The Hourglass Model Revisited

Yosephine Susanto  
Nanyang Technological University

Andrew G. Livingstone  
University of Exeter

Bee Chin Ng

Erik Cambria  
Nanyang Technological University

Abstract—Recent developments in the field of AI have fostered multidisciplinary research in various disciplines, including computer science, linguistics, and psychology. Intelligence, in fact, is much more than just IQ: it comprises many other kinds of intelligence, including physical intelligence, cultural intelligence, linguistic intelligence, and EQ. While traditional classification tasks and standard phenomena in computer science are easy to define, however, emotions are still a rather mysterious subject of study. That is why so many different emotion classifications have been proposed in the literature and there is still no common agreement on a universal emotion categorization model. In this work, we revisit the Hourglass of Emotions, an emotion categorization model optimized for polarity detection, based on some recent empirical evidence in the context of sentiment analysis. This new model does not claim to offer the ultimate emotion categorization but it proves the most effective for the task of sentiment analysis.
The remainder of this paper is organized as follows: next section discusses the main emotion models proposed in the literature; later, the revised version of the Hourglass model is presented in detail; then, an evaluation of the model on three sentiment analysis datasets is provided; finally, the last section offers concluding remarks.

RELATED WORK
Emotion research has increased significantly over the past few years thanks to the recent developments in the field of AI. The question, in fact, is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any [6]. One of the earliest efforts in developing an emotion model was made by Shaver et al. [7]. They first selected a group of words and had them classified as emotion words and non-emotion words. This step resulted in 135 emotion words, which were then annotated based on their similarity and grouped into categories so that inter-category similarity was minimized but intra-category similarity maximized. Using the typical prototyping approach, they managed to develop an abstract-to-concrete emotion hierarchy and discovered six emotions on the hierarchy’s lowest level: joy, love, surprise, sadness, anger, and fear. This emotion study implied that most emotions are fuzzy or indistinct and they are combinations of these six basic emotions, which cannot be further divided.

Later, Ortony and Turner argued against the view that basic emotions are psychologically primitive [8]. They proposed that all emotions are discrete, independent, and related to each other through a hierarchical structure, hence there is no basic set of emotions that serve as the constituents of others. Having refuted the existence of basic emotions, Ortony, Clore, and Collins introduced their own emotion model (termed OCC from the initials of the three authors) [9]. The OCC model classifies emotions into 22 emotion types. The hierarchy contains three branches, namely consequences of events (e.g., pleased or displeased), actions of agents (e.g., approving or disapproving), and aspects of objects (e.g., liking or disliking). A number of ambiguities of the emotions defined in the OCC model were later identified and discussed by Steunebrink et al. [10], who extended the model to 24 emotion categories.

A few years after the original OCC model was proposed, Mehrabian proposed the Valence/Arousal model [11], a popular model in psychology that places specific emotion concepts in a circumflex model of core affect defined by two basic dimensions: Arousal, which ranges from high to low, and Valence, which varies from positive to negative. Another very popular model, based on facial expressions, was later proposed by Ekman [12]. The model only consists of six emotions (anger, fear, disgust, joy, sadness, and surprise) but turned out to be one of the most used models in the literature for its simplicity and applicability. Many subsequent models are based on Ekman’s model, e.g., Plutchik’s wheel of emotions [13]. Likewise, the Hourglass of Emotions [5] is a reinterpretation of Plutchik’s model for sentiment analysis. Many more models have been proposed in the literature [14], mostly to adapt previous models to different disciplines, modalities, or applications.

THE REVISITED MODEL
After almost a decade of using the Hourglass model [5] in the context of sentiment analysis, we realized that this presents several issues, namely:

- uncanny color associations;
- presence of neutral emotions;
- absence of some polar emotions;
- wrong association of antithetic emotions;
- low polarity scores for compound emotions;
- absence of self-conscious or moral emotions.

Uncanny color associations
While this was not a matter that affected the accuracy of sentiment analysis, it has been a pressing issue for a while since many researchers in the community questioned the choice of some colors of the Hourglass, e.g., blue for surprise, green for fear, and purple for both sadness and disgust. In line with recent studies on the association between colors and emotions [15], we assigned tendentially warm colors to positive emotions and cold colors to negative ones (Figure 1). This also ensures a better distinction between different emotions (e.g., sadness and disgust are now blue and green, respectively) and an enhanced organization of the model (positive emotions now reside in the upper part of the Hourglass while negative ones are at the bottom).
Wrong association of antithetic emotions

One of the main advantages of having an emotion categorization model is to be able to classify unknown concepts based on known features. For example, if the model did not contain the emotion discomfort, it could look up its opposite (comfort) and flip its polarity to obtain the polarity of the unknown concept. This mechanism works well in the new model, as emotions are now organized with respect to their polarity (Table 2), but it generated a lot of errors in the previous version of the Hourglass, as this contained wrong associations of antithetic emotions, e.g., anger and fear (which are both negative) or surprise and anticipation (which are opposite in terms of meaning but not in terms of polarity).

Low polarity scores for compound emotions

The main goal of sentiment analysis is to calculate the polarity value (positive or negative) of a piece of text, an image or a video. In many applications, polarity intensity also plays an important role for classification and decision-making. The old Hourglass model had a big shortcoming in this sense: to make sure the polarity value stayed between -1 (extreme negativity) and +1 (extreme positivity), a static normalization factor was introduced. Such a normalization factor, however, made the polarity intensity of most concepts very low. Concepts with high intensity were not the ones with high emotional charge but rather those that were associated with compound emotions (e.g., hatred) because of more dimensions active at the same time (e.g., anger and fear).

Presence of neutral emotions

One of the main problems with the previous model was the presence of ambiguous emotions (e.g., distraction [16]) and, especially, neutral emotions, e.g., surprise. Here, we do not want to debate whether surprise is an emotion or not [4] but we definitely do not want it in a model that is catered for sentiment analysis as this will lead to the wrong categorization of all concepts (words and multi-word expressions) that are semantically associated with it. Surprise, in fact, only becomes polar when coupled with positive or negative emotions (Table 1).

Absence of some polar emotions

Another issue with the original model was the absence of some important polar emotions, e.g., calmness and eagerness. All the concepts associated with such emotions, e.g., deep_breath or volunteer, were going undetected by the model and, hence, miscategorized as neutral. This issue extended to germane emotions, e.g., enthusiasm and bliss, and concepts associated with them, e.g., ambition or meditation.

Table 1. Examples of compound emotions.
To this end, we replaced the old normalization factor with a new dynamic quantity that is directly proportional to the number of active dimensions:

$$p_c = \frac{I_c + T_c + A_c + S_c}{|\text{sgn}(I_c)| + |\text{sgn}(T_c)| + |\text{sgn}(A_c)| + |\text{sgn}(S_c)|}$$

(1)

where \( c \) is an input concept, \( p \) is the polarity value of such concept, \( I \) is the value of Introspection (the \textit{joy}-versus-\textit{sadness} dimension), \( T \) is the value of Temper (the \textit{calmness}-versus-\textit{anger} dimension), \( A \) is the value of Attitude (the \textit{pleasantness}-versus-\textit{disgust} dimension), and \( S \) is the value of Sensitivity (the \textit{eagerness}-versus-\textit{fear} dimension). Before, a negative concept (e.g., \textit{death}) associated with a strong emotion (e.g., \textit{grief}) would not result in a high (negative) polarity because its affective intensity would have been divided by 3. Now, that same intensity remains intact because the denominator of the polarity formula is equal to 1, since only one dimension (Introspection) is active. The denominator will actually be equal to 1 for most concepts, as most concepts are only associated with one emotion; it will be equal to 2 for concepts that are associated with bidimensional emotions like \textit{love} (\textit{joy}-plus-\textit{pleasantness}) and submission (\textit{fear}-plus-\textit{pleasantness}); it will be equal to 3 for those few concepts that are associated with tridimensional emotions like bittersweetness (\textit{sadness}-\textit{anger}-plus-\textit{pleasantness}); finally, it will be 4 for those very rare concepts that are associated with compound emotions that span all dimensions like jealousy (\textit{anger}-\textit{fear}-\textit{sadness}-\textit{disgust}).

Absence of self-conscious or moral emotions

The old Hourglass model systematically excluded what are commonly known as self-conscious or moral emotions such as pride, prejudice, guilt, shame, embarrassment or humiliation. This has been a serious issue as it caused the model to be unable to recognize this pretty large subset of emotions and, hence, the polarity (and the concepts) associated with them. We solved this issue by encapsulating such emotions as subdimensions of Attitude (Table 3).

Emotions like pride and confidence, in fact, can be interpreted as positive Attitude (\textit{pleasantness} and \textit{acceptance}, respectively) directed at oneself. Likewise, embarrassment and guilt rep-

---

**Table 2.** New emotion classification with five sample emotion words for each category.

<table>
<thead>
<tr>
<th>Introspection</th>
<th>Temper</th>
<th>Sensitivity</th>
<th>Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECSTASY</td>
<td>BLISS</td>
<td>ENTHUSIASM</td>
<td></td>
</tr>
<tr>
<td>Elation</td>
<td>Placidity</td>
<td>Deulath</td>
<td></td>
</tr>
<tr>
<td>Joy</td>
<td>Calmness</td>
<td>Eagerness</td>
<td></td>
</tr>
<tr>
<td>Contentment</td>
<td>Serenity</td>
<td>Responsiveness</td>
<td></td>
</tr>
<tr>
<td>Melancholy</td>
<td>Serenity</td>
<td>Anxiety</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>Antoyance</td>
<td>Fear</td>
<td></td>
</tr>
<tr>
<td>Grief</td>
<td>ANGER</td>
<td>Terror</td>
<td></td>
</tr>
</tbody>
</table>

---

- **Introspection** includes emotions like elation, joy, contentment, melancholy, and grief.
- **Temper** includes placidity, calmness, serenity, annoyance, anger, and rage.
- **Sensitivity** includes placidity, calmness, serenity, annoyance, anger, and rage.
- **Attitude** includes elation, joy, contentment, melancholy, and grief.
resent negative Attitude (dislike and disgust, respectively) directed at oneself. Similarly, magnanimity and sociability can be considered positive Attitude (delight and pleasantness, respectively) towards others, while humiliation and malevolence represent negative Attitude (disgust and loathing, respectively) towards others.

EVALUATION

We tested the new Hourglass model against some of the above-mentioned emotion categorization models on three sentiment benchmarks: the Blitzer Dataset [17], the Movie Review Dataset [18], and the Amazon dataset [19]. The first consists of product reviews in seven different domains and contains 3,800 positive sentences and 3,410 negative ones. The second is about movie reviews and is composed of 4,800 positive sentences and 4,813 negative ones. Finally, the Amazon dataset contains the reviews of 453 mobile phones, which were split into sentences and labeled as positive, neutral, or negative. The final dataset contains 48,680 negative sentences and 64,121 positive ones.

We used these three datasets to compare how the new Hourglass model performs on the task of polarity detection in comparison with the models proposed by Shaver [7], Ekman [12], Plutchik [13], the OCC models [9, 10], and the previous Hourglass model [5] (Table 4). For this experiment, we considered sentiment analysis as a binary classification problem (positive versus negative) and, hence, we left out models that focus more on intensity, e.g., the Valence/Arousal model.

The evaluation was performed by connecting the concepts of SenticNet [20], a commonsense knowledge base for sentiment analysis, to a positive or negative polarity via the emotions of each model and by using sentic patterns [19] to calculate the polarity of each sentence in the datasets. Sentic patterns model sentences as electronic circuits: sentiment words are ‘sources’ while other words are ‘elements’, e.g., very is an amplifier, not is a logical complement, rather is a resistor, but is an OR-like element that gives preference to one of its inputs (Figure 2). Thus, for each emotion model, a polarity was firstly assigned to each concept encountered in a sentence based on its connections with positive or negative emotions in the graph of SenticNet and, secondly, sentic patterns were used to calculate the final polarity of the sentence.

As expected, the accuracy of text sentiment analysis using the models of Ekman and Shaver is low as both are based on facial expressions and, hence, cover a very limited set of emotions. Ekman’s model, in particular, is not very good for detecting polarity from text because, unlike Shaver’s model, it is unbalanced (as it consists of 2 positive emotions and 4 negative ones).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ekman’s model [10]</td>
<td>66.87%</td>
<td>65.93%</td>
<td>59.53%</td>
</tr>
<tr>
<td>Shaver’s model [6]</td>
<td>67.12%</td>
<td>66.73%</td>
<td>60.89%</td>
</tr>
<tr>
<td>Plutchik’s model [11]</td>
<td>86.94%</td>
<td>85.79%</td>
<td>80.91%</td>
</tr>
<tr>
<td>Hourglass model [3]</td>
<td>88.27%</td>
<td>88.12%</td>
<td>82.75%</td>
</tr>
<tr>
<td>OCC model [7]</td>
<td>89.15%</td>
<td>88.73%</td>
<td>84.76%</td>
</tr>
<tr>
<td>OCC model revisited [8]</td>
<td>90.41%</td>
<td>89.41%</td>
<td>85.93%</td>
</tr>
<tr>
<td>Hourglass model revisited</td>
<td>94.92%</td>
<td>93.29%</td>
<td>89.85%</td>
</tr>
</tbody>
</table>

Table 4. Comparison of emotion models on three datasets for sentiment analysis.
CONCLUSION

Affective neuroscience and twin disciplines have clearly demonstrated how emotions and intelligence are strictly connected. Some prominent researchers have also questioned the possibility of emulating intelligence without taking emotions into account. Emotions, however, are rather elusive entities and, hence, are difficult to categorize.

In this paper, we reviewed major emotion models and proposed a new version of the Hourglass model, a biologically-inspired and psychologically-motivated emotion categorization model for sentiment analysis.

This model represents affective states both through labels and through four independent but concomitant affective dimensions, which can potentially describe the full range of emotional experiences that are rooted in any of us. The new version of the model provides a better color representation of emotions; it excludes neutral emotions (e.g., surprise) and includes some important polar emotions that were previously missing (including self-conscious and moral emotions); it better categorizes emotions in order to ensure that antithetic emotions are mirrored; finally, it calculates the polarity associated with natural language concepts with higher accuracy.

In the future, we plan to test the validity of the new Hourglass model on different domains (beyond product reviews) and different modalities (beyond text). We also plan to develop mechanisms to dynamically customize the model according to different cultures, personalities, age group, sex, and user preferences.

ACKNOWLEDGEMENTS

The authors would like to thank Prof Andrew Ortony for the insightful discussions on emotion categorization (especially on the inclusion of calmness and the exclusion of surprise) and Jocelyn, Victor and Alvin Choong for their support in designing the new graphics of the model.

REFERENCES


Yosephine Susanto is a PhD student at Nanyang Technological University. Her research interests include affective computing and multilingual sentiment analysis. Susanto received a master in applied linguistics from Atma Jaya Catholic University of Indonesia. Contact her at yosephin001@e.ntu.edu.sg.

Andrew G. Livingstone is a senior lecturer at University of Exeter. His research interests lie in social identity, emotion, group processes and intergroup relations. Livingstone was awarded a PhD in social psychology from University of Exeter. Contact him at a.livingstone@exeter.ac.uk.

Bee Chin Ng is an associate professor at Nanyang Technological University. Her research interests include psycholinguistics and sociolinguistics aspects of language acquisition in multilingual contexts. Ng received a PhD in linguistics from La Trobe University. Contact her at mbcng@ntu.edu.sg.

Erik Cambria is the corresponding author and an associate professor at Nanyang Technological University. His main research interests are AI and affective computing. Cambria earned his PhD in computing science and mathematics through a joint programme between the University of Stirling and MIT Media Lab. Contact him at cambria@ntu.edu.sg.