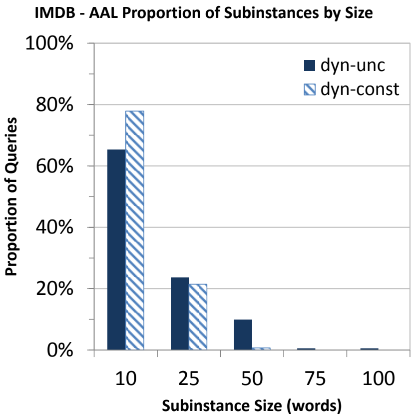
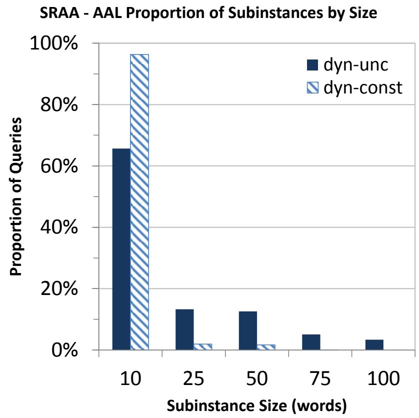


(a) IMDB



(b) SRAA

Figure 4: Proportion of subinstance sizes selected by dynamic AAL.

		Obs.	Exp.
IMDB	const	15%	48%
	unc	34%	45%
SRAA	const	2%	50%
	unc	37%	45%

Table 2: The percentage of observed neutral labels for `dynamic-unc` and `dynamic-const`, compared with what is expected for subinstances of the observed sizes.

Uncertainty improves the exploration of AAL algorithms. In general, we find that using uncertainty outperforms constant utility. For example, in Figure 3, `dynamic-unc` outperforms `dynamic-const` on both datasets. This supports prior work showing the benefits of uncertainty sampling. Additionally, we find that the uncertainty term of the objective balances the neutrality term, enabling a better exploration of the sample space. For example, in the SRAA dataset, Figure 3(b) shows that `dynamic-const` stops learning after a few rounds of active learning. Interestingly, analysis of this result showed that the $Q(z|\mathbf{x}_i^k)$ model quickly learned in the first iterations a strong correlation between non-neutral response and certain terms, e.g. "GPL" (proper name indicating class "auto") in the subject line; `dynamic-const` selected subinstances with those terms thereafter obtaining non-neutral labels for those subinstances. Thus, $Q(z|\mathbf{x}_i^k)$ was further reinforced to predict the subinstances that contain the word "GPL" as non-neutral. This prevented the neutrality model from exploring other terms and the student did not receive a diverse set of documents to improve learning. Including the uncertainty term encouraged the student to seek a more diverse set of examples, thus avoiding this problem.

This is further supported by Table 2, which shows that `dynamic-const` primarily selects instances for which it is very likely to receive a non-neutral label. E.g., in SRAA, only 2% of instances receive a neutral label for `dynamic-const`; Figure 3(b) shows this to be an ineffective strategy, since the student observes only a homogeneous subset of examples. While `dynamic-unc` increases the rate of neutral labels, this additional cost is worth the greater diversity of labeled data.

Related Work

There has been a significant amount of work on noisy and reluctant oracles for traditional active learning scenarios without anytime capability (Donmez and Carbonell 2008; Donmez, Carbonell, and Schneider 2009; Fang, Zhu, and Zhang 2012; Wallace et al. 2011; Yan et al. 2011). Much of this work considered which instance to show to which oracle, and the oracle's quality is fixed. In the anytime framework that we propose in this paper, the student has control over how much time an oracle should spend on an instance, thus controlling the quality of the label.

There has also been a significant amount of work on cost-sensitive active learning (Settles, Craven, and Friedland 2008; Donmez and Carbonell 2008; Haertel et al. 2008; Kapoor, Horvitz, and Basu 2007; Tomanek and Hahn 2010). The common strategy is to use a utility-cost ratio to determine the most cost-effective instances. We follow the same strategy and use utility-cost ratio, with the additional multiplicative factor of probability of non-neutrality.

We build on our previous work (Ramirez-Loaiza, Culotta, and Bilgic 2013) about the problem of searching over subinstances with some notable differences: i) we conducted user studies to determining annotation time, whereas they assumed a linear cost function, ii) we allow the oracles to return a neutral label and model neutrality.

Conclusions and Future Work

We present an anytime active learning framework in which the student is allowed to interrupt the oracle to save annotation time. User studies were conducted to quantify the relationship between subinstance size, annotation time, and response rate. These were used to inform a large-scale simulated study on two document classification tasks, which showed that although interruption can cause the oracle return neutral labels, interrupting at the right time can lead to significantly more efficient learning. We found that optimal interruption time depends on the domain and proposed a dynamic AAL strategy that is better than or comparable to the best static strategy that uses a fixed interruption time.

In the future, we will expand our model of annotation time to account for lexical content. Moreover, we assumed in this paper that the annotator reads the document serially starting from the beginning and hence created subinstances that correspond to the first k words of the document. As future work, we will consider alternative techniques of interruption, such as structured reading (e.g., the first and last sections of a document) and text summarization to speed up annotation.

References

Donmez, P., and Carbonell, J. G. 2008. Proactive learning: : Cost-sensitive active learning with multiple imperfect oracles. In *Proceeding of the 17th ACM conference on Information and Knowledge Mining - CIKM '08*, 619.

Donmez, P.; Carbonell, J. G.; and Schneider, J. 2009. Efficiently learning the accuracy of labeling sources for selective sampling. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, 259.

Fang, M.; Zhu, X.; and Zhang, C. 2012. Active Learning from Oracle with Knowledge Blind Spot. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence - ICML '12*.

Haertel, R.; Ringger, E.; Seppi, K.; Carroll, J.; and Mcclanahan, P. 2008. Assessing the Costs of Sampling Methods in Active Learning for Annotation. In *ACL*, 65–68.

Kapoor, A.; Horvitz, E.; and Basu, S. 2007. Selective supervision: Guiding supervised learning with decision-theoretic active learning. In *IJCAI*, volume 7, 877–882.

Lewis, D. D., and Gale, W. A. 1994. A sequential algorithm for training text classifiers. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, 3–12.

Maas, A. L.; Daly, R. E.; Pham, P. T.; Huang, D.; Ng, A. Y.; and Potts, C. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 142–150. Portland, Oregon, USA: Association for Computational Linguistics.

Mazzoni, D.; Wagstaff, K.; and Burl, M. 2006. Active learning with irrelevant examples. In *ECML*, 695–702.

Nigam, K.; McCallum, A.; Thrun, S.; and Mitchell, T. 1998. Learning to classify text from labeled and unlabeled documents. In *Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence*, AAAI '98/IAAI '98, 792–799.

Ramirez-Loaiza, M. E.; Culotta, A.; and Bilgic, M. 2013. Towards anytime active learning: Interrupting experts to reduce annotation costs. In *KDD Workshop on Interactive Data Exploration and Analytics (IDEA)*.

Roy, N., and McCallum, A. 2001. Toward optimal active learning through sampling estimation of error reduction. In *International Conference on Machine Learning*, 441–448.

Settles, B.; Craven, M.; and Friedland, L. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, 1–10.

Settles, B. 2012. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool.

Tomanek, K., and Hahn, U. 2010. A comparison of models for cost-sensitive active learning. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 1247–1255.

Wallace, B. C.; Small, K.; Brodley, C. E.; and Trikalinos, T. A. 2011. Who should label what? instance allocation in multiple expert active learning. *In Proc. of the SIAM International Conference on Data Mining (SDM)*.

Yan, Y.; Fung, G. M.; Rosales, R.; and Dy, J. G. 2011. Active learning from crowds. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 1161–1168.