Seeing the Best and Worst of Everything on the Web with a Two-level, Feature-rich Affect Lexicon

Tony Veale
Web Science and Technology Division
KAIST, Daejeon,
South Korea.
Tony.Veale@gmail.com

ABSTRACT
Affect lexica are useful for sentiment analysis because they map words (or senses) onto sentiment ratings. However, few lexica explain their ratings, or provide sufficient feature richness to allow a selective "spin" to be placed on a word in context. Since an affect lexicon aims to capture the affect of a word or sense in its most stereotypical usage, it should be grounded in explicit *stereotype* representations of each word's most salient properties and behaviors. We show here how to acquire a large stereotype lexicon from Web content, and further show how to determine sentiment ratings for each entry in the lexicon, both at the level of properties and behaviors and at the level of stereotypes. Finally, we show how the properties of a stereotype can be segregated on demand, to place a positive or negative spin on a word in context.

Categories and Subject Descriptors
I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *language models, lang. parsing and understanding, text analysis.*

General Terms
Algorithms, Measurement, Human Factors, Languages.

Keywords
Affect lexicons, stereotypes, acquiring common-sense knowledge

1. INTRODUCTION
Just as words have the potential to mean different things in different contexts, so too can their affective intent vary from one context to another. Thus, in some contexts one might feel complimented to be described as *cunning* or *aggressive*, but feel aggrieved and insulted to be so described in another. As concepts become more complex and multifaceted, and accrete more layers of stereotypical associations, their ability to assume different affective profiles in different contexts also increases. We therefore propose a two-level organization for the affective lexicon: at the first level, concepts are associated with a nuanced description of their stereotypical behaviors, as learned from the Web; at the second level, these behaviors are assigned an affective profile.

Consider the word "baby", used to denote a human infant. In some contexts the word carries a positive affect: babies can be cute and adorable, curious and trusting, and an obvious target of love and affection, especially when asleep. Crying babies, however, can be selfish, whining, drooling, hissing, tantrum-throwing little monsters. Both views are stereotypical of human babies, and either can be intended when a speaker uses the term "baby" figuratively, whether to describe a beloved partner or an annoying colleague. This is a matter of conceptual perspective, not of lexical sense, and many other words exhibit a similar affective duality; "teenager" for instance can mean "whining brat" just as easily as "growing adolescent". The concepts *Baby* and *Teenager* are complex and multifaceted, and different uses in context may highlight different stereotypical behaviors of each. Their affective meaning in context is therefore not so much a function of which lexical sense is intended but of which qualities are highlighted, and of the perceived affect of those qualities.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

2. MINING AFFECTIVE STEREOTYPES
We construct the affective stereotype lexicon in two stages. In the first stage, a large collection of stereotypical concept descriptions is harvested from the Web. As in [2], our goal is to acquire a lightweight common-sense representation of everyday concepts. In the second stage, we link these common-sense qualities in a *support graph* that captures how they mutually support each other in their co-description of a stereotypical idea. From this graph we can estimate pleasantness and unpleasantness scores for each property and behavior, and for the stereotypes that exhibit them.

Similes and stereotypes share a symbiotic relationship: the former exploit the latter as reference points for an evocative description, while the latter are perpetuated by their constant re-use in similes, especially on the Web. Expanding on the approach in [3], we use two kinds of query for harvesting stereotypes from the Web. The first, "as ADJ as a NOUN", acquires typical adjectival properties for noun concepts; the second, "VERB+*ing* like a NOUN" and "VERB+*ed* like a NOUN", acquires typical verb behaviors. Rather than use a wildcard * in both positions (ADJ and NOUN, or VERB and NOUN), which yields limited results with a search engine like Google, we generate fully instantiated similes from hypotheses generated via the Google n-grams (Brants and Franz, [1]). Thus, from the 3-gram "a drooling zombie" we generate the query "drooling like a zombie", and from the 3-gram "a mindless zombie" we generate the Web query "as mindless as a zombie".

We generate hundreds of thousands of speculative queries in this fashion, but only those that retrieve one or more Web documents via Google indicate the most promising associations. But this still gives us over 250,000 web-validated simile associations for our stereotypical model. We now filter these candidates manually, to ensure that the contents of the lexicon are of the highest quality (as we plan to re-use the lexicon in a wide variety of applications, it is worth the investment of a few weeks of labor). As a result, we obtain rich descriptions for many stereotypical ideas, such as *baby*, which is described via 163 typical properties and behaviors like *crying*, *drooling* and *guileless*. After this manual phase, the stereotype lexicon maps 9,479 stereotypes to a set of 7,898 properties and behaviors, to yield more than 75,000 pairings.

We construct the second level of the lexicon by automatically linking these properties and behaviors to each other in a support graph. The intuition here is that properties which reinforce each other in a single description (e.g. "as *lush and green* as a jungle" or "as *hot and humid* as a sauna") are more likely to have a similar affect than properties which do not support each other. We first gather all Google 3-grams in which a pair of stereotypical properties or behaviors X and Y are linked via coordination, as in "*hot and humid*" or "*kicking and screaming*". A bidirectional link between X and Y is then added to the support graph if one or more stereotypes in the lexicon contain both X and Y. If this is not so, we consider whether both descriptors ever reinforce each other in Web similes, by posing the Web query "*as X and Y as*". If this query has a non-zero hit set, we also add a link between X and Y.

3. ESTIMATING AFFECT: TWO LEVELS

Let **N** denote this support graph, and N(p) denote the set of neighboring terms to p, that is, the set of properties and behaviors that can mutually support p. Since every edge in **N** represents an affective context, we can estimate the likelihood that a property p is ever used in a positive or negative context if we know the positive or negative affect of enough members of N(p). Thus, if we label enough vertices of **N** with + or – labels, we can interpolate a positive/negative affect score for all vertices p in **N**.

To do this, we build a reference set **–R** of typically negative words, and a set **+R** of typically positive words. Given a few seed members of **–R** (such as *sad*, *disgusting*, *evil*, etc.) and a few seed members of **+R** (such as *happy*, *wonderful*, *pretty*, etc.), we easily find many other candidates to add to **+R** and **–R** by considering neighbors of these seeds in **N**. After just three iterations in this fashion, we populate **+R** and **–R** with approx. 2000 words each.

For a property p we can now define $N^+(p)$ and $N^-(p)$ as follows:

$$(1) \quad N^+(p) = N(p) \cap +\mathbf{R}$$

$$(2) \quad N^-(p) = N(p) \cap -\mathbf{R}$$

We can now assign positive and negative scores to each vertex p by interpolating from reference values to their neighbors in **N**:

$$(3) \quad \mathrm{pos}(p) = \frac{|N^+(p)|}{|N^+(p) \cup N^-(p)|}$$

$$(4) \quad \mathrm{neg}(p) = 1 - \mathrm{pos}(p)$$

If a term S denotes a stereotypical idea and is described via a set of typical properties and behaviors *typical(S)* in the lexicon, then:

$$(5) \quad \mathrm{pos}(S) = \frac{\sum_{p \in \mathrm{typical}(S)} \mathrm{pos}(p)}{|\mathrm{typical}(S)|}$$

$$(6) \quad \mathrm{neg}(S) = 1 - \mathrm{pos}(S)$$

Thus, (5) and (6) calculate the mean affect of the properties and behaviors of S, as represented via *typical(S)*. We can now use (3) and (4) to separate *typical(S)* into those elements that are more negative than positive (putting a negative spin on S) and into those that are more positive than negative (putting a positive spin on S):

$$(7) \quad \mathrm{posTypical}(S) = \{p \mid p \in \mathrm{typical}(S) \wedge \mathrm{pos}(p) > \mathrm{neg}(p)\}$$

$$(8) \quad \mathrm{negTypical}(S) = \{p \mid p \in \mathrm{typical}(S) \wedge \mathrm{neg}(p) > \mathrm{pos}(p)\}$$

Formulae (7) and (8) can be used to view a concept S positively or negatively in given context, to highlight only the properties of S that support such a positive or negative viewpoint.

4. EVALUATION

In the process of populating **+R** and **–R**, we identify a reference set of 478 positive stereotypes (such as *saint* and *hero*) and 677 negative stereotypes (such as *tyrant* and *monster*). When we use these reference points to test the effectiveness of (5) and (6) – and thus, indirectly, of (3) and (4) and of the stereotype lexicon itself – we find that **96.7%** of the positive exemplars are correctly assigned a positivity score greater than 0.5 (thus, $pos(S) > neg(S)$) while **96.2%** of the negative exemplars are correctly assigned a negativity score greater than 0.5 (thus, $neg(S) > pos(S)$).

We can also use **+R** and **–R** as a gold standard for evaluating the separation of *typical*(S) into the distinctly positive and negative subsets *posTypical(S)* and *negTypical(S)*. The lexicon contains 6,230 stereotypes with at least one property in **+R**∪**–R**, and on average, **+R**∪**–R** contains 6.51 of the properties of each of these stereotypes (on average, 2.95 are in **+R** while 3.56 are in **–R**).

Table 1. Average P/R/F1 scores for the affective retrieval of positive and negative properties from 6,230 stereotypes

***Macro Average* (6230 stereotypes)**	**Positive properties**	**Negative properties**
Precision	.962	.98
Recall	.975	.958
F-Score	.968	.968

In a perfect separation, (7) should yield a positive subset that contains only those properties in *typical*(S)∩**+R,** while (8) should yield a negative subset that contains only those in *typical*(S)∩**-R**. Viewing the problem as a retrieval task, in which (7) and (8) are used to retrieve distinct positive and negative property sets for a stereotype S, we report the encouraging results of Table 1 above.

5. ACKNOWLEDGMENTS

This research was supported by the WCU (World Class University) program under the National Research Foundation of Korea, and funded by the Ministry of Education, Science and Technology of Korea (Project No: R31-30007).

6. REFERENCES

[1] Brants, T. and Franz, A. Web 1T 5-gram Version 1. Linguistic Data Consortium, 2006.

[2] Liu, H., Lieberman, H. and Selker, T. A Model of Textual Affect Sensing Using Real-World Knowledge. *Proceedings of the 8th international conference on Intelligent user interfaces*, pp. 125-132. 2003.

[3] Veale, T. Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity. In Proceedings of ACL'2011, the 49th *Annual Meeting of the Association of Computational Linguistics*. June 2011.