A Personalized Affect Response Model for Online News Articles

Kyohei Atarashi^{1*}, Akimi Moriyama^{1*}, Satoshi Oyama^{1,2} and Masahito Kurihara¹

¹Hokkaido University

²RIKEN AIP

{atarashi_k, a.moriyama_ast}@complex.ist.hokudai.ac.jp, {oyama, kurihara}@ist.hokudai.ac.jp

Abstract

With the spread of smartphones and the diversification of Web services, individuals, companies, and organizations can send and receive various types of information to and from anywhere. Although information is provided by many types of media, articles and posts are the most common type: news articles, blog posts, social media posts, etc. Such articles and posts can create unexpected emotional responses in readers and can occur flame war in the worst case. To avoid the risk of occurrence of flame war and to inform the recipient as intended, methods for personalized affect analysis have attracted attention. We present a model based on latent Dirichlet allocation for performing personalized affect analysis of news articles. Each article is assumed to have a distribution of topics, and each reader is assumed to have latent features that represent the strength of the effect of each topic on the reader. A reader responds to an article on the basis of the distribution of the topics it contains and the reader's latent features. Furthermore, the model leverages the articles to which no readers respond for training. The effectiveness of the proposed model was demonstrated using data on the responses, as collected through crowdsourcing, of readers to several online news articles.

1 Introduction

With the spread of smartphones and the diversification of Web services, individuals, companies, and organizations can send and receive various types of information to and from anywhere. Although information is provided by many types of media, articles and posts are the most common type: news articles, blog posts, social media posts, etc. Such articles and posts can create unexpected emotional responses in readers and can occur *flame war* in the worst case.

To avoid the risk of the occurrence of flame war and to inform the recipient as intended, methods for personalized affect analysis, which predict the responses of a reader to an article or post, have attracted attention. By modeling the personal affect responses of a reader, one can recommend articles that are beneficial for that reader and avoid recommending ones that might make the reader feel bad and avoid flame war by predicting the ratio of people who will feel bad after reading the article. Although there are several methods for performing affect analysis, many of them predict *only one affect* for a given article or word [Maas *et al.*, 2011; Ptaszynski *et al.*, 2013; Liu *et al.*, 2003]. However, different readers may have different responses to an article and hence an article that makes one reader feel bad may be beneficial to another reader.

We have developed a probabilistic generative model for personalized affect analysis that combines the latent features (factors) of news articles and their readers. It is an extension of the latent Dirichlet allocation (LDA) model [Blei et al., 2003]. Each article and its readers have their own latent features. An article's latent features can be interpreted as the topics it includes. A reader's latent features can be regarded as parameters representing the strength of each topic's effect on the reader. Unlike other personal response models [Dawid and Skene, 1979; Kajino et al., 2012; Koren et al., 2009; Duan et al., 2014], our model uses the articles to which no readers respond for training (i.e., for inferring the posterior distributions) since the proposed model includes the generating process of articles. The model's effectiveness was demonstrated on the task of predicting the affect responses, which were collected by crowdsourcing, of readers of Japanese news articles.

We next introduce the notation used. In Section 3, we review the LDA model. In Section 4, we present our proposed model for personalized affect analysis. We discuss related work and the differences between the proposed model and previous models in Section 5. We present and discuss our experimental results in Section 6 and conclude in Section 7 with a brief summary.

2 Notation

We denote the set of articles as D. We use N_d for the number of words in the *d*-th article. We use $w_{d,n}$ for the *n*-th word in the *d*-th article. We denote the set of readers as H and use E for the set of affect responses. We represent the affect responses of the *h*-th reader to the *d*-th article as $l_{h,d} \in \{0,1\}^{|E|}$. For $e \in |E|$, $l_{h,d,e} = 1$ means that

^{*}Alphabetic order (equal contribution).

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the *h*-th reader feels the *e*-th affect for the *d*-th article, and $l_{h,d,e} = 0$ means that the *h*-th reader does not. Note that we assume that readers can feel multiple affects for an article; i.e., $\sum_{e=1}^{|E|} l_{h,d,e}$ can be higher than 1. Our goal is to develop a model for personalized affect analysis, i.e., a model for predicting responses $l_{h,d,e}$ that are not observed (unknown). Therefore, in our scenario, all readers do not read all the articles. We represent the set of indices of readers that read and respond to the *d*-th article by H_d and the set of observations of responses $\{l_{h,d,e} \mid d \in [D], h \in H_d, e \in [E]\}$ by *L*.

3 Latent Dirichlet Allocation

The LDA model [Blei *et al.*, 2003] is a statistical model for analyzing documents. Each document is assumed to have a distribution of *topics*; i.e., each document can be represented as a mixture of topics, and each topic has a distribution of words. LDA is based on the assumption that each topic has a word distribution and that the words in a document are generated on the basis of this distribution and on the mixture-ratio of topics in the document. The joint distribution of the LDA is given by

$$p(D, \{\boldsymbol{\phi}_k\}, \{z_{n,d}\}, \{\boldsymbol{\theta}_d\}; \boldsymbol{\alpha}, \boldsymbol{\beta})$$

$$:= \prod_{k=1}^{K} p(\boldsymbol{\theta}_k; \boldsymbol{\beta}) \prod_{d=1}^{|D|} [p(\boldsymbol{\theta}_d; \boldsymbol{\alpha})$$

$$\times \prod_{n=1}^{N_d} p(z_{d,n}; \boldsymbol{\theta}_d) p(w_{d,n}; \phi_{z_{d,n}})],$$
(1)

where K is the number of topics, which is specified by the user, α and β are the parameters of Dirichlet distributions $p(\phi_k; \beta)$ and $p(\theta_d; \alpha)$, and $\{\phi_k\}, \{z_{n,d}\}, \text{ and } \{\theta_d\}$ are latent variables. The process for generating D is as follows:

- 1. For k = 1, ..., K
 - (a) Generate word distribution for k-th topic $\phi_k \sim \text{Dir}(\phi_k; \beta)$.
- 2. For d = 1, ..., |D|
 - (a) Generate topic distribution for *d*-th document $\theta_d \sim \text{Dir}(\theta_d; \alpha)$.
 - (b) For $n = 1, ..., N_d$
 - i. Generate topic $z_{d,n} \sim \operatorname{Cat}(z_{d,n}; \boldsymbol{\theta}_d)$.
 - ii. Generate *n*-th word $w_{d,n} \sim \operatorname{Cat}(w_{d,n}; \phi_{z_{d,n}})$.

where $Dir(\cdot; \cdot)$ is the Dirichlet distribution and $Cat(\cdot; \cdot)$ is the categorical distribution. Figure 1 shows a graphical representation of the LDA model.

In many cases, given *D*, the objective of using the LDA model is to infer the topic posterior distribution of a document and the word distribution of each topic. There are many methods for inferring them, e.g., variational Bayes (VB) [Blei *et al.*, 2003] and Markov chain Monte Carlo (MCMC) [Porteous *et al.*, 2008].

4 Proposed Model

Our proposed model for personalized affect analysis of news articles is an extension of the LDA model. It introduces the



Figure 1: Graphical representation of LDA model.

latent features of readers and readers affect responses on the basis of those features and the topic distributions of the articles.

4.1 Generating Process of Documents and Affect Responses in Proposed Model

We first explain our assumptions for personalized affect analysis. Because the LDA model is simply a basic model for analyzing documents, we extended it for personalized affect analysis. Since the *e*-th affect response of reader *h* to the *d*-th news article $l_{h,d,e}$ is in {0, 1}, we assume that the distribution of $l_{h,d,e}$ is a Bernoulli distribution: Ber $(l_{h,d,e}; p_{h,d,e})$, where $p_{h,d,e}$ is the probability of $l_{h,d,e} = 1$. Because $p_{h,d,e}$ clearly depends on the reader and the article, and each article has a topic distribution θ_d in the LDA model, we assume that the (h-th) reader has his or her own latent features for the *e*-th affect response: $\psi_{h,e} \in \mathbb{R}^K$, which represents the characteristics of the reader. We also assume that $l_{h,d,e}$ is generated from θ_d and the characteristics: $p_{h,d,e} = \langle \theta_d, \psi_{h,e} \rangle$. Because $p_{h,d,e}$ must be greater than 0 and less than 1, we assume that $\psi_{h,e,k} \sim \text{Beta}(\phi_{h,e,k}; \gamma_k)$. Then, $p_{h,d,e} = \langle \theta_d, \psi_{h,e} \rangle$ clearly satisfies the constraint: greater than 0 and less than 1.

The joint distribution of the proposed model, which reflects these assumptions, is given by

$$p(D, L, \{\boldsymbol{\phi}_k\}, \{z_{n,d}\}, \{\boldsymbol{\theta}_d\}, \{\boldsymbol{\psi}_{h,e}\}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \{\gamma_k\})$$

$$:= \prod_{h=1}^{|H|} \prod_{e=1}^{|E|} \prod_{k=1}^{K} p(\psi_{h,e,k}; \gamma) \prod_{k=1}^{K} p(\boldsymbol{\theta}_k; \boldsymbol{\beta}) \prod_{d=1}^{|D|} [p(\boldsymbol{\theta}_d; \boldsymbol{\alpha})$$

$$\times \prod_{n=1}^{N_d} p(z_{d,n}; \boldsymbol{\theta}_d) p(w_{d,n}; \phi_{z_{d,n}})$$

$$\times \prod_{h \in H_d} \prod_{e=1}^{|E|} p(l_{h,d,e}; \langle \boldsymbol{\theta}_d, \boldsymbol{\psi}_{h,e} \rangle)].$$
(2)

The process for generating D and $\{l_{h,d,e}\}$ is as follows:

- 1. For k = 1, ..., K
 - (a) Generate word distribution for k-th topic $\phi_k \sim \text{Dir}(\phi_k; \beta)$.
 - (b) For h = 1, ..., |H|i. For e = 1, ..., |E|

. .



Figure 2: Graphical model representation of proposed model.

- A. Generate parameter for *h*-th reader for *e*-th affect : $\psi_{h,e,k} \sim \text{Beta}(\psi_{h,e,k}; \gamma_k)$.
- 2. For $d = 1, \ldots, |D|$
 - (a) Generate topic distribution for *d*-th article $\theta_d \sim \text{Dir}(\theta_d; \alpha)$.
 - (b) For $n = 1, ..., N_d$
 - i. Generate topic $z_{d,n} \sim \operatorname{Cat}(z_{d,n}; \boldsymbol{\theta}_d)$.
 - ii. Generate *n*-th word $w_{d,n} \sim \operatorname{Cat}(w_{d,n}; .\phi_{z_{d,n}})$.
 - (c) For $h \in H_d$
 - i. For e = 1, ..., |E|
 - A. Generate *e*-th affect response of *h*-th reader to *d*-th article: $l_{h,d,e} \sim \text{Ber}(l_{h,d,e}; \langle \boldsymbol{\theta}_d, \boldsymbol{\psi}_{h,e} \rangle).$

Figure 2 shows a graphical model of the proposed model.

Our model can be interpreted as an extension of the LDA model: the distribution of reader responses $p(l_{h,d,e})$ is introduced naturally. As described in the next section, in many of the existing models for personal responses, articles to which no readers responded, i.e., *unresponded-to articles*, cannot be used for learning (i.e., inferring posterior distributions) because they model only the generation of the distribution of reader responses, $p(l_{h,d,e})$. On the other hand, our model can leverage these unresponded-to articles for learning since it includes the generating process of articles.

4.2 Inference of Posterior Distributions

Given articles D and responses L, the goal of a user using the proposed model is to obtain he posterior distributions of topics $p(\theta_d \mid D, L)$, words $p(\phi_k \mid D, L)$, and reader parameters $p(\psi_{h,e} \mid D, L)$ for all $d \in [|D|]$, $h \in [|H|]$, $e \in [|E|]$, and $k \in [K]$. After inferring the posterior distributions, the proposed model can predict unknown responses $l_{h,d,e}$ that are not included in L; i.e., $l_{h,d,e} \in L_u = \{l_{h,d,e} \mid d \in [|D|]\}, e \in [|E|], h \in [|H|] \setminus H_d\}$. Since the proposed model is an extension of the LDA model, it can infer the posterior distributions by using a VB or MCMC method in a manner similar to that of the LDA model. Generally, VB-based methods are faster than MCMC-based ones but require model-specific derivation and implementation of the algorithm, which are more complex than those of MCMC-based methods. In our experiments, we used the automatic differentiation variational inference (ADVI) [Kucukelbir et al., 2017] method for inferring the posterior distributions. The ADVI method enables the inference of approximated (variational) posterior distributions without model-specific complex derivation and implementation of an algorithm. The approximated posterior distribution is represented as a Gaussian distribution with conversion of some parameters and is optimized by maximizing the evidence lower bound using stochastic gradient ascent with the reparameterization trick [Kingma and Welling, 2014]. After the approximated posterior distributions have been inferred, the user can easily compute the distribution of topics of unknown articles.

5 Related Work

Various methods have been reported for affect analysis. Some have been rule based. Liu et al. presented an affect analysis model based on the Open Mind Common Sense database, which is a generic common sense database [Liu et al., 2003]. Ptaszynski et al. presented an affect analysis system for Japanese narratives based on ML-Ask, which is an automatic affect annotation tool using the Emotive Expression Dictionary [Ptaszynski et al., 2013]. Some have been machinelearning based. Tokuhisa et al. presented one that uses an automatically obtained labeled dataset [Tokuhisa et al., 2008]. Maas et al. proposed a method for learning word representations (word latent features) and a classifier for the affect analysis of documents. These methods are not for personalized affect analysis. Our proposed method does not introduce latent features of words and assumes that the affect responses of readers depends on only the latent features of documents and each reader. Hence, our future work includes the extension of the proposed model by introducing the latent features of words, like [Maas et al., 2011].

Probabilistic models that have been proposed in the crowdsourcing field can be used to predict personal responses [Dawid and Skene, 1979; Kajino et al., 2012; Kajino et al., 2013; Welinder et al., 2010; Duan et al., 2014; Whitehill et al., 2009]. Crowdsourcing services have been used to collect labeled data for supervised learning: a machinelearning user first collects unlabeled data and then asks crowdsourcing workers to label the data. Because crowdsourcing workers are not professional, the quality of their work is generally low. Hence, many methods have been developed for inferring true labels and worker ability from a set of noisy item labels given by multiple workers [Dawid and Skene, 1979; Kajino et al., 2012; Kajino et al., 2013; Welinder et al., 2010; Duan et al., 2014; Whitehill et al., 2009]. In addition to inferring unknown item true labels, these methods can also be used to predict personal responses.

The model presented by Dawid and Skene for aggregating diagnoses from multiple doctors [Dawid and Skene, 1979] has also been used for inferring true labels from a set of noisy labels given by crowdsourcing workers. We call it the DS model. It is based on the assumption that each doctor has his or her own confusion matrix for diagnosis and that their diagnoses (responses) depend on his or her confusion matrix and the unknown true diseases of patients. In the task of aggregating noisy crowdsourced responses, doctors, patients, and diagnoses correspond to workers, items (data), and responses, respectively. There are three main differences between the proposed model and the DS model. (i) The DS model is based on the assumption that each item (document) has one true label (true affect response) while the proposed model does not (it is not reasonable to assume the existence of a true affect response). (ii) The DS model does not consider the characteristics of each item because responses depend on only the true labels and worker characteristics. (iii) The proposed model introduces the generation of articles (items) while the DS model does not, so the proposed model can leverage unresponded-to articles for learning while the DS model can neither leverage nor predict the responses of workers to unresponded-to items.

Kajino et al. proposed the personal classifier (PC) model for learning a classifier directly from noisy crowdsourced labels [Kajino et al., 2012]. The PC model is based on the assumption that (i) each worker has his or her own classifier, (ii) given an item, each worker inputs the feature vector of the item to his or her own personal classifier and labels the item in accordance with the output of the classifier, and (iii) there is a base classier and the parameters of the personal classifiers are noisy versions of the base classifier (noises represent the worker characteristics). The advantages of the PC model are that (i) it produces not only (inferred) true labels but also a classifier, (ii) the optimization problem is convex and thus easy to solve, and (iii) it can predict the responses of workers to unresponded-to items, unlike the DS model. However, unlike the proposed model, the PC model cannot use unresponded-to items for learning.

Methods for modeling personal responses have also been presented for recommender systems [Koren *et al.*, 2009]. Matrix factorization (MF) [Koren *et al.*, 2009] is a commonly used method for recommender systems. It is based on the assumption that each item and the user has its own latent features (multi-dimensional vector) and models the responses of the *h*-th user to the *d*-th item as the dot product of the latent features of the user and those of the item. Random variables θ_d and ψ_h correspond to the latent features of the item and user, respectively. A similar model was proposed in the crowdsourcing area [Welinder *et al.*, 2010]. Unlike the proposed model, the MF method cannot leverage and predict responses of workers to unresponded-to items.

Duan *et al.* presented probabilistic models for estimating multiple labels for emotions generated by crowdsourcing workers [Duan *et al.*, 2014]. Their models are extensions of the DS model. Because their models, like the DS model, do not use item (document) information (i.e., unlike the proposed model), they cannot leverage and predict responses of workers to unresponded-to items.

6 Evaluation

We experimentally evaluated the effectiveness of our proposed model for predicting personal affect responses.

anger	sadness	joy	displeasure	surprise	fear
1,738	1,742	1,861	1,248	1,067	1,836

6.1 Dataset

We used 770 articles taken from the livedoor news corpus dataset¹. To collect reader responses to these articles, we used the Lancers Crowdsourcing Service². There were 95 readers, and each one responded to at least ten articles We defined the affect responses as $E = \{anger, sadness, joy, displeasure, surprise, fear\}$. We asked the readers to label each article they read with at least one affect response. There were 220 articles with responses (i.e., $|\{d \mid d \in [D], H_d \neq \emptyset\}| = 220$). We used 30 responses for each article ($H_d = 30$ for such d). The mean, median, and mode of the number of responses per readers were 69 30, and 10, respectively. The number of responses (i.e., $|\{l_{h,d,e} = 1 \mid d \in [D], h \in H_d\}$) for each affect for the 220 articles is shown in Table 1.

Since we wanted to evaluate the prediction performances of different methods, we split the 220 responded-to articles into 200 training articles and 20 testing articles. The remaining 550 articles were used as unresponded-to articles to train the proposed model.

6.2 Models Compared

Since we wanted to demonstrate the effectiveness of the proposed model on the task of predicting personal affect responses, we compared its performance with those of the PC and IPC models.

- **PC**: personal classifier model [Kajino *et al.*, 2012]. We used logistic regressions as personal classifiers and a base classifier similarly to [Kajino *et al.*, 2012]. The workers had |*E*| personal classifiers for each affect response.
- **IPC**: independent personal classifier model. Each reader has his or her own personal classifier, as in the PC model, but does not have a base classifier, unlike the PC model. That is, in the IPC model, personal classifiers are learned independently, so the IPC model is a baseline model.
- **Proposed**: proposed model described in Section 4. We set the number of topics per article *K* to 5. Since the proposed model can use unresponded to articles for learning, we compared its performance with and without the unresponded to articles, i.e., the remaining 550 articles. We call the proposed model with unresponded-to articles **Proposed-Unresponded-to** and without unresponded to articles for which reader responses are to be predicted can be obtained in advance, and the proposed model can leverage such test articles for training similar to methods for transductive learning [Vapnik, 1998]. We call the proposed model with test article information

¹https://www.rondhuit.com/download.html#ldcc

²http://www.lancers.jp/

Table 2: ROC-AUC of PC, independent PC (IPC), and proposed models for each affect label and average ROC-AUC.

	anger	sadness	joy	displeasure	surprise	fear	average
PC	0.569	0.539	0.635	0.641	0.563	0.599	0.591
IPC	0.556	0.578	0.565	0.604	0.584	0.554	0.573
Proposed-Transductive	0.655	0.669	0.564	0.581	0.669	0.665	0.634
Proposed-Unresponded-to-Transductive	0.644	0.688	0.602	0.577	0.634	0.676	0.637
Proposed	0.650	0.685	0.570	0.597	0.675	0.661	0.640
Proposed-Unresponded-to		0.687	0.559	0.567	0.643	0.683	0.625

Proposed-Transductive and without test article information **Proposed**. Similarly, we call the proposed model with unresponded-to articles and test articles **Proposed-Unresponded-to-Transductive** and without unresponded-yo articles and with test articles **Proposed-Unresponded-to**.

The DS model [Dawid and Skene, 1979] and MF model [Koren *et al.*, 2009] (and similar models [Welinder *et al.*, 2010]) were not compared because they cannot predict the responses of readers to unresponded-to articles. We evaluated them by using the area under the receiver operating characteristic curve (ROC-AUC) on the 20 test articles.

6.3 Results and Discussion

The results are shown in Table 2. The average ROC-AUC values among affect labels for all versions of the proposed model were higher than those for the PC and IPC models. Unfortunately, using unresponded-to articles and/or test articles was not effective, possibly due to the small numbers of unresponded-to and test articles. Future work includes investigating the effects of using larger numbers of each. Furthermore, the mode number of responses per reader was 10, which is insufficient for learning the latent features of a reader. As with the clustering PC model [Kajino *et al.*, 2013], clustering readers on the basis of their latent features is a promising way to efficiently learn the latent features of readers. Future work also includes extending our model to include reader clustering.

7 Summary

Our proposed model for personalized affect analysis is a natural extension of the LDA model. Personal affect responses are obtained from the latent features of articles and readers, which are easy to interpret. Testing demonstrated that the proposed model outperforms existing models on the task of predicting personal responses.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers JP15H02782 and JP18H03337, the Telecommunications Advancement Foundation, and Global Station for Big Data and Cybersecurity, a project of Global Institution for Collaborative Research and Education at Hokkaido University.

References

- [Blei *et al.*, 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.
- [Dawid and Skene, 1979] A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statics)*, 28(1):20–28, 1979.
- [Duan *et al.*, 2014] Lei Duan, Satoshi Oyama, Haruhiko Sato, and Masahito Kurihara. Separate or joint? estimation of multiple labels from crowdsourced annotations. *Expert Systems with Applications*, 41(13):5723–5732, 2014.
- [Kajino et al., 2012] Hiroshi Kajino, Yuta Tsuboi, and Hisashi Kashima. A convex formulation for learning from crowds. In Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI 2012), pages 73–79, 2012.
- [Kajino et al., 2013] Hiroshi Kajino, Yuta Tsuboi, and Hisashi Kashima. Clustering crowds. In Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI 2013), pages 1120–1127, 2013.
- [Kingma and Welling, 2014] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *Proceedings* of the International Conference on Learning Representations, 2014.
- [Koren *et al.*, 2009] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.
- [Kucukelbir et al., 2017] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M Blei. Automatic differentiation variational inference. *The Journal* of Machine Learning Research, 18(1):430–474, 2017.
- [Liu et al., 2003] Hugo Liu, Henry Lieberman, and Ted Selker. A model of textual affect sensing using real-world knowledge. In Proceedings of the 8th international conference on Intelligent user interfaces, pages 125–132. ACM, 2003.
- [Maas et al., 2011] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, HLT '11, pages 142–150, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

- [Porteous et al., 2008] Ian Porteous, David Newman, Alexander Ihler, Arthur Asuncion, Padhraic Smyth, and Max Welling. Fast collapsed gibbs sampling for latent dirichlet allocation. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 569–577, 2008.
- [Ptaszynski *et al.*, 2013] Michal Ptaszynski, Hiroaki Dokoshi, Satoshi Oyama, Rafal Rzepka, Masahito Kurihara, Kenji Araki, and Yoshio Momouchi. Affect analysis in context of characters in narratives. *Expert Systems with Applications*, 40(1):168–176, 2013.
- [Tokuhisa *et al.*, 2008] Ryoko Tokuhisa, Kentaro Inui, and Yuji Matsumoto. Emotion classification using massive examples extracted from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 881–888. Association for Computational Linguistics, 2008.
- [Vapnik, 1998] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [Welinder *et al.*, 2010] Peter Welinder, Steve Branson, Pietro Perona, and Serge J Belongie. The multidimensional wisdom of crowds. In *Advances in neural information processing systems*, pages 2424–2432, 2010.
- [Whitehill *et al.*, 2009] Jacob Whitehill, Ting-fan Wu, Jacob Bergsma, Javier R Movellan, and Paul L Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, pages 2035–2043, 2009.