

Toward Comprehensive Understanding of a Sentiment Based on Human Motives

Naoki Otani and Eduard Hovy Language Technologies Institute, Carnegie Mellon University, USA.

What is the reason for a sentiment value?

Many studies have focused on the aspect & valence of a sentiment

Everything is always cooked to perfection. 😊 (FoodQuality, P)
The waiter was rude at times. 😡 (Service, N)

but have paid little attention to the *reasons* for holding a sentiment.

Aspects are typically limited to properties of entities
and do not show *why* and *how* such aspects cause sentiments.
→ *Clues* to better respond to the sentiments.

We assume that a sentiment is triggered by whether the holder's *motive* is satisfied. (Li&Hovy,17)

→ Human motive detection as the first step.

Everything is always cooked to perfection. 😊

Feeling satisfied [Self-fulfilment]

The waiter was rude at times. 😡

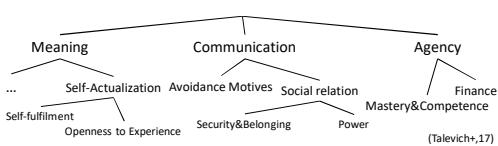
Being treated well [Social Relation]

Highlights

- Define six basic motives covering a wide range of topics in reviews.
- Annotate 1,600 restaurant & laptop reviews by crowdsourcing.
- Empirically show that underlying motives are universal across domains.

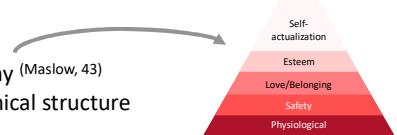
Representation of Human Motives

We use the hierarchical taxonomy defined by Talevich et al. (2017)



Why?

- Comprehensive (161 motives)
- Links to well-known theories like Maslow's hierarchy (Maslow, 43)
- Easy to adjust the granularity based on the hierarchical structure



Maslow's hierarchy is used in several NLP studies
(Ding&Riloff,18; Rashkin+18)

Annotation by Crowdsourcing

Based on preliminary annotations, we choose seven motive categories.

Self-Fulfilment	Finding meaning in life or feeling satisfied with one's life.
Embrace & Explore Life (merged into self-fulfillment)	Being entertained and exploring a new thing
Appreciating Beauty	Enjoying fine visual design/arts/natural beauty or being creative.
Social Relation	Being treated well by others or belonging to a social group.
Health	Being physically healthy/good.
Ambition & Ability	Being competent/knowledgeable, keeping things in order, or being efficient.
Finance	Saving money or getting things worth the financial cost.

Restaurant and Laptop reviews, 800 sentences w/≤ 25 tokens from each domain.

"The quality of the food was perfect."

"The wine list is extensive."

"A beautifully designed dreamy restaurant."

"Everyone was cheerfully cooperative."

"The fish was not fresh, and the rice tasted old."

"I've waited over one hour for food."

"The prices are high, but I felt it was worth it."

Data: SemEval 2016 Aspect-based Sentiment Analysis

Restaurant and Laptop reviews, 800 sentences w/≤ 25 tokens from each domain.

Label distribution:

Use MACE (Hovy+13) to obtain gold standard labels.

Quality Control:

worker qualification + redundancy (3 workers/text)

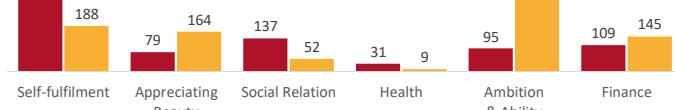


- Labels on the first 200 texts* are corrected by the first author.
- We qualify workers whose F1 score is higher than 0.5,

Annotation agreement (Krippendorff's α): 0.51 (restaurant), 0.61 (laptop)

Label distribution:

Use MACE (Hovy+13) to obtain gold standard labels.



Human Motive Detection

Given a sentence, predict relevant motives. (Multi-label classification)

Predictive Models:

Linear SVM
SVM only uses surface word forms.

- Bag of n -gram vectors ($n=1,2,3$)
- Scale n -gram counts by TFIDF

Multi-layer Perceptron (MLP)



Encoding a series of word embeddings into a sentence embedding

- SWEM (Shen+18) average and max-pooling
- CNN (Kim,14) with window size 3,4, and 5
- BiLSTM: Two-layers Bidirectional LSTM

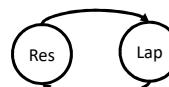
Training:

Transfer learning across domains

Human motives will be universal across domains although distributions can be different.

$$\mathcal{L}' = \mathcal{L}_{in} + \lambda \mathcal{L}_{out}$$

In-domain Out-domain



Weighted loss function (MLP)

for handling highly skewed class distributions.

$$\mathcal{L} = - \sum_{(x,y) \in D} \sum_{c \in \mathcal{C}} [w_c y_c \log \text{MLP}_c(x) + (1 - y_c) \log(1 - \text{MLP}_c(x))]$$

w_c : class weight (Morik+, 99) $w_c = \frac{\#(y \neq c)}{\#(y=c)}$ $c \in \mathcal{C}$

Experiments

Settings:

Primary measure: macro-F1

3-fold CV: in each fold

- Train/valid/test = 1:1:1
- Tune hyperparameters using the train and valid sets.
- Train a model on the train+valid set and evaluate on the test set.

100D GloVe embeddings (Pennington+, 14)

Trained on 6B tokens from Wikipedia and En Gigaword.
Embedding layers are fixed during training.

Results: Out-of-domain training data helps.

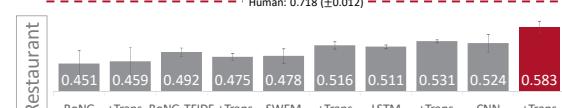
↑ The precision of MLP classifiers.

This indicates the universality of underlying motives across domains.

SVM performs poorly in terms of recall.

Surface-level features are insufficient.

Large gap between the classifiers and human.



Error Analysis: Errors concentrate on ambiguous examples.

Text

I had to ask her three times before she finally came back with the dish I've requested. [restaurant service] Ambition & Ability (Being efficient) Self-fulfilment, Social Relation

English must have been his third or fourth language. [laptop customer service]

Gold

Social Relation

Prediction

Ambition & Ability