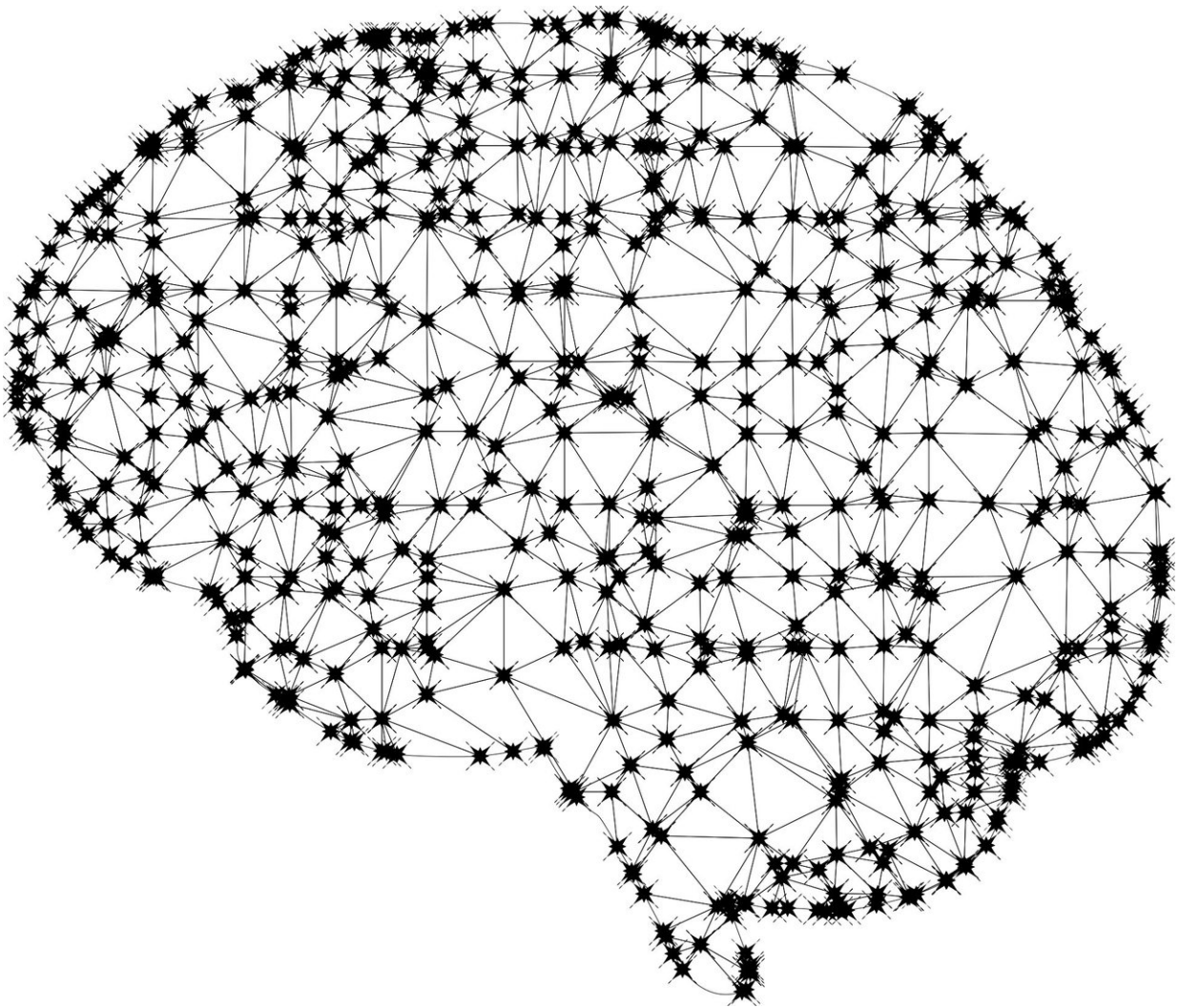# CruzAffect: a feature-rich approach to characterize happiness

March 6 2019, by Ingrid Fadelli

Credit: CC0 Public Domain

A team of researchers at UC Santa Cruz have recently developed a new machine learning approach to characterize happiness, called CruzAffect. Their approach, presented in a paper pre-published on arXiv, can be applied to different models for affective content classification, including both traditional classifiers and deep learning convolutional neural networks (CNN).

This recent study builds on previous research that explored how people convey first-person affect and happiness. In one study, the same researchers found that people tend to describe situations, such as 'my friend bought me flowers', or 'I got a parking ticket', from which other humans can readily infer their implicit affective reactions. They concluded that compositional semantics can provide strong evidence of the sentiment associated with a given event.

X	Y	E_{hve}	E_{lck}	Example
+	+	+	-	I have a new kitten.
+	-	-	+	I got a parking ticket.
-	+	-	+	My rival got a prize.
-	-	+	-	My rival got a reprimand.

X have/lack Y

Table 1: Functions for verbs of possession.

Credit: Wu et al.

In another study, the researchers tried to ground people's linguistic descriptions of events on theories of well-being and happiness. By

analyzing a corpus of private micro-blogs extracted from an application called Echo, they examined the extent to which different theoretical accounts could explain the variance in the happiness scores that Echo users gave to daily events in their lives.

"It's challenging to generalize an affective event and associate it with well-being theories," Jiaqi Wu, one of the researchers who carried out the study, told TechXplore. "In our past research, we noticed that there isn't a single theory that can predict the sentiment of all affective events. The aim of our recent work was to identify specific compositional semantics that characterize the sentiment of events and attempt to model happiness at a higher level of generalization. However, finding generic characteristics for modeling well-being remains challenging."

The primary aim of the recent study carried out by Wu and her colleagues was to investigate the effectiveness of feature-rich traditional machine learning methods and deep learning methods for affective content classification. To achieve this, they identified a series of features that characterize happiness in affective content and applied them to a traditional classifier, XGBoosted forest, and a CNN.

"Our project, called CruzAffect, includes the development of two different models: a traditional machine learning method (i.e. XGBoosted forest) and a deep learning CNN with GloVe embedding," Wu said. "We utilize syntactic features, emotional features, and profile features, and their performance is stable for different affective content classification tasks."

Essentially, the researchers evaluated the performance of two different machine learning models for affective content classification (XGBoosted forest and a CNN), both of which analyzed content based on the features that they had previously identified. These include:

- Syntactic Features: part of Speech, nouns, verbs, adjectives and adverbs, use of questions, tense and aspect information.
- Emotional Features: Linguistic Inquiry and Word Count (LIWC) v2007, emotion lexicon, subjectivity lexicon, level of factual and emotional language.
- Word Embedding: GloVe 100 dimension word vectors for word representation.
- Profile Features: age, country, gender, marital status, parenthood, etc.

These features allowed the researchers to uncover essential indicators of social involvement and control that different people might exercise during happy moments. In their study, they trained both the XGBoosted and CNN model with supervised learning on a dataset of 10,000 labeled textual snippets. They also trained the models to generate pseudo-labels for 70,000 unlabeled snippets using a bootstapped semi-supervised approach, as this allowed them to broaden their dataset. All of these textual snippets were extracted from the HappyDB database.
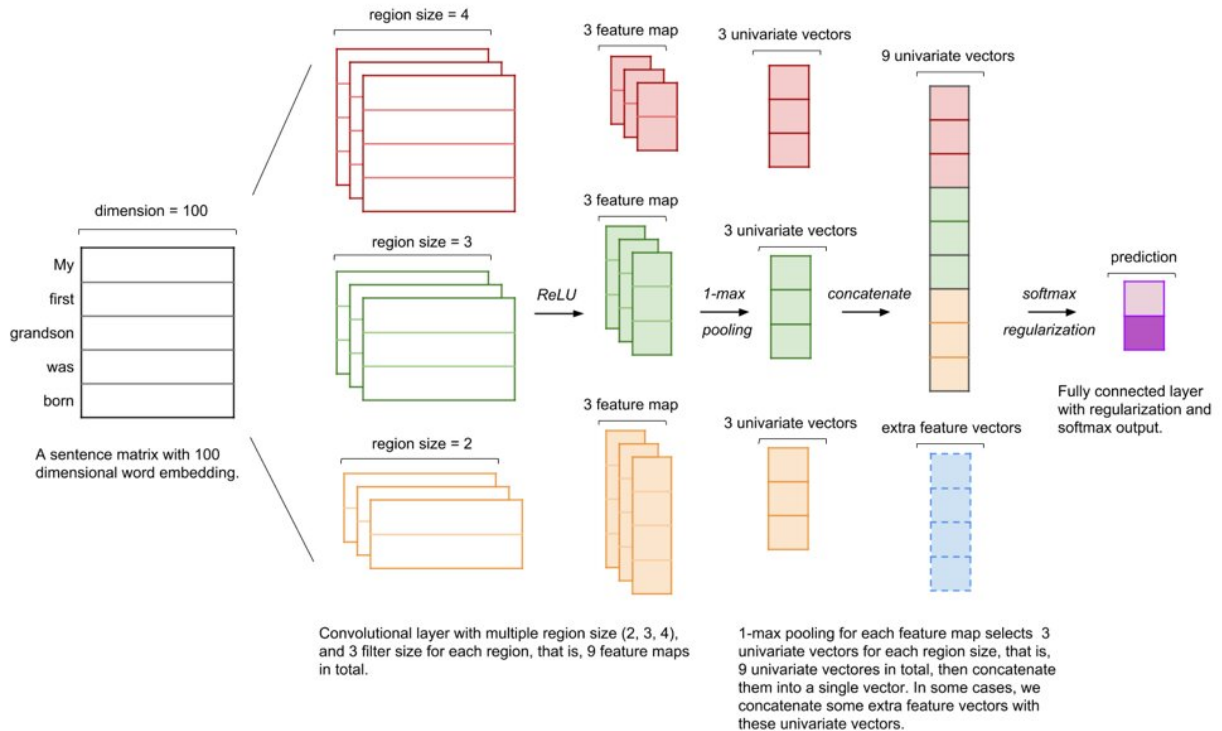
CNN Architecture. Credit: Wu et al.

"The meaningful findings of our study include the interesting syntactic patterns that repeat over different domains," Wu said. "Such linguistic patterns are likely to be associated with well-being theories. We also find that the features that include expert knowledge, such as LIWC dictionary can improve the performance of traditional model as well as the deep learning model in the affective content classification tasks."

The researchers evaluated the XGBoosted forest and CNN models on the binary classification of agency and social labels, as well as on the multi-class prediction of concept labels. Their evaluations yielded promising results, suggesting that the features identified by them are

particularly effective for classifying affective content. Although the CNN based model performed better on multi-class classification tasks, the traditional machine learning model achieved comparable results using the features that they had previously identified.
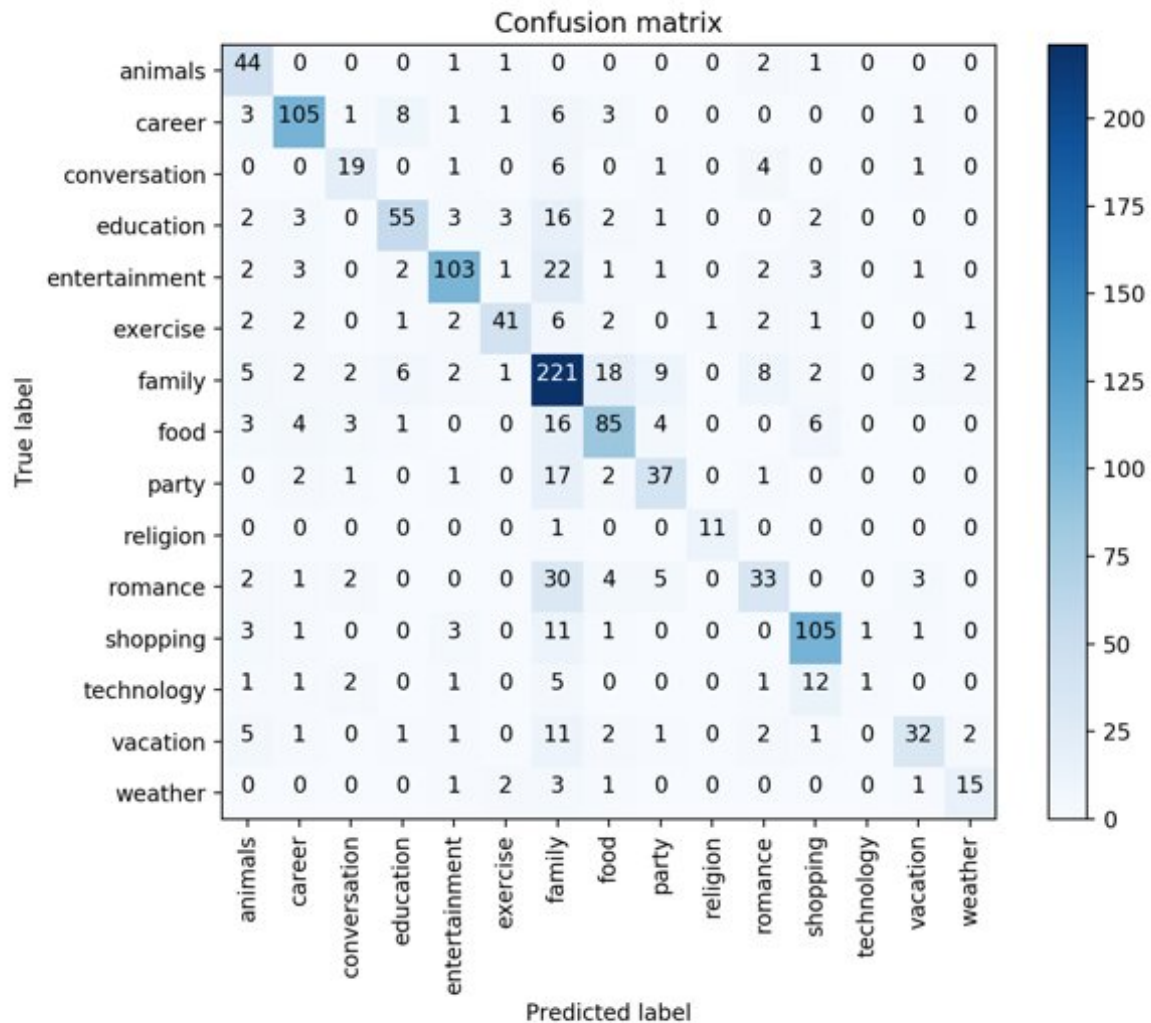
The confusion matrix of the best CNN model with syntactic, emotional and profile features in 10-fold cross-validation to predict the concepts characteristic. Credit: Wu et al.

The study carried out by Wu and her colleagues uncovered general themes that are recurrent in people's descriptions of happy moments. In the future, their findings could inform the development of new models for affective classification tasks, allowing researchers to effectively predict well-being and happiness by analyzing the content of textual snippets.

"I will now explore the cross-domain affective event analysis, and investigate a better model to ground the linguistic descriptions of events that users experience in theories of well-being and happiness," Wu said. "After understanding the relation between the affective content and well-being theories, we might be able to collect general affective events that are highly related to the well-being."

The team of researchers who carried out the study. Credit: Wu et al.

More information: CruzAffect at AffCon 2019 shared task: A feature-rich approach to characterize happiness. arXiv:1902.06024 [cs.CL]. arxiv.org/abs/1902.06024

Learning lexico-functional patterns for first-person affect. arXiv:1708.09789v1 [cs.CL]. arxiv.org/pdf/1708.09789.pdf

Linguistic reflexes of well-being and happiness in echo. arXiv:1709.00094v1 [cs.CL]. arxiv.org/pdf/1709.00094.pdf