Detecting Expressions of Blame or Praise in Text

Udochukwu Orizu and Yulan He
School of Engineering and Applied Science
Aston University, UK
orizuus@aston.ac.uk, y.he@cantab.net

Abstract

The growth of social networking platforms has drawn a lot of attentions to the need for social computing. Social computing utilises human insights for computational tasks as well as design of systems that support social behaviours and interactions. One of the key aspects of social computing is the ability to attribute responsibility such as blame or praise to social events. This ability helps an intelligent entity account and understand other intelligent entities' social behaviours, and enriches both the social functionalities and cognitive aspects of intelligent agents. In this paper, we present an approach with a model for blame and praise detection in text. We build our model based on various theories of blame and include in our model features used by humans determining judgment such as moral agent causality, foreknowledge, intentionality and coercion. An annotated corpus has been created for the task of blame and praise detection from text. The experimental results show that while our model gives similar results compared to supervised classifiers on classifying text as blame, praise or others, it outperforms supervised classifiers on more finer-grained classification of determining the direction of blame and praise, i.e., self-blame, blame-others, self-praise or praise-others, despite not using labelled training data.

Keywords: blame/praise detection, text classification, social computing.

1. Introduction

With the prevalence of social networking sites, the need for social computing is constantly growing. Social computing utilises human insights for computational tasks as well as design of systems that support social behaviours and interactions (Parameswaran et al, 2007). In recent years, social computing is becoming key in many research areas and technological systems such as learning, human-computer interaction, entertainment and many more. One of the key aspects of social computing is the ability of an entity to infer the social behaviour of not just itself but of other entities as well (Parameswaran et al, 2007; Mao et al, 2011). This inference includes the ability to pass judgement and determine if an entity is blameworthy or praiseworthy and allocate blame or praise where appropriate (Tognazzini et al, 2014). Blame and praise are closely related, to blame an entity is to hold that entity morally responsible for doing something of a negative outcome while praise is to hold that entity morally responsible for doing something of a positive outcome (Eshleman, 2014) . Detection of blame/praise can be used in a variety of applications such as identifying entities holding moral responsibilities in multi-agent systems, helping with the detection of emotions such as guilt, remorse, admiration, and shame, and many more (Ortony, 1990).

In this paper, we focus on detecting blameworthiness and praiseworthiness based on the "Path Model of Blame" presented in (Malle et al, 2014). In particular, we propose an approach by adapting the original Path Model of Blame and combine natural language processing techniques for the detection of blame or praise expressed in text. To the best of our knowledge, this is the first piece of work exploring an automated approach for blame/praise detection from text. In order to evaluate our proposed approach, we have created an annotated corpus by labelling each sentence as expressing "*blame*" or "*praise*" from the ISEAR[1] (International Survey On Emotion Antecedents And Reactions) dataset. We have also provided annotations at the more finer granularity level to further distinguish the direction of blame/praise, i.e., "*self-blame*", "*blame-others*", "*self-praise*" or "*praise-others*". Our experimental results show that while our model gives similar results compared to supervised classifiers on classifying text as "*blame*", "*praise*" or "*others*", it outperforms supervised classifiers on more finer-grained classification of determining the direction of blame and praise, despite using no labelled training data.

In the rest of the paper, we first summarise theories of blame and then discuss some related work in the area of blame detection. We subsequently present our proposed approach for blame/praise detection. We explain how we create the annotated dataset and discuss experimental results. Finally, we conclude our paper and outline future directions.

2. Theories of Blame

Blame is in the family of "moral judgements". It deals with evaluating agents for their involvement in events to determine if an agent is blameworthy or praiseworthy. As discussed in (Malle et al, 2014), in the family of moral judgments one needs to distinguish at least three types :

1. Setting and attesting to norms for example avowing one norm as overriding another or stating an imperative;
2. Evaluating events, outcome of events and behaviours in relation to norms, e.g. judging an event as bad or good;
3. Performing agent evaluations to see there involvement in norm-related events, for example judging someone as blameworthy or morally responsible.

[1] `http://www.affective-sciences.org/researchmaterial`

The key difference between the three types of judgment include the following: *Type 1* is directly involved with norms, while *Types 2* and *3* are more judgments through evaluation in relation to those norms. Furthermore, *Type 2* focuses on events, while *Type 3* focuses on agents. Blame falls into the category of *Type 2* and *3* (Malle et al, 2014). Blame is cognitive in relation to the process that leads to a judgement of blame; blame is also social in relation to the act of showing a judgement of blame to different entities.

There are a variety of theories of blame, which could be organised in different dimensions depending on purpose. We consider here just two dimensions. First, we could categorise theories of blame according to the content of blaming attitudes (Malle et al, 2014). This dimension covers theories that force of blame is located in judgements of ill will and those that say blame is an emotional response to ill will (Malle et al, 2014; Tognazzini et al, 2014). The other option is to categorise theories of blame according to those psychological states or dispositions that are identified with blame. We get four category accounts of blame which include: *Cognitive* theories of blame say that blame is an evaluation or judgement an entity makes about an agent in relation to attitudes or actions; *Conative* theories of blame emphasise motivational elements, like desires and intentions, as essential to blame; *Emotional* theories sees blame as an emotional expression; and *Functional* accounts identify blame by its functional role and can be more flexible than other three categories (Tognazzini et al, 2014).

2.1. Blameworthiness

One's action is blameworthy, when they are found to be morally responsible for some wrong doing. In contrast, they are praiseworthy for doing something right (Malle et al, 2014; Tognazzini et al, 2014). The path model to blame helps answering the question: "*When is it appropriate for X to blame Y?*". The answer is that "*Only when Y deserves it*". Hence in order for an agent to be blameworthy, certain conditions must be satisfied:

- **Moral Agency**: As we stated earlier, being to blame is not adequate for being blameworthy. According to Gary Watson, earthquakes and mosquitoes can be to blame for various negative outcomes, but neither can be blameworthy because neither can react effectively and competently in moral matters (Eshleman, 2014). There is a wide acceptance that blameworthy agents must have the capacity to reason about and execute a decision, thus the agent must be a moral agent (Tognazzini et al, 2014). This mean that entities such as earthquakes and floods cannot be moral agents.

- **Freedom**: In addition to having the general capacity for practical reasoning, it is often thought that an individual is blameworthy only if on the occasion in question, exercises free will. "*I couldn't help it*" or "*I was forced to do it*" are excuses are often enough to render blame inappropriate. Free will is seen as the ability to control by process of selection which two possible futures obtain. Our vulnerability to coercion, situational pressures and manipulation which robs us of our freedom, provides us with an exemption from blame (Eshleman, 2014; Tognazzini et al, 2014).

2.2. Path Model of Blame

We propose an approach for blame detection based on the "Path Model of Blame" proposed by Malle et al. (2014), The model expresses that inside of the theoretical structure currently in place in standard social cognition gives rise to blame judgments. Blame judgments involves information which is important to other concepts and verifying the meeting of various required criteria. Blame seems to be centered around events and outcomes. According to Malle et al. (2014), the model applies equally well to both events which are time-extended processes and outcomes which the result of events.

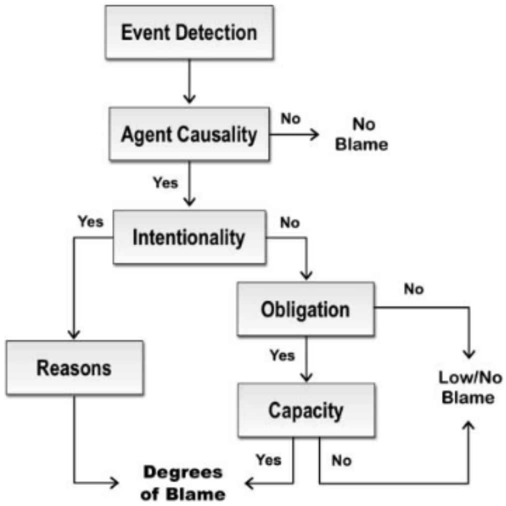


Figure 1: The Path Model of Blame (Malle et al, 2014).

Figure 1 shows the "Path Model of Blame". On the hierarchy of en route to blame, the logic has to proceed along particular paths, as represented in the Figure 1. From the structure, blame emerges if the perceivers first detect an event has violated the perceived norm (*Event Detection*); after the detection of the negative event, it has to be identified that an agent caused the event (*Agent Causality*). If no (moral) agent has caused the negative event, blame cannot be established. The Path Model states that the causal involvement of an agent falls into two categories, either intentional or unintentional. On the intentional path, if the negative event in question is evaluated and found to be intentional, the perceiver must now consider reasons for this action (*Reasons*). Blame is present, but the degree of blame is dependent on the reasons. If the agent is found to have caused the event unintentionally, the perceiver considers the degree of *obligation* and *capacity* (includes capacity to foresee or foreknowledge of the event) of the agent had to prevent the negative event. According to the Path Model, it is only when an agent is found to have both the obligation and capacity will the agent be blamed for the negative actions.

Adding the capacity for practical reasoning to the power of free will, one ends up with a morally responsible agent.

There are some subtleties here which are ignored in this paper.

3. Related Work

We did not find much research in the area of blame/praise detection in text. We however present here work which is close to what we are doing. Attribution research looks at how one makes sense of the world by attributing behaviour and events to their causes. It is basically ascribing a cause to an event as well as the judgements made (Lagnado and Channon, 2008). There are other works not directly related to text processing but useful in understanding computational models for blame/praise detection in text. In Mao et al. (2011), they adopted the shaver model terminology and represented causal knowledge in hierarchies which allow a conscience description of the causal relationship between events and states. An example was presented by using the online text data crawled from 25,103 web pages from news outlets related to Al-Qaeda. A set of manually defined linguistic patterns and rules was used to extract actions and the action preconditions and effects which were then used to represent causal knowledge.
Although the research presented in (Mao et al, 2011) shed a light in analysing the casual relationship between events and states, their reliance of manually defined linguistic patterns for identification and extraction of actions and action precondition and effects limits the scope of study since it involved heavy manual effort. Their approach is also domain dependent and cannot be generalised to other application areas.

4. Our Approach

The detection of blame/praise from text can be casted as a classification problem, where given a sample text, we aim to learn a model which is able to classify the text as expressing "*blame*", "*praise*" or "*others*". We also explore a finer-grained distinction of the direction of blame and praise, i.e., "*self-blame*", "*blame others*", "*self-praise*" and "*praise others*".
We started to explore the use of the Path Model of Blame (Malle et al, 2014) for the detection of blame/praise from text. As we are not concerned with the identification of the degree of blame but the existence of blame, there is no need to determine "*Reasons*" as in the original Path Model of Blame. Also, instead of identifying "*Intentionality*", "*Capacity*" and "*Obligation*", we replace them with "*Foreseeability*" and "*Coercion*". According to the path model in Figure 1, "*Capacity*" deals with the ability of the moral agent to have known about the actions and its effects before hand, in other words its foreseeability. Foreseeability refers to an agent's foreknowledge about actions and their effects. Clearly we can say that intentionality entails foreknowledge (Mao et al, 2011). Various other papers in this area (Mao et al, 2011; Guo et al, 2009; Mao et al, 2005; Shaver, 1985) state that there is a close inter-play between intentionality and foreseeability. As such, we replace "*Intentionality*" and "*Capacity*" with "*Foreseeability*". According to the Webster dictionary to *coerce* is "to make (someone) do something by using force or threats". In Figure 1, obligation deals with the extent to which a moral agent had ability to prevent a negative event. In this case, the perceiver is considering "could the agent have been forced to carry out an action" or "was the agent tricked into carrying out an action". Thus, coercion covers not only cases where the agent was forced but also where the agent was tricked to execute an action. Typically coercion is thought to carry with it the implication to diminish the targeted agent's freedom and responsibility (Anderson, 2014). We therefore replace "*Obligation*" with "*Coercion*". The revised Path Model of Blame is illustrated in Figure 2.

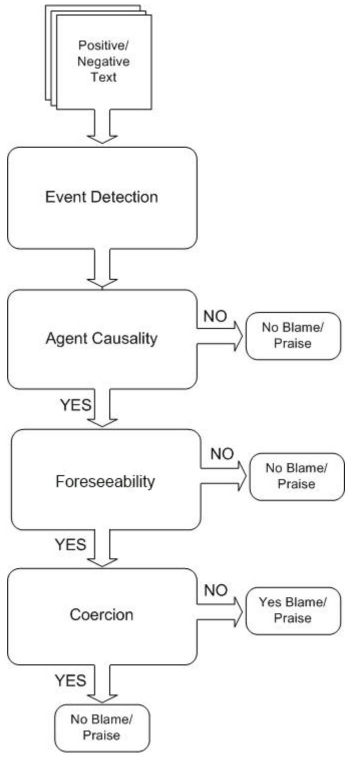


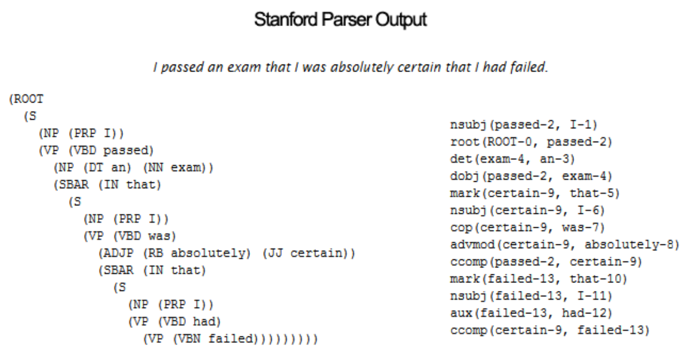
Figure 2: Revised Path Model of Blame.

It is generally acknowledged that potentially blameworthy entity must be capable of reflecting upon, reasoning about, and executing a decision. On the off chance that an entity does not have these requirements, it is exempted from blame (Tognazzini et al, 2014). Thus we use named entity recognition focusing on entities of persons, organisation and country. We also identify the use of pronouns representing persons based on the Part-of-Speech (POS) tagging results.
We first pre-process text by carrying out sentence splitting and tokenisation[2], POS tagging[3], named entity recognition (NER)[4], word sense disambiguation (WSD)[5], dependency parsing[6], and polarity detection using majority voting based on the lexicon matching results obtained with three senti-

[2] `http://nlp.stanford.edu/software/tokenizer.shtml`
[3] `http://nlp.stanford.edu/software/tagger.shtml`
[4] `http://nlp.stanford.edu/software/CRF-NER.shtml`
[5] `http://www.nltk.org/`
[6] `http://nlp.stanford.edu/software/stanford-dependencies.shtml`

ment lexicons, SentiWordNet (Esuli et al, 2005), AFINN (Hansen et al, 2011) and the Subjectivity Lexicon (Wilson et al, 2005). Negation is also considered during polarity detection.

In order to use the revised Path Model of Blame for the detection of blame/praise from text, we need to first detect events and then determine "*Agent Causality*", "*Foreseeability*" and "*Coercion*". In the following, we describe how each of the steps can be performed.


Stanford Parser Output

I passed an exam that I was absolutely certain that I had failed.

```
(ROOT
  (S
    (NP (PRP I))
    (VP (VBD passed)
      (NP (DT an) (NN exam))
      (SBAR (IN that)
        (S
          (NP (PRP I))
          (VP (VBD was)
            (ADJP (RB absolutely) (JJ certain))
            (SBAR (IN that)
              (S
                (NP (PRP I))
                (VP (VBD had)
                  (VP (VBN failed))))))))))
```

```
nsubj(passed-2, I-1)
root(ROOT-0, passed-2)
det(exam-4, an-3)
dobj(passed-2, exam-4)
mark(certain-9, that-5)
nsubj(certain-9, I-6)
cop(certain-9, was-7)
advmod(certain-9, absolutely-8)
ccomp(passed-2, certain-9)
mark(failed-13, that-10)
nsubj(failed-13, I-11)
aux(failed-13, had-12)
ccomp(certain-9, failed-13)
```

Figure 3: An example dependency parse result.

- **Event Detection**. We look at the "`verb+object`" combination as identified using the Stanford dependency parser and take note of the agent of the verb. We use the majority voting mechanism mentioned above for polarity detection. Negatively or positively valenced events are extracted from sentences expressing negative or positive polarity respectively.

  In the example shown in Figure 3, we see that the event detected by the "`verb+object`" pattern is "passed exam". And the agent of the verb "passed" here is "I". We then detect the polarity of the event by searching for positive or negative words modifying the event taking into account of negation. In this example, the verb "passed" carries a positive polarity. As such, the event is considered as a positively valenced event.

- **Agent Causality**. Here, one must establish that a moral agent caused an event. We first make use of a popular explicit intra-sentential pattern for causation expression which is "`NP verb NP`" where NP is a noun phrase (Girju, 2003) and then we identify the agent within the noun phrase. If the intra-sentential pattern is not found we consider verbs in the text that belong to the CAUSE class and the CAUSE-TO semantic relation which are defined in the WordNet. In order for "Agent Causality" taking the value "True", the agent must be a person entity (including pronouns).

  In the example shown in Figure 3, we see that the intra-sentential pattern "`NP verb NP`" is present and the dependency parse result shows that verb "passed" is associated with the subject "I" (first person pronoun). This tells us that the agent is a moral agent within the context of the sentence.

  For all the self categories ("*self-blame*" or "*self-praise*"), the agent must be a first person pronoun. For other categories ("*blame others*" or "*praise others*"), the agent must not be a first person pronoun, but must be one of the following: a pronoun, a person, country or organisation as identified using the NER tool.

- **Foreseeability**. We rely on a set of verbs which indicate foreseeability. These include verbs of communication as suggested in (Mao et al, 2011) and other verb classes which include verbs of creation, verbs of consumption, verbs of competition, verbs of possession and verbs of motion. These classes of verbs are defined in the WordNet[7] and can be identified by looking at the WordNet sensekey of the verbs.

  Example: *When I did not* ***speak*** *the truth.*

  In the example above, the communication verb "speak" indicates that the subject "I" had foreknowledge of the event of "speaking the truth".

- **Coercion**. To identify coercion, we look at the extension verb classes presented in (Kipper et al, 2006) focusing on verbs in the URGE (13 members), FORCE (46 members) and FORBID (17 members) classes.

  Example: *I was* ***forced*** *to quite the job in the city.*

  In the example above , using word sense disambiguation, the verb "forced" is of sense "*to cause to do through pressure or necessity, by physical, moral or intellectual means*". The agent "I" in this case did not willingly quite the job and the sentence does not mention who forced the agent. Thus, the sentence is classified as "*Others*" (i.e., no blame or praise).

5. Corpus Creation

We created our data from the the ISEAR dataset, which was collected during the 1990s by a large group of psychologists by asking nearly 3,000 participants from different cultural background about their emotional experiences. This dataset contains 7,660 comments, each of which is labelled with one of the seven emotions (*joy, fear, anger, sadness, disgust, shame* and *guilt*).

We asked two English-speaking individuals to annotate each comment in the ISEAR dataset as "*blame*", "*praise*" or "*others*". For comments expressing *blame* or *praise*, the annotators further labelled them as "*self-blame*", "*blame others*", and "*self-praise*" and "*praise others*". The annotators were provided with the annotation guidelines and sample annotation results. A web-based interface has been developed to ease the task of annotation. We did not provide them with the ISEAR emotion labels of the comment as we believe that the emotion label information, although might be helpful, will create bias and influence the annotators.

The inter-annotator agreement for our data set is shown in Table 5.. There are several agreement measures which have been proposed in the literature (Artstein et al, 2008). We measure the reliability of the annotation results by using the kappa (k) coefficient (Cohen, 1960), which is defined as $k = A_o - A_e/1 - A_e$ where A_o is the observed agreement, and A_e is the expected agreement by chance. We obtained

[7] https://wordnet.princeton.edu/man/lexnames.5WN.html

a k score of 0.62. Using the scales for interpreting Kappa provided in (Landis et al, 1977) and (Green, 1997) in terms of strength of agreement, our score can be interpreted as a good and substantial agreement.

Annotator 2 \ Annotator 1	Blame	Praise	Others	Total
Blame	**3483**	222	279	3984
Praise	227	**778**	305	1310
Others	348	299	**1719**	2366
Total	4058	1299	2303	7660

Table 1: Annotation agreement matrix.

It can be observed from Table 5. that, 45% of the comments in ISEAR are labeled as "*blame*" and 10% as "*praise*". We only keep the comments where both annotators reach an agreement. On fine grained labelling, we had a discrepancy of about 17%. We then got both annotators to re-examine these 17% comments to reach an agreement. Our final dataset consists of 57.1% self blames and 42.9% blames directed towards others within the blame context; and 66.2% self praises and 33.8% praises directed towards others in the praise context.

6. Experiments

In this section, we present the evaluation results of our blame/praise detection approach and compare it with supervised learning approaches trained on the "bag-of-words" features. Experiments for the supervised classifiers were carried out using Weka[8] with documents pre-processed with stopword removal, POS tagging and extraction of n-grams up to trigrams. We report the results using 10-fold cross validation. Note that such a comparison is not fair since our approach does not make use of any labelled data.

Class	NB			SVM			Our Approach		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
blame	**0.74**	0.60	0.66	0.73	**0.80**	**0.76**	0.73	0.76	0.75
praise	0.40	0.38	0.39	**0.60**	0.44	0.51	0.43	**0.63**	**0.52**
others	0.47	**0.66**	0.55	0.61	0.57	**0.59**	**0.63**	0.53	0.58

Table 2: Classification results of *blame*, *praise* or *others*.

It can be observed from Table 2 that supervised SVM performed the best in classifying *blame*, but only slightly outperforms our approach by about 1% in F-measure. Supervised NB gives much worse results with F-measure lower than that of SVM. Our approach achieves similar performance as SVM on the *praise category* and outperforms NB by 13% in F-measure. The dataset has a higher number of negative comments and this is reflected in the results obtained on the *blame* category having better performance than those from *praise* across all classifiers.

For fine-grained classification, it can be observed from Table 3 that SVM performed better than NB on all categories. However, our approach performed better than both SVM and NB. In classifying into *self-blame* (blame directed towards oneself) and *blame others* (blame directed towards other people), our approach performs better than SVM and NB with an average F-measure difference of about 6% and 12% respectively. In the *self-praise* and *praise others* categories, our approach performs better than both SVM and NB with an average F-measure difference of about 20% and 17%.

Class	NB			SVM			Our Approach		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
self-blame	0.54	0.40	0.46	0.52	0.58	0.55	**0.58**	**0.59**	**0.59**
blame others	0.43	0.49	0.46	0.50	0.47	0.48	**0.53**	**0.68**	**0.57**
self-praise	0.44	0.42	0.43	**0.54**	0.38	0.45	0.49	**0.51**	**0.50**
praise others	0.16	0.33	0.25	0.22	0.12	0.16	**0.49**	**0.53**	**0.51**
others	0.51	0.52	0.51	0.56	**0.62**	**0.59**	**0.63**	0.53	0.58

Table 3: Classification results of *self-blame*, *self-praise*, *blame-others*, *praise-others* or *others*.

ISEAR dataset contains personal experience expressed by a wide range of participants and hence might contain lots of informal and ill-grammatical text. Our experimental results show that our approach performs reasonably well on such a dataset. Our approach relies on results generated from a series of NLP tasks such as POS tagging, word-sense disambiguation, dependency parsing and polarity detection in order to be able to assign values to a set of variables for blame/praise detection. Thus, any error that occurs will be propagated down the pipeline process.

One main reason for the poor performance of the supervised approaches in fine-grained classification can be attributed to their inability of distinguishing between various types of pronouns. All approaches struggled with sentences which had no "`object`" or "`intransitive sentences`". For our approach this made it difficult for event detection which relies on the use of the "`verb+object`" pattern. Furthermore, failure in detecting the polarity of text will make it impossible for our approach to identify the underlying blame/praise category. However, our approach performs reasonably well especially on fine-grained classification of detecting the directions of blames or praises. This is very useful in not only identifying the entity responsible of blame or praise but also for inferring emotions such as guilt, remorse, anger and many more. For example, for the sentence "*When I caused problems for somebody because he could not keep the appointed time and this led to various consequences.*", the Agent ("I") is blameworthy and hence we can infer that the sentence expresses an emotion of *guilt*.

7. Conclusions

In this paper, we have proposed a rule-based approach built upon the Path Model of Blames for detecting expressions of blame and praise in text. Experimental results on our dataset show that our approach gives similar performance compared to supervised classifiers when classifying text as *blame* or *praise*. For fine-grained classification of identifying the direction of blame and praise, our approach outperforms the supervised methods by a large margin of 14% in F-measure compared to NB and 13% in F-measure compared to SVM.

In future, we will test our approach with informal short text such as tweets and social media posts by considering hashtags, emojis and the language specifically in such environments. We will also consider using our approach to

[8] http://www.cs.waikato.ac.nz/ml/weka/

bootstrap more training examples to iteratively improve the performance of supervised classifiers for blame/praise detection.

8. Bibliographical References

Anderson, S.: Coercion. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Spring 2014 edn. (2014)

Artstein Ron, Poesio Massimo: Inter-coder agreement for computational linguistics. In: Computational Linguistics. pp. 555–596. (2008)

Brand, A.: Social cognition, emotions, and the psychology of writing. Journal of Advanced Composition pp. 395–407 (1991)

Brand, A.G.: Hot cognition: Emotions and writing behavior. Journal of Advanced Composition pp. 5–15 (1985)

Coughlan, S.: Spelling mistakes 'cost millions' in lost online sales (2011 (accessed February 21, 2014)), `http://www.bbc.co.uk/news/education-14130854`

Cohen Jacob: A coefficient of agreement for nominal scales. In: Educational and psychological measurement. pp. 37–46. (1960)

Esuli, A., Sebastiani, F.: Determining the semantic orientation of terms through gloss classification. In: CIKM. pp. 617–624 (2005)

Eshleman, A.: Moral responsibility. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Summer 2014 edn. (2014)

Green Annette M: Kappa statistics for multiple raters using categorical classifications. In: Proceedings of the 22nd annual SAS User Group International conference. pp. 4. (1997)

Girju Roxana: Automatic detection of causal relations for question answering. In: Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering. pp.76–83. (2003)

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter 11(1), 10–18 (2009)

Hansen, L.K., Arvidsson, A., Nielsen, F.Å., Colleoni, E., Etter, M.: Good friends, bad news-affect and virality in twitter. In: Future information technology, pp. 34–43. Springer (2011)

Indurkhya, N., Damerau, F.J.: Handbook of natural language processing, vol. 2. CRC Press (2010)

Klavans, J., Kan, M.Y.: Role of verbs in document analysis. In: Proceedings of the 17th international conference on Computational linguistics-Volume 1. pp. 680–686. Association for Computational Linguistics (1998)

Kipper, K., Korhonen, A., Ryant, N., Palmer, M.: Extending verbnet with novel verb classes. In: Proceedings of LREC. vol. 2006, p. 1. Citeseer (2006)

Lagnado, D.A., Channon, S.: Judgments of cause and blame: The effects of intentionality and foreseeability. Cognition 108(3), 754–770 (2008)

Landis J Richard , Koch Gary G: The measurement of observer agreement for categorical data. In: Biometrics. pp. 159–174. (1977)

Mao, W., Ge, A., Li, X.: From causal scenarios to social causality: An attributional approach. IEEE Intelligent Systems 26(6), 0048–57 (2011)

Malle, B.F., Guglielmo, S., Monroe, A.E.: A theory of blame In:Psychological Inquiry. pp.147–186. (2014)

Ortony, A.: The cognitive structure of emotions. Cambridge university press (1990)

Parameswaran, M., Whinston, A.B.: Social causality and responsibility: Modeling and evaluation. Intelligent Virtual Agents, 191–204 (2005)

Parameswaran, M., Whinston, A.B.: Social computing: An overview. Communications of the Association for Information Systems 19(1), 37 (2007)

Picard, R.W.: Affective computing (2000)

Shaver, K.: The attribution of blame: Causality, responsibility, and blameworthiness. Springer. (1985)

Tognazzini, N., Coates, D.J.: Blame. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Summer 2014 edn. (2014)

Wilson T., Wiebe J., Hoffmann P: Recognizing contextual polarity in phrase-level sentiment analysis. In: HLT-EMNLP. pp. 347–354. (2005)

Yi Guo,Nan H.,Zhiqing S.: An Algorithm for Detection of Cognitive Intentionality in Text Analysis. Information Technology and Computer Science 108(2), 176–179 (2009)