MFQ-2016-1

by Mhd Furqan

Proceeding International Conference of Computer, Environment, Social Science, Engineering and Technology (ICEST) 2016
Medan, Indonesia, May 23-25, 2016
ISBN 979 458 877 6

SENTIMENT ANALYSIS USING CONTEXT BASED FUZZY LINGUISTIC HEDGES

Hartono[1], Dahlan Abdullah[2], Dedy Hartama[3], Roslina[4], Mhd Furqan[4], Muhammad Zarlis[5], Zakarias Situmorang[6]
Department of Computer Science, University of Sumatera Utara
e-Mail: hartonoibbi@gmail.com, dahlan.unimal@gmail.com, dedyhartama@yahoo.com, roslinanich@gmail.com, mhdfurqan@gmailcom, m.zarlis@yahoo.com, zakarias65@yahoo.com

ABSTRACT

Sentiment analysis refers to the inference of people's views, positions and attitudes in their written or spoken texts. We Present Context Based Fuzzy Linguistic Hedges, a novel approach for sentiment analysis which has proven effective both for regular texts and texts with a high degree of noise. We have proposed novel function that emulate the effect of different linguistic hedges by using fuzzy function and incorporated them in the sentiment classification task. Our paper using SentiWordNet Tool for determining the initial sentiment value.

Keywords: Sentiment Analysis, Fuzzy Linguistic Hedges, SentiWordNet Tool

1. Introduction

Sentiment analysis refers to the inference of people's views, positions and attitudes in their written or spoken texts. Before the coining of the term, the field was studied under names such as subjectivity [1], point of view [2] and opinion mining [3]. Nowadays, the field is rapidly evolving due to the rise of new platforms such as blogs, social media and user-generated reviews [4]. entiment analysis is considered a challenging natural language processing (NLP) problem [5]. We Present Context Based Fuzzy Linguistic Hedges, a novel approach for sentiment analysis which has proven effective both for regular texts and texts with a high degree of noise. We have proposed novel function that emulate the effect of different linguistic hedges by using fuzzy function and incorporated them in the sentiment classification task.

2. Sentiment Analysis

Sentiment Analysis can be considered a classification process as illustrated in Fig. 1 [6].

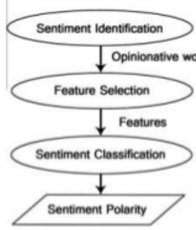

Fig 1. Sentiment Analysis Process

3. Fuzzy Linguistic Hedges

Due to its high commercial importance, mining and summarizing of user reviews are a widely studied application. The two main tasks involved in opinion mining regardless of the application are (1) identiacation of opinion-bearing phrases/sentences from free text and (2) tagging the sentiment polarity of opinionated phrases. The descriptors such as adjectives or adverbs describing the features present in an opinion sentence mainly indicate the polarity of the expressed opinion. However, the strength and polarity of the opinionated phrases are also affected by the presence of linguistic hedges such as modifiers (e.g., "not"), concentrators (e.g., "very," "extremely"), and dilators (e.g., "quite," "almost," and "nearly"). Zadeh developed the concept of fuzzy linguistic variables and linguistic hedges that modify the meaning and intensity of their operands [7, 8]. Recent papers in this aeld have also pointed out that the task of opinion mining is sensitive to such hedges and taking the effect of linguistic hedges into consideration can improve the efficiency of the sentiment classiacation task [9, 10, 11].

4. Context Based Fuzzy Linguistic Hedges

This process can be best explained using an example. Let us consider a sentence that appears in a paragraph of a transcribed call that was tagged as negative: ''I'm always confused when I get my bill. I can't find... I have problems with things there''. Every sentences's word may have a positive or negative meaning. The adjective "unpredictable" may have a negative sentiment in a car review as in "unpredictable steering", but it could have a positive sentiment in a movie review as in "unpredictable plot". So we must use contex based in determining positive or negative statement. We classify a new user review based on its fuzzy sentiment score whose computation requires three steps: (1) extract features, associated descriptors, and hedges from the review based on FOLH table lookup, (2) identify the polarity and initial value of the feature descriptors based on SentiWordNet score and (3) calculate overall sentiment score using fuzzy functions to incorporate the effect of linguistic hedges. As discussed earlier, we consider the SentiWordNet score of a feature descriptor as its initial fuzzy score . If the descriptor has a preceding hedge, its modified fuzzy score is calculated using:

$$f(\mu(s)) = 1-(1-\mu(s))^{\delta} \quad \text{.........................(1)}$$

Similar to Zadeh's proposition [10], if the hedge is a concentrator, we choose which gives us modiaed fuzzy concentrator score as indicated in (2), while if the hedge is a dilator we choose which gives us modiaed fuzzy dilator score as indicated in (3)

$$f_c(\mu(s)) = 1-(1-\mu(s))^{2} \quad \text{.........................(2)}$$

$$f_d(\mu(s)) = 1-(1-\mu(s))^{1/2} \quad \text{.......................(3)}$$

ISBN 979 458 877 6

Let and indicate the initial sentiment values of a feature descriptor which are to be modified using the proposed functions for fuzzy linguistic hedges. From Property 1 it becomes clear that both the concentrator and dilator fuzzy functions are strictly increasing in the interval. Moreover, as indicated by Property 2, the dilator function decreases the value of the input sentiment variable while the concentrator function increases its value. Property 3 indicates that even a3er applying the fuzzy functions the output value remains in the normalized range of [0,1].

Let represent the complete feature set of a product. Suppose that a user review has comments on a subset of the feature set. Further, let H represent the subset of which is preceded by concentrator or dilator linguistic hedges, while represents the subset of not preceded by these hedges.

Now, the average fuzzy sentiment score is calculated as shown in [4].

$$\beta_{\text{avg}} = \frac{\sum_{i=1}^{|H|} p_i f\left(\mu_i(s)\right) + \sum_{k=1}^{|N|} p_k \mu_k(s)}{|F|}. \quad \ldots\ldots\ldots\ldots..(4)$$

The Value of B_N Calculating using:

$$\beta_N = \frac{\beta_{\text{avg}} + 1}{2}. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(5)$$

Once the value of B_N is computed, the opinion class can be determined using the following rule set:

if $\beta_N \geq 0$ and $\beta_N \leq 0.25$, then C = "very negative," else
if $\beta_N > 0.25$ and $\beta_N < 0.5$, then C = "negative," else
if $\beta_N = 0.5$, then C = "neutral," else
if $\beta_N > 0.5$ and $\beta_N \leq 0.75$, then C = "positive," else
if $\beta_N > 0.75$ and $\beta_N \leq 1$, then C = "very positive."

5. Implementation of Context Based Fuzzy Linguistic Hedges

Example:
I'm always confused when I get my bill. I can't find... I have problems with things there

Initial sentiment value are determinng using the SentiWordNet Tool

- Confused initial sentiment value for negative is 0.00065 and for positive is 0.000041
- My Bill initial sentiment value for negative is 0.0013 and for positive is 0..00048
- Problems initial sentiment value for negative is 0.00055 and for positive is 0.00038

for Positive Statement the value of Bavg that can be counted (without contentrator or dilator):
assume that Polarity positive statement is 1 dan -1 for negative statement

For Negative Statement
Bavg = (-1 x 0.00065 + -1 x 0.0013 + -1 x 0.00055) / 3
= -0.000833

For positive statement
Bavg = (1 x 0.000041 + 1 x 0.00048 + 1 x 0.00038) / 3
= 0.00030033

Because of the value of Negative Statement is greater than positive statement, the meaning is in the negative statement.

We will give another examples that using Fuzzy Linguistic Hedge. There are a statement:
""The call quality is extremely good and navigation is comfortable but the body is somewhat fragile."

According to that statement, Table 1 show us the partial feature orientation table with linguistic hedges for smartphone products

Table 1. The partial feature orientation table with linguistic hedges for smartphone products

Feature	Descriptors with positive polarity ($P = 1$)	Descriptors with negative polarity ($P = -1$)	Linguistic hedges (if present)		
			Modifier (inverter)	Concentrator	Dilator
Call quality	Good, excellent, and satisfactory	Poor, bad	Not, never	Very, extremely	Quite, hardly
Body/design/build	Sleek, lightweight, thin, slim, beautiful, sturdy, striking, and gorgeous	Heavy, bulky, and fragile		Very, absolutely	Somewhat, quite, and almost
Screen/touchscreen/display/retina display	Nice, great, sensitive, awesome, clear, and bright	Dull, bad, and spotty	Not	Highly, incredibly	Quite
Camera/phone camera/digital camera	Awesome, good, superior, and high- resolution	Low resolution, inferior	Not	Very, positively	
User interface	Friendly, attractive, good, and lovely	Bad, poor	Not so	Highly, very	More or less
Navigation	Comfortable, intuitive, easy, and fast	Bad, difficult, slow, and jumpy		Very, significantly	Quite

In the above sentence, *[.n]* indicates noun, *[.a]* indicates adjective, and *[.v]* indicates verb. [us, *"call quality"* can be interpreted as a noun feature which is described by the adjective descriptor *"good."* Similarly, *"navigation"* and *"body"* are features described by the descriptors *"comfortable"* and "*fragile*," respectively. Moreover, the descriptor "*good*" is preceded by the concentrator hedge *"extremely"* and the descriptor "*fragile*" is preceded by the dilator hedge "*somewhat,*" while the descriptor *"comf or tabl e"* has no preceding hedge in this particular review sentence.

Let us revisit the smartphone review sentence *"The body is fragile."*

the SentiWordNet score for adjective "*fragile*" (as used to describe *body* in the smartphone review sentence) is given by the triplet (P: 0, O: 0.375, and N: 0.625) which indicates its positive, objective, and negative score. Since the negative sentiment value is highest in the triplet, "*fragile*" is assigned a polarity of "-1" that indicates negative orientation and an initial sentiment intensity value "0.625" which is used in the next phase.

ISBN 979 458 877 6

the initial sentiment score for the descriptor *"fragile"* obtained using SentiWordNet is μ(s) = 0.625. If this descriptor is preceded by a concentrator linguistic hedge, for example, *"very fragile,"* then its modified fuzzy score is obtained using (2) as *Fc*(μ(s)) = 0.893. Similarly, if this descriptor is preceded by a dilator linguistic hedge, for example, *"somewhat fragile,"* then its modiaed fuzzy score is obtained using (3) as *Fd*(μ(s)) = 0.3876.

$$\beta_{avg} = \frac{-1x0.625 + -1x0.8593}{0.8593} = \frac{-1.4843}{0.8593} = -1.66$$

$$\beta_N = \frac{-1.66+1}{2} = \text{-}0.33 = 0.33$$

Karekan β_N >0,25 dan β_N >0,3 maka termasuk ke dalam kategori *negative*

6. Discussion

Our new approach can be determining the sentiment in the sentence. For the future, we would like to build an enhanced opinion mining system that calculates the weight of an opinion bye establishing its authenticity.

6. Conclusions

The conclusion that can be drawn from this study are as follows.

1. The sentiment analysis must be adapted to the context that the sentence is using. It is because different context can make the meaning of the sentence become different.
2. The proposed functions for emulating fuzzy linguistic hedges could be successfully incorporated into the statement classification task.

References

[1] R.W. Langacker, Observations and speculations on subjectivity, *Iconicity Syntax* 1 (985) (1985) 109

[2] J. Scheibman, *Point of View and Grammar: Structural Patterns of Subjectivity in American English Conversation*, vol. 11, John Benjamins Publishing, 2002

[3] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Found. Trends Inform.* Ret. 2 (1–2) (2008) 1–135

[4] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, *in: Proceedings of the Workshop on Languages in Social Media*, Association for Computational Linguistics, 2011

[5] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, Proceedings *of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, Association for Computational Linguistics, 2002, pp. 79–86

[6] Medhat, W., A. Hassan, and H. Korashy. Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal* Vol. 5, 2014, pp. 1093-1113

[7] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning-II, *Information Sciences*, vol. 8, no. 4, part 3, 1975, pp. 301–357

[8] V. N. Huynh, T. B. Ho, and Y. Nakamori, A parametric representation of linguistic hedges in Zadeh's fuzzy logic, *International Journal of Approximate Reasoning*, vol. 30, no. 3, pp. 203–223, 2002. View at Publisher

[9] M. K. Dalal and M. A. Zaveri, Semisupervised learning based opinion summarization and classiacation for online product reviews, *Applied Computational Intelligence and Sod Computing*, vol. 2013, Article ID 910706, 8 pages, 2013.

[10] L. Dey and Sk. M. Haque, Opinion mining from noisy text data, *International Journal on Document Analysis and Recognition*, vol. 12, no. 3, pp. 205–226, 2009.

[11] S. Nadali, M. A. A. Murad, and R. A. Kadir, Sentiment classiacation of customer reviews based on fuzzy logic, *in Proceedings of the International Symposium on Information Technology (ITSim' 10)*, pp. 1037–1044, mys, June 2010.

[12] Dalal, M.K., M.A. Zaveri. Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges. *Applied Computing Intelligence and Soft Computing*. 2014.

MFQ-2016-1

ORIGINALITY REPORT

18%	9%	16%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Thien Khai Tran, Tuoi Thi Phan. "Chapter 25 Computing Sentiment Scores of Adjective Phrases for Vietnamese", Springer Science and Business Media LLC, 2016 Publication	3%
2	Mita K. Dalal, Mukesh A. Zaveri. "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews", Applied Computational Intelligence and Soft Computing, 2013 Publication	2%
3	ijcotjournal.org Internet Source	2%
4	Ernesto D'Avanzo, Giovanni Pilato. "Mining social network users opinions' to aid buyers' shopping decisions", Computers in Human Behavior, 2015 Publication	1%
5	ijarcsse.com Internet Source	1%

6	Natalia Vanetik, Marina Litvak. "DRIM: MDL-Based Approach for Fast Diverse Summarization", 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2018 Publication	1 %
7	wic.uchile.cl Internet Source	1 %
8	benjamins.com Internet Source	1 %
9	Walaa Medhat, Ahmed Hassan, Hoda Korashy. "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, 2014 Publication	1 %
10	Reshma V, Ansamma John. "Aspect based summarization of reviews using naïve Bayesian classifier and fuzzy logic", 2015 International Conference on Control Communication & Computing India (ICCC), 2015 Publication	1 %
11	Md. Asimuzzaman, Pinku Deb Nath, Farah Hossain, Asif Hossain, Rashedur M. Rahman. "Sentiment analysis of bangla microblogs using adaptive neuro fuzzy system", 2017 13th International Conference on Natural	1 %

Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2017
Publication

12	Betul Dundar, Suat Ozdemir, Diyar Akay. "Opinion mining and fuzzy quantification in hotel reviews", 2016 International Symposium on Networks, Computers and Communications (ISNCC), 2016 Publication	<1%
13	hal.univ-lorraine.fr Internet Source	<1%
14	Farman Ali, Kyung-Sup Kwak, Yong-Gi Kim. "Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification", Applied Soft Computing, 2016 Publication	<1%
15	Submitted to University of Michigan, Dearborn Student Paper	<1%
16	dspace.bracu.ac.bd:8080 Internet Source	<1%
17	D Abdullah, Tulus, S Suwilo, S Effendi, Hartono. "DEA Optimization with Neural Network in Benchmarking Process", IOP Conference Series: Materials Science and Engineering, 2018 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		