A World of Difference: Divergent Word Interpretations among People

Tianran Hu,[*1] **Ruihua Song,**[†2] **Maya Abtahian,**[*3] **Philip Ding,**[†4] **Xing Xie,**[†5] **Jiebo Luo**[*6]

[*]University of Rochester, [†]Microsoft Research
{[1]thu, [6]jluo}@cs.rochester.edu
[3]maya.r.abtahian@rochester.edu
{[2]song.ruihua,[4]phding,[5]xing.xie}@microsoX.com

Abstract

Divergent word usages reflect differences among people. In this paper, we present a novel angle for studying word usage divergence – word interpretations. We propose an approach that quantifies semantic differences in interpretations among different groups of people. The effectiveness of our approach is validated by quantitative evaluations. Experiment results indicate that divergences in word interpretations exist. We further apply the approach to two well studied types of differences between people – gender and region. The detected words with divergent interpretations reveal the unique features of specific groups of people. For gender, we discover that certain different interests, social attitudes, and characters between males and females are reflected in their divergent interpretations of many words. For region, we find that specific interpretations of certain words reveal the geographical and cultural features of different regions.

Introduction

As Pinker stated in his book (Pinker 2007), *"language is the window into human nature"*. The differences among people, for example, gender (Coates 2015), age (Nguyen et al. 2013), and occupation (Hu et al. 2016b) trigger their different word usages. These divergences in word usages, in turn, reveal the unique features of groups of people, such as their specific interests (Rakesh et al. 2014), personalities (Vinciarelli and Mohammadi 2014), and so on. The previous work in differences among people mostly focuses on word frequency. In this work, we take a step further to study word interpretation. Our findings suggest that the interpretations of words may vary from people to people. More importantly, we discover that divergent word interpretations are also related to the unique features of people, which provides a novel angle to comprehend the differences between social groups.

Intuitively, a word can be interpreted using its semantically close words in a corpus. The idea of our approach is to extract words' semantically closest words as their interpretations from corpora of different groups of people, separately. The widely adopted word embedding techniques are employed to learn the closest words. We then quantify the semantic distance between the multiple interpretations of

each word. The larger the distances is, the more differently a word is interpreted among people. We refer to the semantic distance between the interpretations of a word as its ***divergence score***, and the words of high divergence scores as ***divergent words***.

We apply the approach to studying two widely studied differences among people – gender and regional differences. The discovered divergent words clearly portray the unique features of specific populations. For gender, we observe the different interests between men and women in various aspects. As for the regionally divergent words, we observe that they are usually related to the geographical or cultural features of regions. In this paper, we report the detected gender and regionally divergent words, and summarize the unique features of populations as read from these words.

Related Work

Much work has focused on divergent word usages among people in numerous aspects, such as gender, age, occupation, and region. It is reported that typical male language uses more judgmental adjectives, elliptical sentences, directives, and "I" reference, while typical female language contains more intensive adverbs, references to emotions, uncertainty verbs (Poynton 1989). People in different age groups appear to choose words differently. Nguyen et al. reported that, comparing with younger people, older people talk more about family and work, and use less swear words in their language (Nguyen et al. 2013). Jobs affect the language patterns of people as well (Hu et al. 2016b). In the paper, the authors listed the most used words of several occupations, and their results indicate clear divergences in word usages between people of different jobs. Similarly, different cities may have different preferences in words (Cheng, Caverlee, and Lee 2010), which reflect regional differences between people.

Data Collection and Preprocessing

We collect our dataset via the Twitter open APIs[1]. First, we set up two geo-bounding boxes. Both boxes are two degrees of latitude long, and two degrees of longitude wide. One box is centered at NYC, and the other is centered at the Bay Area. We use the Twitter streaming API to collect users who

---

[1]https://dev.twitter.com/overview/api

have posted tweets within these two areas from June to October 2016, resulting in 0.4 million unique users. For each unique user, we download their at most 3,000 recent tweets with the Twitter timeline API, resulting in around 0.2 billion tweets.

To assign gender tags to each user, we employ the API of genderize.io[2] suggested in (Abbar, Mejova, and Weber 2015). Genderize.io takes a first name as input, and outputs the probabilities of this name being used by males and females. We feed the API with the first names appearing in user profiles to obtain gender tags, and filter out the names with low confidence (probability < 0.7). We obtain around 32 thousand males, and 30 thousand females. To figure out user locations, we follow the approach suggested in (Sadilek and Kautz 2013). We collect all the geo-tagged tweets from a user's recent tweets, and filter out the user if the amount of geo-tagged tweets was less than 10, or if less than half of the tweets were sent inside our bounding boxes. This step leaves us with roughly 33 thousand NYC users, and 33 thousand Bay Area users.

We further clean these raw tweets by removing hashtags, mentions, URLs, retweets, and short tweets (less than 10 words). We also exclude the words that are used by few people (less than 100 people) from these tweets, and set all words to lowercase. After cleaning, we separate these tweets according to user gender and location. Therefore, we obtain four corpora: male and female corpora denoted by C_m and C_f, contains 7.4 million, and 5.4 million tweets, respectively; and NYC and Bay Area corpora denoted by C_N and C_B, contains 8.1 million and 7.9 million tweets, respectively.

Methodology

Our approach consists of two steps. First, it learns group specific interpretations for each word. Second, it quantifies the semantic distance between the interpretations. Without loss of generality, we denote the corpora of two groups of people as C_1 and C_2. These groups of people can be people of different genders, from different regions, or of any other type of difference, and our approach can be easily extended to more than two groups.

Group Specific Word Interpretations

To learn the word interpretations for two groups of people, we feed their corpora to an embedding model, and obtain a word vector space for each group. Hence, using C_1 and C_2, we train two word vector spaces S_1 and S_2, respectively. Each space contains the semantic patterns of a group. Since words' closest words in an embedding space provide effective clues to understanding the words, we take these closest words as their interpretations. Therefore, for each word w, we collect two sets of its top n closest words from S_1 and S_2, respectively, along with the semantic similarities between these words to w. We refer to such sets as w's **Interpreting Sets**, denoted by I_w^1 and I_w^2.

The left part of Figure 1 shows an example of extracting the interpretations of *bitter* for males and females. In two
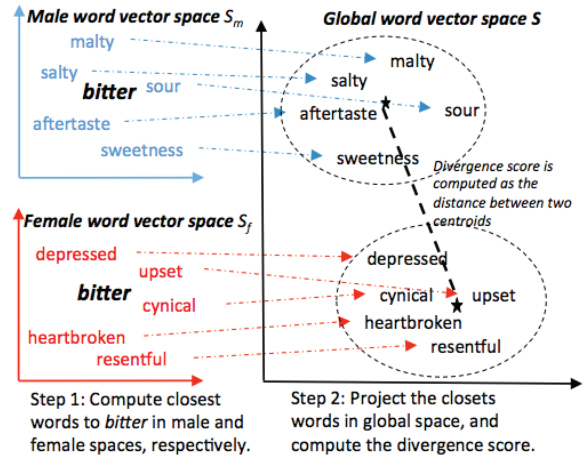
Figure 1: Illustration of the calculation of the divergence score of *bitter* from male and female corpora. For the brevity of the illustration, we do not note the similarities between closest words and word *bitter* on the plot.

word vector spaces of the two genders, we find the neighbors of the word, and obtain its two interpreting sets I_{bitter}^m and I_{bitter}^f. For example, set $n = 3$, then the two sets would be as follows:

$$I_{bitter}^m = \{(salty, 0.7), (sour, 0.6), (aftertaste, 0.6)\}$$

$$I_{bitter}^f = \{(upset, 0.6), (depressed, 0.6), (cynical, 0.6)\}$$

where the value after each word denotes its semantic similarities to *bitter*. The interpreting sets of a word represent its group specific understandings. If w is interpreted similarly by two groups, I_w^1 and I_w^2 should be similar as well. Otherwise, the difference between two sets reflects the divergence between w's interpretations.

Quantified Interpretation Divergences

To quantify the interpretation divergence, we first project the words in both interpreting sets to a global word vector space S. In this space, we then compute the semantic distance between these two groups of words. The global space S is trained using both corpora ($C_1 + C_2$), and contains the overall language patterns of both groups. To be specific, we find the corresponding word vectors in S for words in I_w^1 and I_w^2, respectively. We then compute the centroids of two groups of words in the global space as the weighted average of the word vectors, denoted by $cent(I_w^1)$ and $cent(I_w^2)$. The similarities between words and w are the averaging weights. These two centroids are used as the representatives of the interpreting sets in S, as suggested in (Reisinger and Mooney 2010). The cosine distance between two centroids is then calculated to measure the semantic difference between the interpreting sets. We refer to this distance as w's divergence score between two groups of people. Note that, the calculation in a global space takes the semantic meanings of closest words into account. Meanwhile, the weighted average assigns different importance to the words in interpreting sets.

The right part of Figure 1 shows an example of quantifying the differences between male and female interpreting sets of *bitter*. Centroids (denoted by asterisks in the figure) of interpreting sets are first computed. For example, $cent(I_{bitter}^m)$ is an aggregated word vector in S, and calculated as:

$$\frac{0.7 \times v(salty) + 0.6 \times v(sour) + 0.6 \times v(aftertaste)}{0.7 + 0.6 + 0.6}$$

where $v(salty)$ denotes $salty$'s word vector in S. The cosine distance between $cent(I_{bitter}^m)$ and $cent(I_{bitter}^f)$ is then computed as *bitter*'s divergence score.

Gender Divergent Words

We manually check the male and female interpretations (words in two interpreting sets) of the top 100 gender divergent words, in terms of divergent score. We observe that people of one gender are more likely to relate words to the specific interests of this gender.

Sports Males are generally more interested in sports than females (Deaner, Balish, and Lombardo 2015). This conclusion can be derived from the observation that many words associated with sports by men are interpreted quite differently by women. Such words include *title*, *finals*, *draft*, and many others. The closest words to *title* for males are *championship*, *contender*, and *undefeated*. This interpretation indicates that males are more likely to connect the word with championship titles. Differently, female closest words to the word are *draft*, *rewrite*, and *novel*, implying article and book titles. *Finals* is related to the final tournaments in sport for males (closest words: *playoffs*, *tourney*, and *semifinals*). For females, it refers to final exams (closest words: *midterms*, *midterm*, and *assignments*). Males are more like to connect the word *draft* with the player selection procedures of professional sport teams, with closest words such as *drafted*, *rounders*, and *76ers*. For females, the word is closest to *manuscript*, *revision*, and *outline*, implying literary drafts.

Fashion Females are more into fashion than males, as reported in (Goldsmith, Moore, and Beaudoin 1999). The word *styles*, for example, is more related to dressing styles for females. They connect this word with *nail*, *hair*, *lipsticks*, and *color*. Different for males, this word is more related to design styles, and closest to *themes*, *designs*, and *typefaces*. The difference also reflects in their interpretations of *shoots*. Females are more likely to use the word to mean shooting pictures, with closest words such as *posing*, *photoshoot*, and *boudoir*. Males refer to this word as firing guns (closest words: *kills*, *shooting*, *fatally*). Another example is *navy*. Females interpret this word as a style and color, and tend to use dressing related terms to interpret it, such as *collar*, *vest*, *berets*, and *striped*. Not surprisingly, males relate this word to military (closest words: *seal*, *marine*, and *airforce*).

IT & Video Games According to previous work (Gefen and Straub 1997), men show more interest in IT than women, and many gender divergent words provide evidence of it. An obvious example is *windows*. Males tend to refer to the word as the operating system introduced by Microsoft, and its closest words include names of other operating systems such as *win8*, *win7*, *ubuntu*, and *linux*. For females, *windows* is more related to the architectural structure, and closest to *sunroof*, *doors*, and *curtains*. *Bugs* is also assigned divergent interpretations by two genders. It stands for software bugs for male (closest words: *fixes*, *kernel*, *xcode*, and *glitches*), while it is associated with insects by females (closest words: *mosquitoes*, *ants*, *rodents*, and *spiders*).

Men's enthusiasm for video games is clearly reflected in their interpretations of some words that relate to this topic. *Destiny*, for example, is related to the famous online game for male, since its closest words are *halo* (video game series), *bungie* (vedio game company), and *xbox* (video game console). Similarly, word *steam* is associated with the video game digital distribution software by males, indicated by closest words such as *valve* (video game company that developed Steam), *xbox*, and *hearthstone* (online game). The explanations of these two words for females are not related to video games at all. Female closest words to *destiny* are *fate*, *greatness*, and *greatness*, and the closest words to *steam* are *compressor*, *dust*, and *heating*.

Regionally Divergent Words

In this section, we discuss words with divergent interpretations between NYC and the Bay Area. The differences between regions such as lifestyles (Hu et al. 2016a), eating habits, and cultures (Silva et al. 2014) are reported in many studies. In this work, we also manually check the top 100 regionally divergent words. We observe that there are mainly two types of regionally divergent words: one type reflects geographical differences, and the other reflects the cultural differences between regions.

Geographical Differences between Regions

Place names account for a large part of this type of divergent words. Such words include *queens* (Queens Borough, NYC), *Jose* (San Jose City, Bay Area), *castro* (Castro District, Bay Area), *buffalo* (Buffalo City, NY state), and so on. An interesting example of this type is *berkeley*. In NYC, the word is referred to as the famous university, and closest to other famous universities such as *princeton*, *cornell*, *dartmouth*, and *columbia*. This word is more likely to be interpreted as the Berkeley City in the Bay Area with closest words such as *pasadena*, *rockridge*, *cerrito* (all three are place names in the Bay Area).

There are also words, which although are not place names, are assigned geographical interpretations in certain regions. *Bart*, meaning Bay Area Rapid Transit in the Bay Area, is a typical example. Its closest words in the Bay Area are *train*, *muni* (Muni Transit), *subway*. In NYC, it is more referred to as a first name. Interestingly, its interpreting set contains *marge*, *homer*, *lisa*, and so on (Bart, Homer, Marge, Lisa are members of the Simpson family in the popular cartoon TV show).

Cultural Differences between Regions

More importantly, some regionally divergent words reveal the unique cultural features of regions. We summarize the features in several aspects, such as IT, finance, and so on.

IT The engineering culture in the Bay Area is clearly reflected in the observations that people there are more likely to relate many words to technologies. Such words include *react*, *code*, *port*, and many others. *React* in the Bay Area is interpreted as the reacting of software, indicated by closest words such as *webpack*, *filesystem*, and *objc* (short for Objective-C). In NYC, the word is interpreted literally with closest words such as *respond*, *confuse*, *reacting*, and so on. *Code*, not surprisingly, in the Bay Area is associated with programming, as interpreted by technical terms such as *compiler*, *parser*, *regex*, and *html*. In NYC, this word is more related to coupon codes or discount codes (closest words: *coupon*, *discount*, and *promocode*). As to *port*, this word means harbor in NYC, with closest words such as *terminal*, *ferry*, and *dock*. In the Bay Area, although one of the closest word in its interpreting set is *tacoma* (Tacoma Port, the Bay Area), its other closest words are all related to hardware ports such as *ethernet*, *connector*, and *eero* (Eero Inc., a Wi-Fi company).

Fashion & Arts Word interpretations in NYC, one of the world's great fashion and art capitals, are biased to fashion and arts too. Beside *carpet*, which we discussed in the introduction, such words include *model*, *string*, *metal*, and so on. In the Bay Area, *model* usually refers to mathematical models, and its closest words are *freemium* (a pricing strategy), *differentiator*, and *hybrid*. The word in NYC is connected with fashion models, and closest to *designer*, *supermodel*, and *stylist*. The popularity of music in NYC is clearly reflected in some regionally divergent words. For example, *string* is associated with string instruments in NYC (closest words: *strumming*, *guitars*, and *revolver*). Differently, in the Bay Area, the word is referred to as the data structure, which is closest to *dict*, *array*, and *ascii*. Similarly, *metal* is highly related to heavy metal music in NYC, with closest words such as *thrash* (thrash metal, a type of heavy metal music), *sevenfold*, and *metallica* (both are names of heavy metal bands). In the Bay Area, this word is interpreted as its literal meaning, and closest to *brass*, *rubber*, and *aluminum*.

Conclusion & Future Work

In this paper, we have discussed divergent word interpretations among people. To detect these words, we propose an approach based on word embedding models, which quantifies differences in word interpretations. We apply the approach to two types of widely studied differences between people – gender and regional differences. The discovered divergent words reveal the unique features of specific demographic groups of people, such as gender specific interests, cultural differences between regions, and so on. In the future, we would like to apply the approach to other types of human differences, such as age and occupation, and study the divergent word interpretations of these groups of people. Another possible direction is "difference between differences". It is interesting to study if some types of differences are more likely to lead to divergent word interpretations than others.

References

Abbar, S.; Mejova, Y.; and Weber, I. 2015. You tweet what you eat: Studying food consumption through twitter. In *ACM CHI*, 3197–3206. ACM.

Cheng, Z.; Caverlee, J.; and Lee, K. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *ACM CIKM*, 759–768. ACM.

Coates, J. 2015. *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.

Deaner, R. O.; Balish, S. M.; and Lombardo, M. P. 2015. Sex differences in sports interest and motivation: An evolutionary perspective.

Gefen, D., and Straub, D. W. 1997. Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS quarterly* 389–400.

Goldsmith, R. E.; Moore, M. A.; and Beaudoin, P. 1999. Fashion innovativeness and self-concept: a replication. *Journal of Product & Brand Management* 8(1):7–18.

Hu, T.; Bigelow, E.; Luo, J.; and Kautz, H. 2016a. Tales of two cities: Using social media to understand idiosyncratic lifestyles in distinctive metropolitan areas. *IEEE Transactions on Big Data*.

Hu, T.; Xiao, H.; Luo, J.; and Nguyen, T.-v. T. 2016b. What the language you tweet says about your occupation. In *AAAI ICWSM*.

Nguyen, D.; Gravel, R.; Trieschnigg, D.; and Meder, T. 2013. " how old do you think i am?" a study of language and age in twitter. In *AAAI ICWSM*.

Pinker, S. 2007. *The stuff of thought: Language as a window into human nature*. Penguin.

Poynton, C. 1989. *Language and gender: Making the difference*. Oxford University Press, USA.

Rakesh, V.; Singh, D.; Vinzamuri, B.; and Reddy, C. K. 2014. Personalized recommendation of twitter lists using content and network information. In *ICWSM*.

Reisinger, J., and Mooney, R. J. 2010. Multi-prototype vector-space models of word meaning. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 109–117. Association for Computational Linguistics.

Sadilek, A., and Kautz, H. 2013. Modeling the impact of lifestyle on health at scale. In *ACM WSDM*, 637–646. ACM.

Silva, T. H.; de Melo, P. O.; Almeida, J.; Musolesi, M.; and Loureiro, A. 2014. You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in foursquare. *arXiv preprint arXiv:1404.1009*.

Vinciarelli, A., and Mohammadi, G. 2014. A survey of personality computing. *Affective Computing, IEEE Transactions on* 5(3):273–291.