Teaching Interactively to Learn Emotions in Natural Language

Rajesh Titung
Rochester Institute of Technology
New York, USA
rt7331@rit.edu

Cecilia O. Alm
Rochester Institute of Technology
New York, USA
coagla@rit.edu

Abstract

Motivated by prior literature, we provide a proof of concept simulation study for an understudied interactive machine learning method, machine teaching (MT), for the text-based emotion prediction task. We compare this method experimentally against a more well-studied technique, active learning (AL). Results show the strengths of both approaches over more resource-intensive offline supervised learning. Additionally, applying AL and MT to fine-tune a pre-trained model offers further efficiency gain. We end by recommending research directions which aim to empower users in the learning process.

1 Introduction

We examine Machine Teaching (MT), an understudied interactive machine learning (iML) method under controlled simulation for the task of *text-based emotion prediction* (Liu et al., 2003; Alm et al., 2005; Alm and Sproat, 2005; Aman and Szpakowicz, 2007; Alm, 2010; Bellegarda, 2013; Calvo and Mac Kim, 2013; Mohammad and Alm, 2015). This problem intersects with *affective computing* (Picard, 1997; Calvo et al., 2015; Poria et al., 2017), and a family of language inference problems characterized by human *subjectivity* in learning targets (Alm, 2011) and semantic-pragmatic meaning (Wiebe et al., 2004). Both subjectivity and the lack of data for learning to recognize affective states motivate iML techniques. Here, we focus on resource efficiency. Our findings from simulations provide directions for user experiments.

Human perception - and thus human annotators' interpretation - is influenced by human factors such as preferences, cultural differences, bias, domain expertise, fatigue, time on task, or mood at annotation time (Alm, 2012; Amidei et al., 2020; Shen and Rose, 2021). Generally, experts with long-standing practice or in-depth knowledge may also not share consensus (Plank et al., 2014). Inter-subjective disagreements can reflect invalid noise artifact (detectable by humans) or *ecologically valid* differences in interpretation.

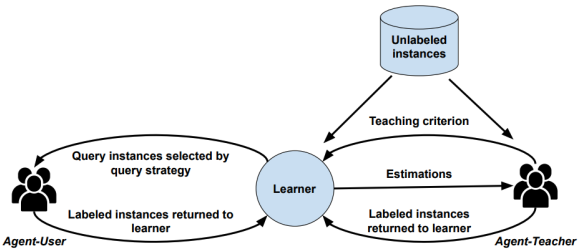


Figure 1: Comparison of interactive *Active Learning* (left) with *Machine Teaching* (right). Training instances are labeled by the Agent-User (in AL) or the Agent-Teacher (in MT).

Holzinger (2016) define iML methods as algorithmic procedures that "*can interact with agents and can optimize their learning behavior through these interactions* [...]" (p. 119). In our study, the stakeholders in the learning process are models (*learners*) and humans (*agent-users* or *agent-teachers*). Tegen et al. (2020) posit that iML involves either *Active Learning* (AL) or interactive *Machine Teaching* (MT),[1] based on humans' role in the learning loop. In AL, the learning algorithm uses *query strategies* (e.g., triggered by *uncertainty*) to iteratively select instances from which it learns (Settles, 2009) if licensed by a *budget*; with a human agent who annotates upon learner request. In contrast, in MT, the teacher (user) who possesses problem knowledge instead selects the instances to be labeled and uses them to train the learner (Zhu, 2015). Initial, foundational MT research focused on constructing a minimal, ideal set of training data, striving for optimality in the data the learner is presented with to learn from. Interactive MT assumes human agent interaction with the learner (Liu et al., 2017), for enabling time- and resource-efficient

[1] We use conventions from Tegen et al. (2020) where MT means an iterative, interactive implementation of Machine Teaching. MT here is not Machine Translation.

Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing, pages 40 - 46

model convergence. Following the training by error criterion described in Tegen et al. (2020), if the learner is unable to predict the right answer, and the budget allows, the human teacher instructs the learner with the label. Thus, AL leverages measures to wisely choose instances for human labeling and subsequent learning, whereas MT capitalizes on the teacher's knowledge to wisely select training instances and proceed to learn when the criterion to teach is met (cf. Figure 1).

2 Related Work and Background

Olsson (2009) discussed AL for NLP tasks, while Schröder and Niekler (2020) discussed deep learning with AL. Our study also builds on Tegen et al. (2020)'s use of simulation to study AL query strategies and MT assessment and teaching criteria. Lu and MacNamee (2020) reported on experiments where transformer-based representations performed consistently better than other text representations, taking advantage of the label information that arises in AL. An et al. (2018) also suggested assigning a varying number of instances to label per human oracle based on their capability/skills and the amount of unlabeled data, which reduced the time required by the deep learner without negatively impacting performance. We comparatively study iML in the fine-tuning stages. Bai et al. (2020) emphasized language-based attributes like reconstruction of word-based cross-entropy loss across words in sentences toward instance selection. To ensure improved experimental control and avoid confounding variables, we focus on uncertainty-based strategies for AL.

MT deals with a teacher designing a well-reasoned, ideally optimal, training set to drive the learner to the desired target concept/model (Zhu, 2015; Zhu et al., 2018). While there has been some progress in the use of MT, its application in NLP is present in its earliest form with little empirical exploration or refinement. MT has been explored mostly in computing security, where the teacher is a *hacker/advisor* who selects training data to adjust the behavior of an adaptive, evolving learner (Alfeld et al., 2016, 2017). Tegen et al. (2020) reported that MT could greatly reduce the number of instances required, and even outperformed most AL strategies. These findings are compelling and motivate exploring MT's potential in NLP, which, however, has some distinct characteristics, including high-dimensional data impacted by scarcity. MT's possibilities in NLP are thus as of yet largely unknown. We begin here by focusing on controlled experimental simulations to examine resource-efficiency and performance in text-based emotion prediction, whereas future work will take a step closer to ecological validity in interactive MT with real-time agent-teachers.

Overall, several prospects can be noted for NLP with interactive Machine Learning (iML):

- Human knowledge and insights can be leveraged to make the *search space substantially smaller* by systematic instance selection (Holzinger, 2016), achieving adequate performance with fewer training instances.
- In a setting where learning occurs online or continually (Tegen et al., 2019), iML enables *sustained learning* over time, with new or updated data offered to the learner. This especially makes sense for natural language tasks which by nature are characterized by linguistic change.
- Using iML can enable model *customization* to specific users, schools of thought, and enable privacy-preserving models (Bernardo et al., 2017), e.g., for deploying NLP on edge devices.
- IML enables users to directly influence the model (Amershi et al., 2014), and interactive techniques can aid agents to *catch bias or concept drift early* in the development process.
- The iML paradigm enables an *initial state with limited data* (or even a *cold start*), which applies to NLP for underresourced languages, low-data clinical problems, etc., including NLP for affective computing since many affective states remain understudied (Alm, n.d.).
- By learning more resource-efficiently, iML has potential to *lower NLP's carbon footprint*.

While iML is promising, issues include:

- Humans users or teachers are *not necessarily willing or available* to provide input or feedback to a system (Donmez and Carbonell, 2010).
- The iML setup is not immune to *catastrophic forgetting* (Holzinger, 2016) in online learning.
- Human factors introduce technical considerations that may impact interaction and performance success; for instance, the learning set-up should accommodate *human fatigue* (Darani and Kaedi, 2017; Llorà et al., 2005).

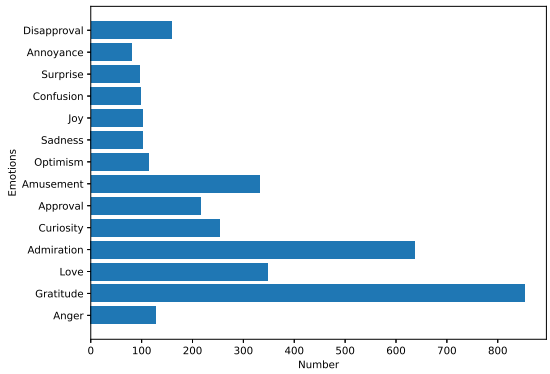


Figure 2: Class imbalance for the 14-class emotion data.

3 MT/AL for Emotion Prediction

Text-based emotion data are subject to variation and ambiguity, which adds to the difficulty in the annotation process, compounded with data scarcity for capturing many affective states. IML methods can be a means to deal with data limitations.

In this study, we used a subset of the GoEmotions dataset (Demszky et al., 2020) which consists of emotion labels for Reddit comments. We prioritized resource-efficiency as the primary experimental variable over exploring impact on target concept ambiguity. Figure 2 shows the imbalanced distribution of emotion classes in this subset. The training and test sets comprised approximately 2800 and 700 instances respectively. In all experiments, the learner was trained initially with 10% of the training set while the remaining 90% was reserved as an unlabeled pool of data which were gradually added to the training set in each iteration.[2] The simulated 'user' had access to the labels of the instances from the unlabeled dataset whenever required via dataset lookup.

3.1 AL vs. MT for Emotion Prediction

We compared the effect of AL and MT strategies and further compared to offline supervised machine learning, referred to as *all-in-one batch*.

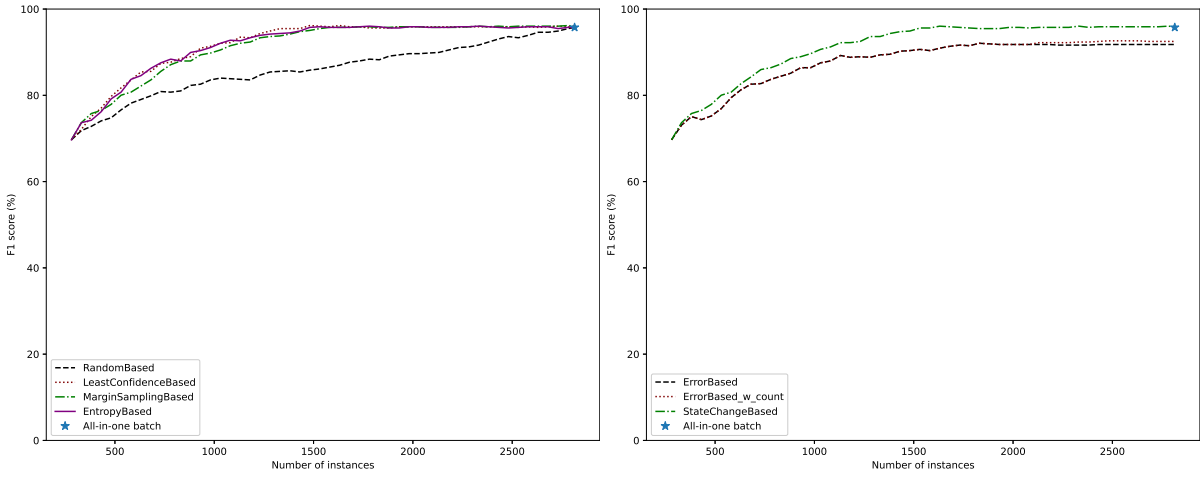
Motivation In our AL experiment, the learner queried the instances using versions of *uncertainty sampling* or a *random* approach. In the *least confident* strategy, the learner selects instances for query for which it has the least probability of prediction in its most probable class; in *margin sampling*, instances with the smallest difference between its top-two most likely classes; and in *entropy*, with the largest entropy (Olsson, 2009; Tegen et al., 2020).

In MT, the agent-teacher chooses instances (Zhu, 2015), which are then labeled and used to teach the learner (Tegen et al., 2020). We simulated the *margin sampling*-based AL query strategy as a teacher to select a set of instances. Moreover, *error-based* and *state change* are two *teaching criteria* used by Tegen et al. (2020) for initiating teaching. In the error-based method, the teacher proceeds to teach based on correctness of the learner's estimation, i.e., supplying the learner with the correct label for wrong estimations. We introduce a modification termed *error-based training with counting* where the teacher continues to provide labeled instances to the learner when all estimations are accurate in two consecutive iterations to ensure periodic model updating. In the state change-based criterion, the teacher provides a label for the instance if the current instance's real class label differs from the prior instance's class label. When no label is given, the learner assumes the instance's label is the same as the last label given by the teacher.

Methods We focus on transportability and opted for sklearn's Linear SVM with hinge loss given its lean computational character (Buitinck et al., 2013; Chang and Lin, 2011). Both setups were trained on CPUs, with MT using state change as teaching criterion taking the longest time (around 40 min).

Results and Discussion Panel (a) in Figure 3 shows the result for AL strategies. The performance on emotion prediction in text is more resource-efficient and uses less data with AL. The query strategies achieved the performance equivalent to learning with the full batch of training data after using just around half of the data with AL, and all perform better than random selection. A Wilcoxon's Rank Sum Test (Wilcoxon, 1992) for independent samples compared random against other query strategies. This indicated a significant difference in their performance with $p < 0.05$. Panel (b) shows the MT results for three teaching criteria. State change improves over the error-based approach, while the error-based approach with counting slightly enhances the regular error-based approach because of the modification introduced. We also observe that since we used margin-based AL

[2]For the Huggingface transformers library 20% of the training set was held out as a validation set before this 90-10 split. For sklearn, attempts at hyperparameter tuning–for the C parameter, dual/primal problem and tolerance values for stopping criteria–used a genetic algorithm without meaningful performance difference, and results are provided with defaults, with class weights initiated as the inverse of the frequency of each class.



(a) AL query strategies

(b) MT teaching criteria

Figure 3: Text-based emotion prediction with (a) AL query strategies or (b) MT teaching criteria. The all-in-one batch option (green star) signifies resource-inefficient offline batch training.

as a teacher for selecting instances, the result mirrors margin sampling-based AL in panel (a). Moreover, we note that error-based teaching saturates, potentially reflecting that state change-based teaching is more capable of dealing with imbalanced data (Tegen et al., 2020). Overall, the encouraging results motivate us to plan to assess utility in a real-time MT scenario with a human teacher and deeper study of teacher variations for data selection and revised teaching criteria for initiating training.

3.2 Fine-tuning with AL and MT

Motivation Previous results showed that MT and AL can build better models more efficiently with annotation savings (time and cost). Here, we explore if fine-tuning a pre-trained model – a frequent and often performance-boosting approach in NLP – that uses iML concepts can improve results further.

Methods We fine-tune a pre-trained BERT model (Devlin et al., 2019) to emotion prediction in text using Huggingface (Wolf et al., 2020), with a max. sequence length of 80 (since comments tend to be quite short). Based on prior observations, we analyze fine-tuning performance with AL for the least confident and margin sampling strategies, and with MT for the error-based and error-based with counting teaching criteria.

Results and Discussion Figure 4 shows the outcomes for fine-tuning BERT interactively. The results show performance close to 96%, which is good for this subjective task. Moreover, AL matched the offline training performance using less than half of the available instances. We note that

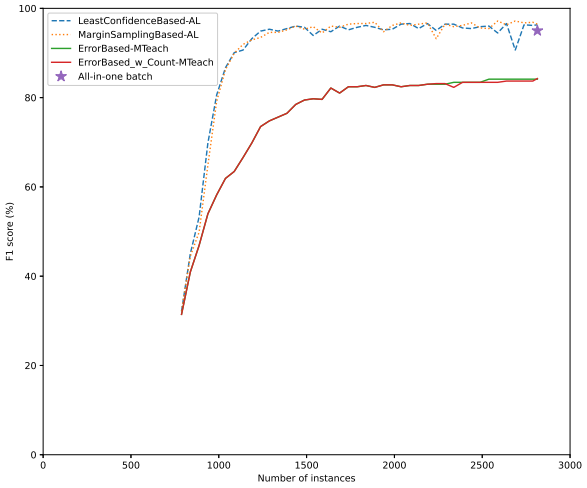


Figure 4: Text-based emotion prediction when using AL or MT in fine-tuning with BERT.

convergence for fine-tuning also required somewhat less data than in the prior SVM-based experiment, as shown by the steeper slope of performance increment. Yet how to better leverage MT in conjunction with fine-tuning, or transfer techniques generally, remains a key priority in continued study.

4 Discussion

We showed that iML efficiently produces desired results for text-based emotion prediction. MT remains understudied and should be further explored for NLP tasks. Fine-tuning a pre-trained model with AL can leverage the strengths of both approaches with small datasets. In addition to experiments detailed above, we explored training the learner *incrementally* (online training) versus in a

non-incremental setup (the learner is trained using accumulated training set up to the most recent query). The incremental approach experiences *catastrophic forgetting* but requires very little time for learner updating and can thus work well under low memory usage, e.g., for a life-long learning setting or edge devices.

5 Conclusion

Our study on text-based emotion prediction demonstrated the potential of both MT and AL methods. We offered initial experimentation with MT and AL for this problem, and based on promising results under controlled simulation, next steps will focus on real-time user/teacher interactions, a broader set of teaching criteria, and new forms of training instance selection. In addition, we are interested in exploring heavily understudied affective states, which are currently not covered sufficiently or not covered at all in annotated emotion corpora. We also suggest focused research on specialized teachers in NLP tasks toward better selection of training data. Teachers who assess the learner and decide the right time to offer an adequate set of new information may also help create more robust or interpretable learners which evolve over time.

Ethics Statement

A limitation of this work is that it did not consider linguistic characteristics of the pre-trained models (Bai et al., 2020). We used an artificial teacher in MT and did not deeply examine hybrid MT-AL strategies, although we used an AL approach as teacher in the MT setup. Still, this work may stimulate NLP researchers to consider the benefits of AL and MT, especially for challenging subjective NLP tasks such as text-based emotion prediction (Alm, 2011). Additionally, continued work can explore how the findings apply in the context of other corpora, including with multimodal data.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. DGE-2125362. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Scott Alfeld, Xiaojin Zhu, and Paul Barford. 2016. Data poisoning attacks against autoregressive models. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 1452–1458. AAAI Press.

Scott Alfeld, Xiaojin Zhu, and Paul Barford. 2017. Explicit defense actions against test-set attacks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 1274–1280. AAAI Press.

Cecilia Ovesdotter Alm. 2010. Characteristics of high agreement affect annotation in text. In *Proceedings of the Fourth Linguistic Annotation Workshop*, pages 118–122, Uppsala, Sweden. Association for Computational Linguistics.

Cecilia Ovesdotter Alm. 2011. Subjective natural language problems: Motivations, applications, characterizations, and implications. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 107–112. Association for Computational Linguistics.

Cecilia Ovesdotter Alm. 2012. The role of affect in the computational modeling of natural language. *Language and Linguistics Compass*, 6(7):416–430.

Cecilia Ovesdotter Alm. n.d. Linguistic data resources for computational emotion sensing and modeling.

Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Cecilia Ovesdotter Alm and Richard Sproat. 2005. Emotional sequencing and development in fairy tales. In *Affective Computing and Intelligent Interaction*, pages 668–674, Berlin, Heidelberg. Springer Berlin Heidelberg.

Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Mautner P. Matoušek V., editor, *Text, Speech and Dialogue TSD 2007*, pages 196–205. Springer.

Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120.

Jacopo Amidei, Paul Piwek, and Alistair Willis. 2020. Identifying annotator bias: A new IRT-based method for bias identification. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4787–4797, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Bang An, Wenjun Wu, and Huimin Han. 2018. Deep active learning for text classification. In *Proceedings of the 2nd International Conference on Vision, Image and Signal Processing*, ICVISP 2018, New York, NY, USA. Association for Computing Machinery.

Guirong Bai, Shizhu He, Kang Liu, Jun Zhao, and Zaiqing Nie. 2020. Pre-trained language model based active learning for sentence matching. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1495–1504, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Jerome R. Bellegarda. 2013. Data-driven analysis of emotion in text using latent affective folding and embedding. *Computational Intelligence*, 29(3):506–526.

Francisco Bernardo, Michael Zbyszynski, Rebecca Fiebrink, and Mick Grierson. 2017. Interactive machine learning for end-user innovation. In *AAAI Spring Symposia*, pages 369–375.

Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. 2013. API design for machine learning software: Experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122.

Rafael Calvo, Sidney D'Mello, Jonathan Gratch, and Arvid Kappas, editors. 2015. *The Oxford Handbook of Affective Computing*. Oxford University Press.

Rafael A. Calvo and Sunghwan Mac Kim. 2013. Emotions in text: Dimensional and categorical models. *Computational Intelligence*, 29(3):527–543.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Zahra Sheikhi Darani and Marjan Kaedi. 2017. Improving the interactive genetic algorithm for customer-centric product design by automatically scoring the unfavorable designs. *Human-centeric Computing and Information Sciences*, 7(38).

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics"*, pages 4040–4054, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pinar Donmez and Jaime Carbonell. 2010. *From Active to Proactive Learning Methods*, volume 262, pages 97–120. Springer Berlin Heidelberg.

Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, 3(2):119–131.

Hugo Liu, Henry Lieberman, and Ted Selker. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, IUI '03, page 125–132, New York, NY, USA. Association for Computing Machinery.

Weiyang Liu, Bo Dai, Ahmad Humayun, Charlene Tay, Chen Yu, Linda B. Smith, James M. Rehg, and Le Song. 2017. Iterative machine teaching. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 2149–2158. JMLR.org.

Xavier Llorà, Kumara Sastry, David E. Goldberg, Abhimanyu Gupta, and Lalitha Lakshmi. 2005. Combating user fatigue in IGAs: Partial ordering, support vector machines, and synthetic fitness. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, GECCO '05, page 1363–1370, New York, NY, USA. Association for Computing Machinery.

Jinghui Lu and Brian MacNamee. 2020. Investigating the effectiveness of representations based on pre-trained transformer-based language models in active learning for labelling text datasets. *arXiv e-prints*, page arXiv:2004.13138.

Saif Mohammad and Cecilia O. Alm. 2015. Computational analysis of affect and emotion in language. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, Lisbon, Portugal. Association for Computational Linguistics.

Fredrik Olsson. 2009. A literature survey of active machine learning in the context of natural language processing. Technical report, Swedish Institute of Computer Science.

Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA.

Barbara Plank, Dirk Hovy, and Anders Søgaard. 2014. Linguistically debatable or just plain wrong? In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 507–511, Baltimore, Maryland. Association for Computational Linguistics.

Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37:98 – 125.

Christopher Schröder and Andreas Niekler. 2020. A survey of active learning for text classification using deep neural networks. *arXiv*, 2008.07267.

Burr Settles. 2009. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison.

Qinlan Shen and Carolyn Rose. 2021. What sounds "right" to me? Experiential factors in the perception of political ideology. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1762–1771, Online. Association for Computational Linguistics.

Agnes Tegen, Paul Davidsson, and Jan A. Persson. 2019. Towards a taxonomy of interactive continual and multimodal learning for the internet of things. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 524–528, New York, NY, USA. Association for Computing Machinery.

Agnes Tegen, Paul Davidsson, and Jan A. Persson. 2020. A taxonomy of interactive online machine learning strategies. In *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part II*, volume 12458 of *Lecture Notes in Computer Science*, pages 137–153. Springer.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.

Frank Wilcoxon. 1992. Individual comparisons by ranking methods. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Methodology and Distribution*, pages 196–202. Springer New York, New York, NY.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Xiaojin Zhu. 2015. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).

Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N. Rafferty. 2018. An overview of machine teaching. *CoRR*, abs/1801.05927.