CSE 584 – MACHINE LEARNING: TOOLS AND ALGORITHMS
HOMEWORK #1

Naga Aravind Kundeti
PSU ID: 970551363

PAPER 1: Active Learning and Machine Teaching for Online Learning: A Study of Attention and Labelling Cost

1. What problem does this paper try to solve, i.e., its motivation?

This research paper, *"Active Learning and Machine Teaching for Online Learning: A Study of Attention and Labelling Cost,"* tackles the big expense and hard work of annotation. This gets important when dealing with data that comes in real-time.

The issue stems from most traditional active learning and machine learning methods assuming unrealistic actions from human commentators. For instance, active learning and machine learning view humans; active learning considers humans passive, while machine learning assumes annotators monitor. An ideal solution would maintain good categorization performance while minimizing the work and attention human annotators need to dedicate to annotation.

2. How does it solve the problem?

The researchers present an interactive system for online learning that combines machine learning with active learning methods. This approach allows both parts to boost each other. The system deepens our grasp of labeling costs looking at not just the price of labeling but also the time and work facing the annotators.

This setup chooses when to ask the human annotator for labels on the fly, based on the current labeling budget and expected labeling expenses. This helps to even out the workload for the human annotator. To get the most out of both focus and effort, the research also explores methods from random sampling to uncertainty sampling.

3. A list of novelties/contributions?

- **Interactive Online Learning approach:** This new method takes into account both labeling cost and attention cost, which were handled.
- **Combination of Active Learning and Machine Teaching:** To strike a better balance between system and annotator interaction, the framework combines the system-driven active learning method with the human-driven machine teaching approach.

- **Attention Cost Modeling:** This work adds the idea of "attention cost" to active learning. It defines the mental effort required from the annotator and gives a deeper understanding of human involvement.
- **Cold Start Settings Experiments:** The team tested the developed framework on real-world time-bound data streams. This provided valuable insights into the trade-offs between performance, effort, and attention in cold start settings.

4. What do you think are the downsides of the work?

- The framework is said to be of a new generation but the emphasis is on the theory construction and the outcome of the experiments.
- This may mean that there is little investigation of generalizability to the real-word and, hence, the promise may be small scale of this proposed approach.
- As such, the study further the preceding hypotheses; however, it simplifies the actions initiated by human annotators.
- Further interpretative research about this way in which people are using machine learning systems has to be carried out since there is a possibility that they are not as straightforward as they seem.
- As highlighted by the study, the costs of labeling and querying are not fixed; it is for this reason that in most real applications, these may not be fixed and may rather change with time, adversely affecting systems' performance.
- Further research on such aspects of human behavior is needed to make the system even more useful in reality: other operational expenses which they refer to as dynamic costs.

PAPER 2: Active Learning System in the Context of the Social Semantic Web

1. What problem does this paper try to solve, i.e., its motivation?

The paper is *"Active Learning System in the Context of the Social Semantic Web"* and the objective of the paper is to explain the various challenges that are prevalent in traditional eLearning scenario and how they are being handled in the online learning environment. The idea for this research stems from the nature of online education where not only the content but also the learner has to be managed in order to achieve the desired results. The authors present an Active Semantic Learning System (ASLS) that utilizes the characteristics of the Semantic Web technologies for delivering individual and, at the same time, more effective learning content. With the addition of semantic annotations and domain ontologies to the proposed system, the delivery of learning aids is optimized according to students' requirements and preferences while promoting interactivity.

In addition to that, the paper also focuses on the aspects of merging social learning strategies into eLearning systems. It also shows how the active semantic learning system helps the integration of various data sources and user generated content from the Social Web to enrich the context for learning. This integration will enhance the advance of knowledge base that will not only help the learners to grasp concepts but also activates the learners. The authors elaborate the idea that using semantic technologies it is possible to enhance the factors of content reuse and interoperability, as a result making the online learning more effective and attractive.

2. How does it solve the problem?

The authors introduce the Active Semantic Learning System (ASLS) which uses Semantic Web technologies and Social Web in learning environment as an improvement towards personalized learning process. Competencies, quizzes, and learning objects are marked up using ontologies employed in the system. The ASLS operates thus as a kind of personal assistant that helps to filter, select and selectively present to the learner different kinds of knowledge depending on the learner's profile.

The system incorporates the Semantic Kernel that links a variety of learning modules and performs the reasoning as a result of the queries based on the SPARQL language. The learning ontologies used in the system apply to different facets of learning including quizzes and learner profiles that can be altered according to the learner's needs.

3. A list of novelties/contributions?

- **Integration of Semantic Web and Social Web:** The system integrates semantic technologies with the-elements-of-sociocultural-approaches-to-learning-and-education social learning strategies, enabling a range of social learning modes to be constructed and deepened for using both machine-readable data and users' content.

- **Semantic Kernel:** A central hub that controls interactions of the learning modules, while receiving, processing and issuing queries for data from various sources.
- **Adaptive Learning Environment:** The system caters for the learner by having his/her characteristics, progress records, and learning goals into consideration to make the learning process enjoyable.
- **Use of Ontologies for Various Learning Objects:** It presents domain ontologies for expressing learning objects, quizzes, skills and learners and incorporates powers of semantic reasoning and inference over the content.
- **Bridge Ontology:** An access layer that accouples content and application ontologies and enables queries and Reasoning about Learning content.

4. What do you think are the downsides of the work?

- Realization of the necessary ontologies for the system is rather time-consuming and may also be a subject to high expenses, provided the increase in the number of websites offering online courses.
- As for the strengths, the system can perform rather well in reasoning among the suggested types of user query; however, the User Interface Module could possibly require further modifications to accommodate more complicated types of user query.
- Incorporation of semantic reasoning within the framework and between different modules and ontologies could be problematic in terms of scalability when number of courses and users increases.
- Data used by the system is imported from outside sources and hence, the system may face problems of inconsistency of data or inadequate information if the data is not updated.

PAPER 3: Addressing the Constraints of Active Learning on the Edge

1. What problem does this paper try to solve, i.e., its motivation?

In the paper entitled *"Addressing the Constraints of Active Learning on the Edge"*, Wei discusses the issues related to using active learning in environments of edge computing. Edge computing enables the implementation of the machine learning model close to the places where the data is generated hence minimizing the transfer of data to the cloud. Nevertheless, active learning procedures are generally not designed to consider various characteristics of edge computations, which are often inferior in terms of hardware capabilities and networking constraints, as well as the geographical distribution of labeled and unlabeled data. The motivation is to produce a framework that will be capable of performing active learning on the edge with these constraints in mind.

2. How does it solve the problem?

The authors put forward a new active learning framework that they have particularly developed for the edge computing application field. This framework is also modular and is built up of the two decision making strategies used in sampling which includes the representative sampling technique where by data points that are selected are deemed to be as diverse and representative of the population as possible and the uncertainty sampling technique where by data points considered as uncertain are given priority during sampling. It also employs a dynamic Bernoulli process to make determination for the uncertainty of model as to whether it should be incorporated when sampling, making the framework versatile to changes in model performances.

In uncertain environment, when the model starts with low performance (cold start), representative sampling is preferred to make worse decision making. The uncertainty sampling occurs more often as is evident when the model updates to higher version. The system intends on being fast, suitable for implementation in different edge devices, and have a low-data requirement and low bandwidth consumption.

3. A list of novelties/contributions?

- **Edge-Compatible Active Learning Framework:** The proposed framework takes into account the limitations that are specific to edge computing including bandwidth and hardware constraints.
- **Dynamic Bernoulli Sampling:** The idea of employing representative and uncertainty sampling Bernoulli process in a dynamic manner is unique, that enables the system to change between different levels of the model performance.

- **Modular Design:** It is highly modular, which means that there is some modularity in the algorithms used, for instance, clustering methods and the manner in which uncertainty is measured can be swapped depending on the preference of the architecture and the amount of computational power available.
- **Cold Start Handling:** This is because the framework is aimed at functioning in cold start cases where the model does not have prior information; the initial iteration will then pre-emptively sample across the entire feature space or a large portion of it.

4. What do you think are the downsides of the work?

- The framework successfully manages to avoid the problem of choosing the most informative points, and provides reasonable accuracy during the initial stages of training of the model but lacks the efficacy compared to more advanced active learning techniques as the accuracy of the model improves.
- Nevertheless, even within the relatively simple module-based structure of the framework, applying it and adapting it to particular cases can lead to further complication, especially if one assumes different and higher degrees of uncertainty or employs more complex clustering models.
- The provided experiments were performed on specific datasets, which are far from covering the whole set of possible use cases of edge computing. Extending the experiments to contain a more diverse set of applications would provide even more support for the suggested approach.
- This is because the system uses data imported from external sources, which can be inaccurate and unreliable; this makes the system vulnerable to providing misleading information to its users.
- Ontology creation and management is a difficult task since the growth of the required ontologies that will support the system takes time and resources particularly as the number of online courses grow.

Naga Aravind Kundeti
PSU ID: 970551363