In the Mood for Affective Search with Web Stereotypes

Tony Veale
Web Science & Technology Div.
KAIST, Yuseong,
Daejeon, South Korea.
+353-86-171-4506
Tony.Veale@gmail.com

Yanfen Hao
School of Computer Science,
University College Dublin,
Belfield, Dublin D4, Ireland.
+353-1-716-2912
Yanfen.Hao@gmail.com

ABSTRACT
Models of sentiment analysis in text require an understanding of what kinds of sentiment-bearing language are generally used to describe specific topics. Thus, fine-grained sentiment analysis requires both a topic lexicon and a sentiment lexicon, and an affective mapping between both. For instance, when one speaks disparagingly about a city (like London, say), what aspects of *city* does one generally focus on, and what words are used to disparage those aspects? As when we talk about the weather, our language obeys certain familiar patterns – what we might call clichés and stereotypes – when we talk about familiar topics. In this paper we describe the construction of an *affective stereotype lexicon*, that is, a lexicon of stereotypes and their most salient affective qualities. We show, via a demonstration system called MOOD*finger*, how this lexicon can be used to underpin the processes of affective query expansion and summarization in a system for retrieving and organizing news content from the Web. Though we adopt a simple bipolar +/- view of sentiment, we show how this stereotype lexicon allows users to coin their own nuanced moods on demand.

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
Affective computing, lexicon design, lexical affect, common-sense knowledge, knowledge acquisition from Web texts.

1. INTRODUCTION
Context exerts a powerful yet largely unseen influence on our interpretation of natural language words and utterances. It is context that primes our expectations, to focus our attention on just those shades and senses of a word that are relevant. In context, a word seems to mean just what it is intended to mean, and carry just the right emotional overtones and mood. But viewed out of context, the mapping of words to affect is never quite so direct. Just as words can have many senses, so too can they have a multiplicity of affective uses. Good writers craft their sentences to bring out just the right senses and moods of their words, and to focus our attention on the qualities the speaker wants highlighted. When it comes to retrieving these texts from the Web, we need to give the search user the same power to suggest sense and mood.

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

For instance, a search user might require that only texts that express a certain mood are retrieved for a given query, or that the texts which are returned are ranked according to the degree to which they exhibit this mood. An affective lexicon, such as the one described here, allows a system to perform an *affective query expansion*, so that a single mood term (like *crazy*) can be expanded into a large family of related terms that carry a similar affect (such as *mad*, *unbalanced*, *risky*, etc.). But users, as the authors of queries, should be able to exploit the same nuances of lexical affect as the authors of the retrieved texts themselves. For instance, the word *crazy* can carry a positive or a negative sentiment: though usually negative, it can also be given an overtly positive spin to describe *risky, quirky, brilliant, adventurous, creative* and *artful* endeavors. Our affect lexicon must not only provide us with a default sentiment for each entry (based on corpus analysis) but must also allow us to override this default. In this way, we can place our own affective spin on a query term, to e.g., search for texts that exhibit the mood ***+crazy*** or ***–crazy***.

We describe the workings of such a system, called MOOD*finger*, in section 4. But first, section 2 outlines how MOOD*finger*'s lexicon of affective stereotypes, called MOOD*prism*, is constructed from Web content, while section 3 shows how additional affective norms are extracted from the Google n-grams.

2. THE MOOD*prism* AFFECT LEXICON
We construct the MOOD*prism* lexicon in two stages. In the first stage, a large collection of stereotypical concept descriptions is harvested from the Web. These descriptions capture the most typical properties and behaviors of many everyday concepts. In the second stage, we link these properties and behaviors in a *support graph* that captures how these elements mutually support each other in the description of a complex idea. From this graph we can estimate pleasantness and unpleasantness scores for each property and behavior, and for the stereotypes that exhibit them.

Since stereotype representations are acquired from the Web, and reflect an open-ended common-sense view of word meaning (as opposed to the narrow semantic meanings found in dictionaries), the affect lexicon is very much grounded in common-sense knowledge, much like the approach of Liu *et al*. [2] Likewise, though we focus here on the estimation of simple positive and negative affect scores for each property and each stereotype, we show in the next section how the interconnectedness of these properties allows us to provide users with the ability to specify their own complex ad-hoc moods. These moods capture a wide range of emotions, more fine-grained even than those in [4].

Veale & Hao [5] identify a symbiotic relationship between similes and stereotypes: the former exploit the latter as a reference point for an evocative description, while the latter are perpetuated by their constant use and re-use in similes, especially on the Web. We can thus build a rich model of stereotypical beliefs by harvesting similes on a large scale from the Web. We expand here upon the method we first outlined in [5], and use two kinds of query for harvesting similes. The first, "as ADJ as a NOUN", acquires typical adjectival properties for nouns; the second, "VERB+*ing* like a NOUN" and "VERB+*ed* like a NOUN", acquires typical verb behaviors for nouns. 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 text fragments from hypotheses generated using the Google n-grams (Brants and Franz, [1]). Thus, given the 3-gram "a shambling chimp", we generate the Web query "shambling like a chimp", and for the 3-gram "a hairy ape" we generate the query "as hairy as an ape".

We generate hundreds of thousands of speculative queries in this fashion, and those that retrieve one or more Web documents via Google indicate the most promising associations. But this still gives us over two hundred thousand promising candidates 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 pairs 9,479 stereotypes with a choice of 7,898 properties and behaviors, to yield over 75,000 pairings in total.

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 still add a link between X and Y.

Let N(p) denote the set of neighboring terms to p in the support graph. In other words, N(p) denotes the set of properties and behaviors that can mutually support p. Intuitively, if we know the positive or negative affect of enough members of N(p), we can estimate the affect of p. More generally, if we label enough elements of the graph with + or – labels, we can estimate a positive/negative affect score for all the elements in the graph **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 p as follows:

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

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

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:

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

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

That is, we simply calculate the mean affect of the properties and behaviors of S, as represented in the lexicon via *typical(S)*. Note that (5) and (6) are simply gross defaults. One can always use (3) and (4) to separate the elements of *typical(S)* into those which are more negative than positive (putting a negative spin on S) and those which are more positive than negative (a positive spin on S):

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

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

As we will see in sections 3 and 4, this ability to selectively focus on just the positive or the negative qualities of a stereotype is particularly useful for the affective expansion of user queries.

2.1 Evaluating the MOOD*prism* Lexicon

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)*).

The lexicon contains 6,230 stereotypes with at least one property or behavior 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**). We can also use **+R** and **–R** then as a gold standard for evaluating the separation of *typical*(S) into the distinctly positive and negative subsets *posTypical(S)* and *negTypical(S)*. Viewing this separation as a retrieval task, in which (7) and (8) are used to retrieve distinct positive and negative property/behavior sets from each of 6,230 stereotypes, we can report macro-averaged P/R/F1 scores for *posTypical*(S) of (P = .962, R = .975, F1 = .968), and comparable macro-averaged scores for *negTypical*(S) of (P = .98, R = .958, F1 = .968).

3. AFFECTIVE '*TALKING POINTS*'

As concise and highly concentrated bundles of meaning, stereotypes make for potent query terms that can be affectively expanded. For instance, the query "*Iraq terrorism* ***–terrorist***" can be used to retrieve documents about terrorism in Iraq that explicitly allude to the negative qualities and behaviors of terrorists. The retrieved documents will be ranked not just by their relevance to terrorism in Iraq, but by their relative density of properties like *shocking*, *condemned*, *sickening* and over 100 other negative-leaning words that we use to describe a typical *terrorist*.

Terrorist is a highly charged term, one with many highly-charged qualities that it make it a useful element of an affective query. However, many of the query terms we would like to affectively expand are not intrinsically charged, and have few or no affective qualities that one might consider stereotypical. So queries like "Ireland **+economy**" and "London **–city**" require a system to look beyond the internal structure of the specific concepts, to consider the broader domains of *economies* and *cities* more generally, to find the +/- topics that one usually highlights in these domains.

To identify frequent *affective talking points* for a given query topic, we use the Creative IR engine of Veale [6] to match non-literal patterns in the Google n-grams [1]. This pattern, for example, finds negative talking points for *city* in Google 5-grams:

-R*/noun in the* <u>*city*</u> .

Here ***-R****/noun* denotes any noun in our reference set of negative words. Matching terms include: *crime, violence, traffic, pollution, congestion, strife, poverty, homelessness, gangs, chaos* and *riots*. A similar pattern retrieves the following negative talking points for ***-economy*** – *slowdown, pressures, inflation, weakness, decline, imbalance, uncertainty, risk, slump, recession, unemployment and instability* – while the ***+R****/noun* variant retrieves the following positive talking points for ***+economy*** – *growth, confidence, investment, improvement, resources, credit, income, efficiency, capacity, stability, recovery, strength, innovation* and *power*.

As we show next, these +/- talking points are used to expand the terms in a user's query: query terms prefixed with + are expanded with a set of positive talking points; those prefixed with – are expanded with a set of negative talking points; unadorned query terms are expanded with both positive *and* negative talking points.

Creative IR, applied to the Google n-grams, yields a rich set of popular talking points for frequent topics. For instance, the pattern ***-R****/noun of* <u>*religious*</u> ***-R****/noun* allows a retrieval engine to identify the concerns that a user is most likely alluding to with **-religious**: *minorities, extremists, persecution, intolerance, fanaticism, discrimination, prejudice, fundamentalism, superstition*, etc. We might now add this term **-religious** to our query on "terrorism in Iraq" to find articles about the religious causes of this violence. Since many talking points are also stereotypes with rich property-level representations in the MOOD*prism* lexicon, our system can finely analyze the affective coherence of any texts that it retrieves.

4. THE MOOD*finger* AFFECT ENGINE

Veale [6] defines *creative information retrieval* as a form of IR in which there exists a non-literal relationship between the elements of a query and the elements of the retrieved texts. For instance, a query might specify that a matching document should contain a stereotype for a given property, or conversely, contain one or more properties that are typical of a given stereotype. Veale's approach is designed to serve as an IR platform for a wide range of creative language applications – such as metaphor retrieval and simile generation. We extend that platform here, with our two-level affective lexicon MOOD*prism*, to support affective search for news content on the Web. Currently, news articles are crawled from a dozen online newspapers (this number will grow in time) and their textual content is indexed using the *Lucene* system [3]. Hourly updates are also obtained from RSS feeds. Queries to the system are separated into two kinds of terms: regular query terms, which are unadorned keywords or phrases; and *mood terms*, which are terms prefixed either with + or – to indicate their affective "spin" (such as ***–proud*** or ***+cunning***). All terms are automatically expanded using stereotypical knowledge (section 2) and affective *talking points* (section 3), though +/- mood terms are expanded only with elements that have a matching +/- affect.

4.1 Affective Expansion and Ad-Hoc Moods

Mohammad and Yang argue in [4] that humans can reliably categorize words by more than mere +/- polarity. They show e.g. that some words convey *sadness* and *fear* to different degrees, while others suggest a degree of *joy* and even *trust*. While we do not explicitly distinguish different dimensions of mood or emotionality in MOOD*finger*, the system does support a whole lexicon of ad-hoc mood types, via mood terms like **+aggressive**.

A word like *aggressive* implies a wide range of positive qualities that are captured by N^+(*aggressive*), and a broader range of negative qualities that are captured by N^-(*aggressive*). The 171 words in N^+(*aggressive*) convey the up-side of aggressive behavior (e.g. being *quick*, *energetic*, *vigorous* and *determined*) while the 219 words in N^-(*aggressive*) convey the down-side of aggressiveness (e.g. being *violent*, *angry*, *hostile* and *abusive*). A user-query containing **+aggressive** is thus expanded with the elements of N^+(*aggressive*), while a query containing **–aggressive** is expanded using N^-(*aggressive*). When retrieved documents are ranked by query-specific relevance, those documents that exhibit a more pervasive sense of aggression will thus be ranked highest.

A query-specific affective summary is generated for each document, by first identifying the sentences that match the original *un*expanded query, and by then scoring each of these according to how well it also fits the expanded query. Additional weight is given to affective terms in each of these sentences if their +/- polarity matches that of a +/- mood term in the user's query. The top 20 documents are re-ranked according to the quality of the affective summary that can be generated for each, and a *retrieval digest* then compiles these summaries, in re-ranked order, into a single affective summary of the retrieval set.

For each retrieval set, MOOD*finger* generates a *mood cloud* of the most frequently matched affective terms in the expanded query. Thus, for the query "*Europe* ***–anxiety***" the words **debt** and **crisis** are given the most visual salience, while for "*Korea* ***-anxiety***" it is words like **war** and **attack** that receive the most prominence.

4.2 Stereotypes in Context

Because stereotypes serve as conceptual landmarks for anchoring highly evocative and often charged meanings, words denoting stereotypes and their salient properties are thus amongst the most interesting terms that one can find in any text. This is also true of the phrasal level: phrases comprising a stereotype head and a property modifier (not necessarily typical of the head) are dense with meaning and often worthy of special emphasis. Phrases like "*boring movie*", "*clever strategy*" and "*visionary leader*" provide a context for mood-carrying words that individual terms alone (like *boring* or *clever* or *leader*) cannot convey in isolation.

For each retrieval, MOOD*finger* compiles a list of every evocative phrase – i.e. any phrase that combines multiple stereotypes, or any that combines stereotypes and stereotypical properties – that is found in a retrieved document. Only those phrases that contain at least one affect-carrying term from the affective expansion of the user's query are considered relevant, so the resulting list contains only those phrases that capture the theme and mood of the query. This list is then displayed as a phrase cloud, in which the most frequent phrases in the retrieval set have the greatest prominence. Clouds of phrases gleaned from the results of two MOOD*finger* queries are shown in Figure 1 at the end of the paper.

Consider the query "*Steve Jobs* ***+leader***". A great many glowing tributes have been paid to this legendary figure since his passing in 2011, so we can expect a large number of news articles to be retrieved, each dense with the qualities of a good leader, such as *capable, honored, dashing, charismatic, trusted* and *accomplished* (note: *typical(leader)* contains over 50 properties and behaviors). Conversely, "*Steve Jobs* ***-leader***" draws our focus to those texts that allude to the negative qualities of Jobs as a leader, such as *brash*, *ruling* and *exiled*, while , "*Steve Jobs* ***+inventor***" focuses on his creative capabilities. In each case, the resulting document set is effectively summarized by a phrase cloud that shows the most frequent phrases that evoke the mood of the query, such as "*visionary leader*" (Figure 1, bottom) or "*creative genius*" (top). Each cloud provides an affective map of the retrieval set, in which all but the most salient landmarks have been removed.

5. SUMMARY AND CONCLUSIONS

Stereotypes condense a great deal of common-sense knowledge into a single term or idea. We simply need to mention a stereotypical word-concept in context for its rich panoply of salient associations to become active in the minds of an audience. Writers exploit stereotypes to anchor the moods and the meanings of their texts, so it makes sense that language processing and IR should also exploit an explicit model of stereotypicality, at the level of both concepts/words and of properties/behaviors.

MOOD*finger* represents an initial attempt to model an affective lexicon around a stereotypical belief system that has been acquired from the Web, so that these stereotypes can be used to retrieve, filter and rank Web content in the most emotionally-useful ways. We believe there is still a great deal of traction to be gained from stereotypes in the affective processing of Web texts, and continue to research new ways of exploiting Web stereotypes.

6. ACKNOWLEDGMENTS

This research was partly supported by the WCU (World Class University) program under the National Research Foundation of Korea (Ministry of Education, Science and Technology of Korea, Project No: R31-30007), and partly funded by Science Foundation Ireland via the Centre for Next Generation Localization (CNGL).

7. 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] McCandless, M., Hatcher, E. and Gospodnetić, O. *Lucene In Action (2nd Edition)*. Manning Publications. 2010.

[4] Mohammad, S. F. and Yang, T. W. Tracking sentiment in mail: how genders differ on emotional axes. *In Proceedings of the ACL 2011 WASSA workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Portland, Oregon. 2011.

[5] Veale, T. and Hao, Y. Making Lexical Ontologies Functional and Context-Sensitive. *Proceedings of ACL-2007, the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 57–64. June 2007.

[6] 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.

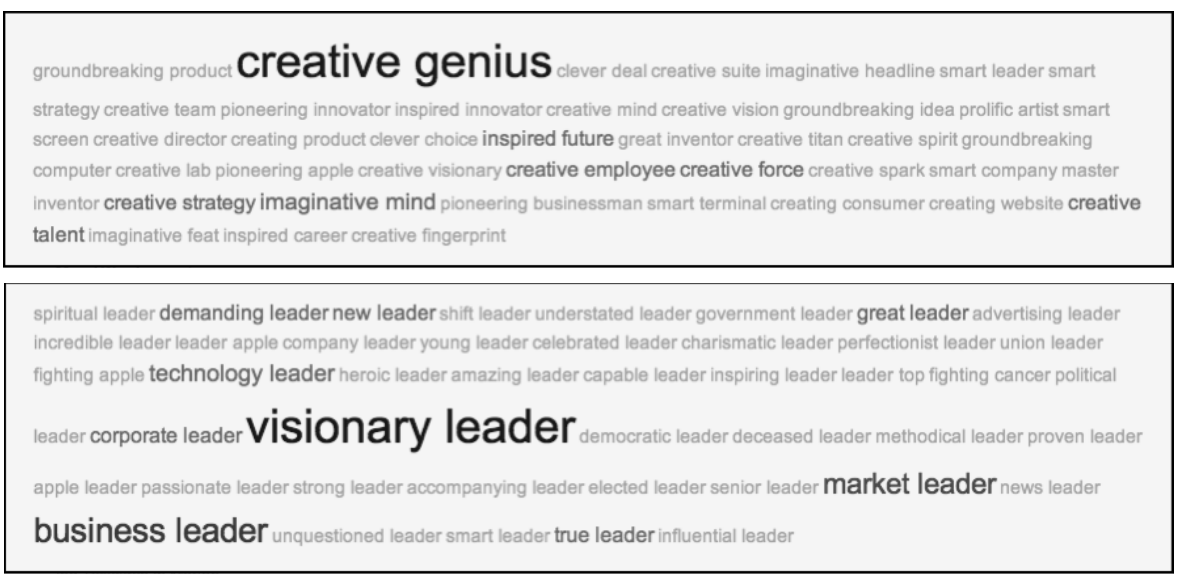

Figure 1. *Top*: A cloud of evocative phrases (stereotypes in context) for the query "*Steve Jobs +inventor*". *Bottom*: a cloud of evocative phrases for the query "*Steve Jobs +leader*". Each phrase is clickable, to retrieve corresponding documents.