EmojiNet: A Machine Readable Emoji Sense Inventory

Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, Derek Doran

Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University, Dayton OH, USA

{sanjaya, lakshika, amit, derek}@knoesis.org

Motivation

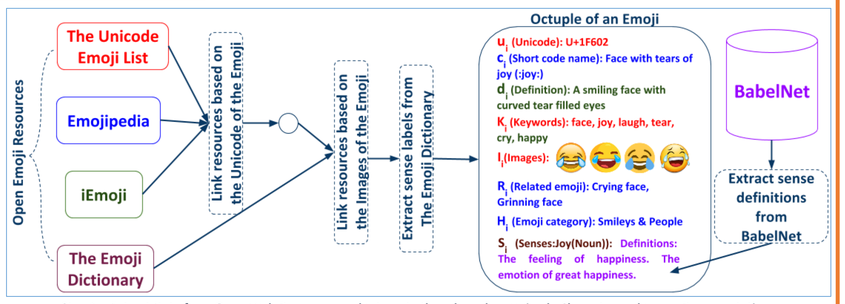
- ✓ Emoji symbols take on different meanings based on the context of a message as no rigid semantics are attached to emoji by design.
- ✓ Hence, for machines to understand emoji, a machine readable sense inventory that lists different emoji senses is needed.
- ✓ **EmojiNet** is a machine readable sense inventory which enables this.

Approach

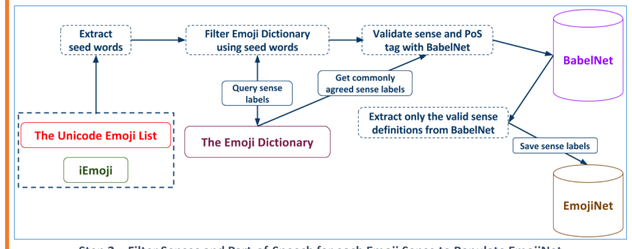
For each emoji e_i in EmojiNet, it records the octuple $e_i = (u_i\ c_i\ d_i\ K_i\ I_i\ R_i\ H_i\ S_i)$, where u_i is the Unicode of e_i, c_i is the short code name of e_i, d_i is the description of e_i, K_i is the set of keywords of e_i, I_i is the set of images of e_i, R_i is the set of related emoji of e_i, H_i is the set of categories of e_i, S_i is the set of emoji senses of e_i.

Emoji Resource	*u*	*c*	*d*	*K*	*I*	*R*	*H*	*S*
Unicode Website	✓	✓	×	✓	✓	×	×	×
Emojipedia	✓	✓	✓	×	✓	✓	✓	×
iEmoji	✓	✓	✓	✓	✓	×	✓	×
Emoji Dictionary	×	×	×	×	✓	×	×	✓

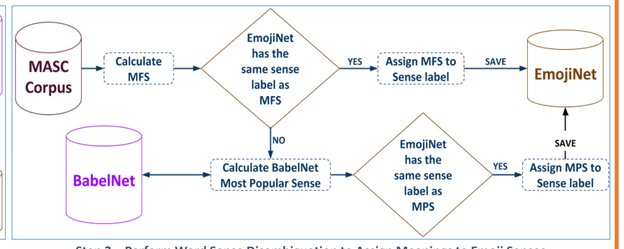
Table 1 – Emoji Data Available in Open Resources



Step 1 – Extract Data from Open Web Resources and Integrate them based on Unicode Character and Image Representations



Step 2 – Filter Senses and Part-of-Speech for each Emoji Sense to Populate EmojiNet



Step 3 – Perform Word Sense Disambiguation to Assign Meanings to Emoji Senses

Real World Examples

😂		🔫		💰	
Laugh(N)	Can't stop laughing 😂	**Anger(N)**	Angry as hell 🔫	**Costly(A)**	Can't buy class la 💰
Happy(N)	Got all A's but 1 😂😄	**Shot(N)**	Ooooooh shots fired! 🔫🔫	**Work Hard(N)**	Up early on the grind 💰
Funny(A)	That was damn hilarious! 😂😂😂	**Kill(V)**	He tried to kill one of my brothers 🔫	**Money(N)**	Earn money when one register /w ur link 💰

Evaluation

- ✓ The Emoji Dictionary was merged with other open resources by matching the images in The Emoji Dictionary with images found in other resources. A **nearest neighborhood-based image processing algorithm** was used for this task and its **accuracy was 98.42%**.
- ✓ Two Word Sense Disambiguation (WSD) algorithms based on **the Most Frequent Sense (MFS) and the Most Popular Sense (MPS)** were used to assign meaning to emoji senses extracted from The Emoji Dictionary. Their **combined WSD accuracy was 85.18%**.

	Correct	Incorrect	Total
Noun	1,271 (83.28%)	255 (16.71%)	1,526
Verb	735 (84.00%)	140 (16.00%)	875
Adjective	725 (90.06%)	80 (9.93%)	805
Total	**2,731 (85.1 8%)**	**475 (14.81%)**	**3,206**

Table 2 –Word Sense Disambiguation Statistics

Emoji Resource	u	c	d	K	I	R	H	S
# of Emoji with each Feature	1,074	845	1,074	1,074	1,074	1,002	705	875
# of Records Stored for each Feature	1,074	845	1,074	8,069	28,370	9,743	8	3,206

Table 3 – EmojiNet Statistics

EmojiNet at Work

Let's disambiguate the sense of the 🙏 emoji in the tweets T1 and T2, where T1:***Pray*** *for my family* 🙏 ***God*** *gained an angel today* and T2: *Hard to win, but we did it man* 🙏 *Lets* ***celebrate***. EmojiNet lists two senses for 🙏 and they are pray(V) and highfive(N). We extract words from sense definitions available in EmojiNet for the above two senses. They are: highfive(N) – {palm, high, hand, slide, **celebrate**, raise, person, head, five} and pray(V) – {worship, thanksgiving, saint, **pray**, higher, **god**, confession}. We then calculate the overlap of the words in tweets with the words extracted from the sense definitions in EmojiNet. This leads us to decide that the 🙏 emoji in T1 refers to pray(V) and the 🙏 emoji in T2 refers to highfive(N).

Future Work

- ✓ Expand EmojiNet sense definitions with words extracted from tweets, using a word embeddings model trained on tweets with emoji.
- ✓ Evaluate the usability of EmojiNet using Emoji Sense Disambiguation and Emoji Similarity Finding tasks and expose EmojiNet as a web service. EmojiNet demo is available at http://emojinet.knoesis.org/

Reference – Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, Derek Doran. EmojiNet: Building a Machine Readable Sense Inventory for Emoji. In 8th International Conference on Social Informatics (SocInfo 2016). Bellevue, WA, USA.

Acknowledgement – We acknowledge partial support from NIDA Grant No. 5R01DA039454-02: "Trending: Social Media Analysis to Monitor Cannabis and Synthetic Cannabinoid Use", NIH Award: MH105384-01A1: "Modeling Social Behavior for Healthcare Utilization in Depression", and Grant No. 2014-PS-PSN-00006 awarded by the Bureau of Justice Assistance.

Link to Paper

Link to Demo