.linkeddata.es/
[16] http://alt.qcri.org/semeval2017/task4/

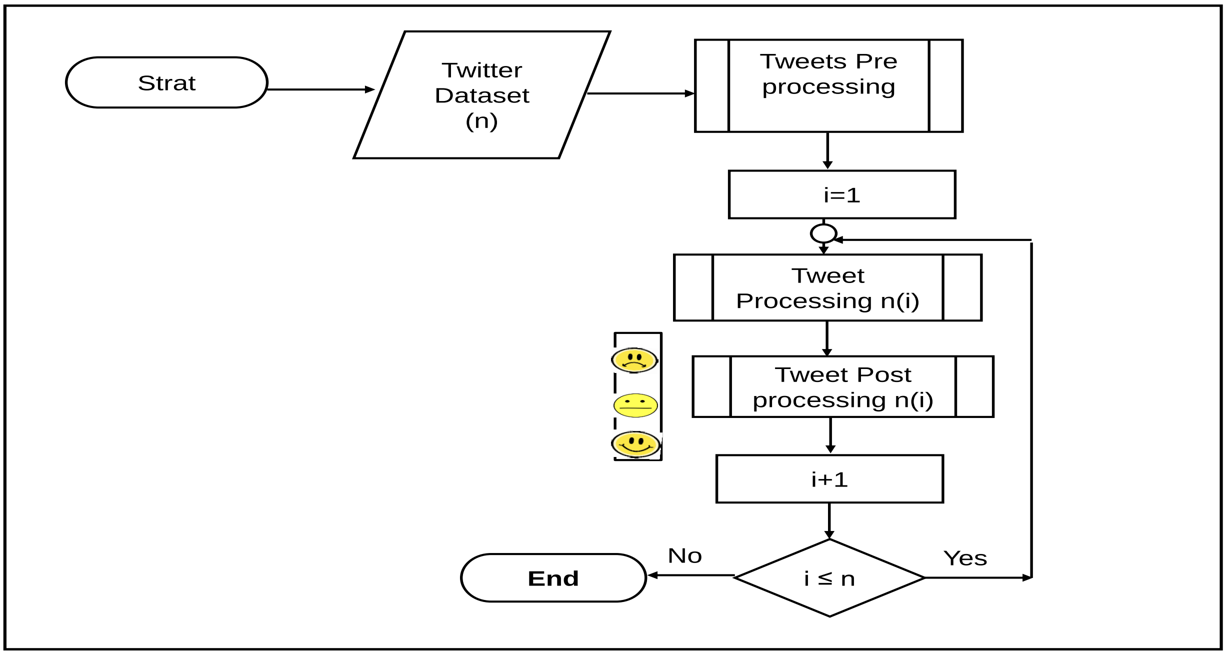

Fig. 5. Tweet polarity calculation algorithm.

in the **tweeet pre-processing** we prepare the twitter dataset by applying a tokenization on the tweets using regular expressions. We took an extra step regarding the data preparation by converting any emojis into text. Since nowadays people frequently express their feelings by using emojis, these symbols have an important effect on the resulting sentiment [25]. We converted the emojis in the data to an equivalent word by using matching rules based on Emoji Sentiment Ranking[17] and Home of Emoji Meaning[18]. In addition, spelling correction was applied to the text and we removed URLs, emails, user handles (@user), and punctuation using regular expression patterns.

Next, the aim of the **tweet processing** is calculating a tweet overall strength score. we applied sentiment analysis on a sentence level. As result, for each tweet we ran the SPARQL query to get the strength score for the tokens that affect the overall sentence sentiment. For example, we ignored the pronouns and the articles in the sentence. If the token did not have a match in the ontology, then we ran the SPARQL query for the current token with the next token. We used the parameterized SPARQL query to apply a query against the ontology (HASO) . We also used the Jena Java API [19], which is a Java framework that provides support for manipulating and querying RDF models as shown in Figure 6. We then calculated the tweet strength score by adding the biggest value from the positive tokens and the smallest value from the negative tokens in the sentence which clarified in Figure 7. We followed the SentiStrength method to calculate the overall sentence's score [26].

Finaly, in the **tweet post processing**, we determine the tweet polarity based on the tweet strength score from the tweet processing block. If the tweet strength score is larger than or equal to 1, then the sentence polarity is positive. If the tweet strength score is less than or equal to -1, then the sentence polarity is negative. Otherwise, the polarity is neutral. Figure shows the tweet post processing process.

Table 2 shows some examples of tweet sentiment analysis results using HASO. In the presented examples, we ran the SPARQL query for each word (token) to get the sentiment strength. The query result may return a positive number, a negative number, or zero based on how we present the word in the ontology. In

[17]http://kt.ijs.si/data/Emoji_sentiment_ranking/
[18]https://emojipedia.org/
[19]https://jena.apache.org/

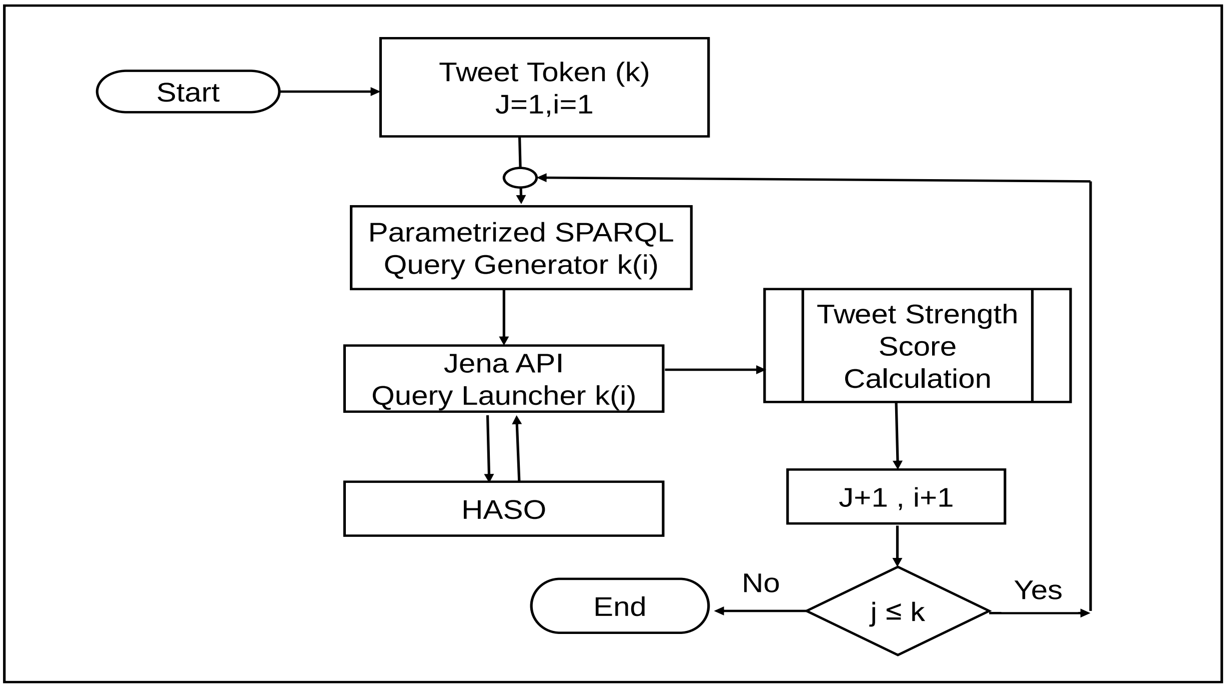

Fig. 6. Tweet processing .

some cases, the queried word may not be present in our ontology. Consequently, the SPARQL query result will be zero. In the column "Word Strength value by SPARQL Query" we present the words that affect the tweet's overall sentiment polarity. We ran the SPARQL query for each word in all of the tweets. In the third example, we have two negative words. However, we consider the negative word with the smallest score "terrorist (-3)" in order to calculate the sentence sentiment score and eventually the sentence sentiment polarity.

Table 2

Example of tweets sentiment analysis result of HASO

Sentence (Tweet)	Word Strength value by SPARQL Query	Tweet Sentiment Result
I watched 25+ minutes of a Facebook live video hatching one of these and I regret every second	HASO: regret (-2)	Negative
Entrepreneurship is about risk-taking, this time it is paying off	HASO : risk-taking (3)	Positive
Hopefully Trump will designate as a terrorist organization and law enforcement can end reign of terror	HASO: Hopefully (2) HASO: terrorist(-3) HASO: enforcement(-2)	Negative

4.2. Ontology vs Machine Learning Sentiment Analysis Result

We employ HASO for sentiment analysis on the data from SemEVal-2017 Task 4 Subtask A. The data contains 12,284 tweets and the goal is to classify the tweets into the three classes positive, negative, and neutral. The data contains tweets related to diverse events such as movies, singers, and politics.

After classifying the tweets with our proposed ontology, we compare our results with the human classification (annotation) that was provided with the test data. Then, we calculate the average recall and F-Score.

There were 38 teams participating in Subtask A. In SemEVal-2017 Task 4 Subtask A, the top 5 teams used deep leaning, and three of the top ten used SVM with various language lexicons. We compare the ontological method with the basic machine learning methods. The ontological method can give high performance results when compared to the basic machine learning method, which is why we chose to compare our proposed ontology results with three of the top-10 teams. The participating teams were ranked according to their average recall (AvgRec) results in Subtask A, where a higher score is better.

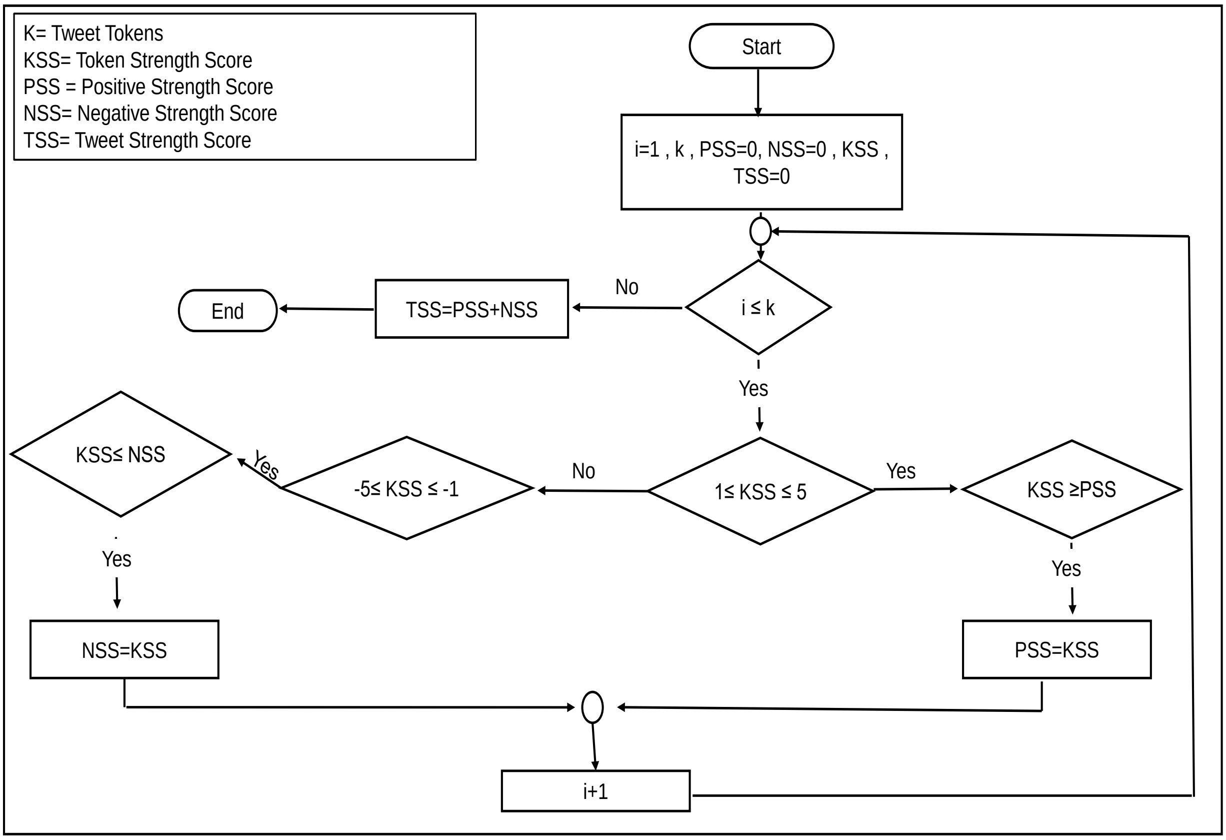

Fig. 7. Tweet strength score calculation

Table 3

Comparison of ontology and machine learning method

System	Avg Rec	F-score
INGEOTEC	0.649	0.645
HASO	0.646	0.637
SiTAKA	0.645	0.628
UCSC-NLP	0.642	0.624

To evaluate the performance of our ontology, we used the two functions `sklearn.metrics.recall` and `sklearn.metrics.f1` to calculate the average recall and the F1 score, respectively[20]. Theses metrics belong to the free python machine learning library scikit-learn. The metrics can handle multilabel classification (positive, negative, neutral).

Table 3 shows the comparison of the sentiment analysis results between our proposed ontology (HASO) and three of the top-10 teams that participated in Subtask A [27]. The results show that our ontological method (HASO) came in second, after the INGEOTEC team, with a small difference of 0.003 points of average recall.

Overall, our method came in 8th place, after INGEOTEC, compared to all 38 participating teams, as shown in [27]. This proves the effectiveness of the ontological method compared to the machine learning method in the area of sentiment analysis. HASO is designed to cover a wide range of sentiment words from language dictionaries and psychological-based resources. With the ontology ability and performance in the sentiment analysis area , the requirements of manual annotation in machine learning approaches can be resolved. Moreover, Ontology has the ability to represent the domain vocabularies that can have different representation in another domain and context [28].

[20] http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics

5. Conclusion

In this work we proposed the development of a Human Affective States Ontology. Moreover, we present the HASO modularization process to produce manageable and scalable modules. The development of the HASO ontology incorporates various sentiment lexicons and psychological-based resources in order to capture a wide range of sentiment terms. The experimental evaluation on the semeval 2017-Subtask A dataset demonstrates the effectiveness of the HASO in detecting sentiment compared to the machine learning method. With the ontology's ability and performance in the area of sentiment analysis, we no longer need manual annotation in machine learning approaches. We aim to compare the sentiment analysis performance on the SemEval-2017 Task 4, Subtask A dataset between our proposed ontology and SentiStrength. Since the sentiment analysis uses natural language processing to identify the human affective states it incorporates more features from natural language processing, which may improve the overall results of the ontology sentiment analysis. In the future, we intend to use more NLP features and further extend the ontology with more language dictionaries and psychology based resources, as well as evaluate the ontology with a different dataset.

References

[1] K. Ravi and V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, *Knowledge-Based Systems* **89** (2015), 14–46.

[2] A. Abirami and V. Gayathri, A survey on sentiment analysis methods and approach, in: *Advanced Computing (ICoAC), 2016 Eighth International Conference on*, IEEE, 2017, pp. 72–76.

[3] L. Floridi, *The Blackwell guide to the philosophy of computing and information*, John Wiley & Sons, 2008.

[4] C. Roussey, F. Pinet, M.A. Kang and O. Corcho, An Introduction to Ontologies and Ontology Engineering, in: *Ontologies in Urban Development Projects*, Vol. 1, Springer, London, 2011, pp. 9–38, DOI:10.1007/978-0-85729-724-2_2.

[5] R. Abaalkhail, B. Guthier, R. Alharthi and A. El Saddik, Survey on ontologies for affective states and their influences, *Semantic Web* **.**(Preprint) (2017), 1–18.

[6] K. Wójcik and J. Tuchowski, Ontology Based Approach to Sentiment Analysis, *June-2014*.

[7] P. Thakor and S. Sasi, Ontology-based sentiment analysis process for social media content, *Procedia Computer Science* **53** (2015), 199–207.

[8] E. Kontopoulos, C. Berberidis, T. Dergiades and N. Bassiliades, Ontology-based sentiment analysis of twitter posts, *Expert systems with applications* **40**(10) (2013), 4065–4074.

[9] R. Nithish, S. Sabarish, M.N. Kishen, A. Abirami and A. Askarunisa, An Ontology based Sentiment Analysis for mobile products using tweets, in: *Advanced Computing (ICoAC), 2013 Fifth International Conference on*, IEEE, 2013, pp. 342–347.

[10] K.M. Sam and C. Chatwin, Ontology-based sentiment analysis model of customer reviews for electronic products, in: *Encyclopedia of Information Science and Technology, Third Edition*, IGI Global, 2015, pp. 892–904.

[11] M.D. Sykora, T. Jackson, A. O'Brien and S. Elayan, Emotive ontology: Extracting fine-grained emotions from terse, informal messages (2013).

[12] M.M. Taye, Understanding semantic web and ontologies: Theory and applications, *arXiv preprint arXiv:1006.4567* (2010).

[13] M. Fernández-López, A. Gómez-Pérez and N. Juristo, Methontology: from ontological art towards ontological engineering (1997).

[14] E. Hudlicka and H. Gunes, Benefits and limitations of continuous representations of emotions in affective computing: introduction to the special issue, *International Journal of Synthetic Emotions* **3**(1) (2012).

[15] N.F. Noy, D.L. McGuinness et al., Ontology development 101: A guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001.

[16] E. Blomqvist, The use of Semantic Web technologies for decision support–a survey, *Semantic Web* **5**(3) (2014), 177–201, DOI:10.3233/SW-2012-0084.

[17] T.R. Gruber et al., A translation approach to portable ontology specifications, *Knowledge acquisition* **5**(2) (1993), 199–220, DOI:10.1006/knac.1993.1008.

[18] A. Merono-Penuela, A. Ashkpour, v.M. Erp, K. Mandemakers and L. Breure, Semantic Technologies for Historical Research: A Survey, *Semantic Web Journal* (2014), 539–564, DOI:10.3233/SW-140158.

[19] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner et al., The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data, in: *Affective computing and intelligent interaction*, Springer, Berlin, Heidelberg, 2007, pp. 488–500, DOI:10.1007/978-3-540-74889-2_43.

[20] C. Laurier, M. Sordo, J. Serra and P. Herrera, Music Mood Representations from Social Tags., in: *ISMIR*, Citeseer, Kobe, Japan, 2009, pp. 381–386.

[21] Z. Obrenovic, N. Garay, J.M. López, I. Fajardo and I. Cearreta, An ontology for description of emotional cues, in: *Affective Computing and Intelligent Interaction*, Springer, Berlin, Heidelberg, 2005, pp. 505–512, DOI:10.1007/11573548_65.

[22] M.C. Suárez-Figueroa, NeOn Methodology for building ontology networks: specification, scheduling and reuse, PhD thesis, Informatica, 2010.

[23] M. d'Aquin, Modularizing ontologies, in: *Ontology Engineering in a Networked World*, Springer, 2012, pp. 213–233.

[24] M. d'Aquin, A. Schlicht, H. Stuckenschmidt and M. Sabou, Criteria and evaluation for ontology modularization techniques, in: *Modular ontologies*, Springer, 2009, pp. 67–89.

[25] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong and U. Kaymak, Exploiting emoticons in sentiment analysis, in: *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ACM, 2013, pp. 703–710.

[26] M. Thelwall, Heart and soul: Sentiment strength detection in the social web with sentistrength, *Proceedings of the CyberEmotions* **5** (2013), 1–14.

[27] S. Rosenthal, N. Farra and P. Nakov, SemEval-2017 task 4: Sentiment analysis in Twitter, in: *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, pp. 502–518.

[28] C.K. Cheng, X. Pan and F. Kurfess, Ontology-based semantic classification of unstructured documents, in: *International Workshop on Adaptive Multimedia Retrieval*, Springer, 2003, pp. 120–131.