Teja Sree Parasa September 8th, 2024

CSE 584: Machine Learning Homework 1

Active Learning

I. Large-Scale Visual Active Learning with Deep Probabilistic Ensembles

Kashyap Chitta, Jose M. Alvarez, and Adam Lesnikowski

Motivation

This paper tackles the problem of how to effectively select and annotate data for training deep neural networks, especially when dealing with large datasets. Traditional methods for active learning aim to choose the most useful data for annotation but often struggle with big datasets because estimating uncertainty accurately is tough. While Bayesian Neural Networks (BNNs) provide good uncertainty estimates, they are too slow and complex for large-scale use. Ensemble methods, which combine multiple models to improve accuracy, don't always give reliable uncertainty estimates. The paper's motivation is to create a practical solution that merges the strengths of BNNs and ensemble methods to improve uncertainty estimation and make active learning more efficient for large datasets.

Solution/Approach

To solve this issue, the paper introduces Deep Probabilistic Ensembles (DPEs), a new approach that regularizes neural network ensembles to mimic BNNs. Instead of just focusing on performance, DPEs are designed to enhance uncertainty estimation, which is key for active learning. The authors use variational inference to develop a regularization technique that combines the benefits of BNNs and ensembles. This method allows DPEs to offer better uncertainty estimates without the heavy computational load of deep BNNs. The paper shows that DPEs need less annotated data to achieve high performance and work better than other methods on large-scale tasks like image classification and segmentation, making active learning more practical and efficient.

Novelties/Contributions

1. Introduction of Deep Probabilistic Ensembles (DPEs), combining ensemble methods with Bayesian Neural Networks (BNNs) using KL regularization for improved uncertainty estimates.
2. Use of KL divergence as a regularization technique to promote diversity among ensemble members and reduce overfitting.
3. Development of a new active learning pipeline with DPEs for image classification, focusing on selecting and labeling samples based on uncertainty measures.
4. Evaluation of different growth parameters (Linear-8, Linear-4, Exponential-4) to understand their effects on model performance and computational efficiency.
5. Analysis of various acquisition functions, with the finding that a simple entropy-based function (Hens) performs competitively.
6. Extension of DPEs to active semantic segmentation, improving annotation efficiency with a class-weighted acquisition function that targets specific classes.

Downsides

1. Training DPEs is computationally demanding and may not scale efficiently to very large datasets or real-time applications.
2. KL regularization introduces approximation challenges and sensitivity to network initialization, potentially impacting performance.
3. The method's effectiveness is primarily demonstrated on image classification and semantic segmentation, leaving its performance on other tasks uncertain.
4. The added complexity of sophisticated acquisition functions and class-weighted approaches may not always be justified, as simpler methods might suffice.
5. The approach relies heavily on empirical tuning, which can be labor-intensive and may not generalize well to different datasets or problem settings.

II. Deep Bayesian Active Learning with Image Data
Yarin Gal, Riashat Islam, Zoubin Ghahramani

Motivation

The main issue this paper tackles is how to effectively use active learning (AL) with deep learning models, particularly for high-dimensional data like images. Active learning aims to minimize the amount of labeled data needed by selecting the most informative samples for labeling. However, integrating active learning with deep learning is challenging because deep learning models typically require large datasets and often lack methods to estimate model uncertainty. This is a problem because active learning relies on being able to work with small amounts of data and needs effective ways to measure and utilize model uncertainty to make smart labeling decisions. This issue is particularly pronounced in high-dimensional areas like image data, where existing active learning methods struggle with scalability and effectiveness.

Solution Approach

To solve this problem, the paper suggests combining Bayesian deep learning methods with active learning techniques. The authors introduce a new active learning framework that uses Bayesian convolutional neural networks (BCNNs), which are capable of representing model uncertainty in their predictions. This approach allows the system to learn more efficiently from smaller labeled datasets by focusing on the most informative samples based on the model's uncertainty. The effectiveness of this method is shown through experiments on image classification tasks using the MNIST dataset and melanoma diagnosis with the ISIC 2016 dataset. The proposed framework significantly reduces the need for labeled data and improves classification accuracy compared to traditional active learning methods and other semi-supervised learning techniques.

List of Novelties/Contributions

1. The paper presents a new framework combining Bayesian Convolutional Neural Networks (CNNs) with active learning, which uses model uncertainty to select the most informative samples, unlike traditional methods.
2. It evaluates various acquisition functions, including BALD, Variation Ratios, Max Entropy, and Mean STD, specifically for Bayesian CNNs, providing a detailed analysis of their effectiveness.
3. The study compares Bayesian CNNs to deterministic CNNs, highlighting how accounting for model uncertainty can improve performance in active learning.
4. The proposed method is tested against existing techniques like kernel-based and semi-supervised approaches, showing that it can be as effective or better with fewer labeled samples.
5. The method is applied to a real-world problem, skin cancer diagnosis, demonstrating its practical use in a critical field where labeling is expensive.
6. The paper includes extensive experiments, testing various acquisition functions and models, and comparing them with state-of-the-art methods.

Downsides

1. The methods used are pretty computationally expensive and time-consuming, which might make it tough to use on large datasets or in practical situations where resources are limited.
2. Relying on Bayesian CNNs adds a lot of complexity and might not be practical compared to simpler, deterministic models that are easier to work with.
3. The experiments were done with relatively small validation sets, which might not give a complete picture of how well the approach works on large or more complicated datasets.
4. There's a risk of overfitting to the specific acquisition functions used in the study, so other potentially better functions might be overlooked.
5. While the approach seems to work well for medical diagnosis, it's not clear how well it would perform in other fields or with more complex tasks.

III. [Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds]
Jordan T. Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, Alekh Agarwal

Motivation

This paper tackles the challenge of making active learning work well with deep neural networks. Active learning helps by picking the most useful examples to label, which is crucial when labeling data is expensive and time-consuming. However, applying active learning to deep neural networks is tricky because these models are very complex and computationally intensive. Most traditional active learning methods either focus on uncertainty or diversity but often struggle to be effective across different neural network types and batch sizes. Additionally, they usually need a lot of tuning, which isn't always practical. The paper aims to solve these issues by creating a method that combines both uncertainty and diversity in a way that's adaptable and doesn't require extensive manual adjustments.

Teja Sree Parasa September 8th, 2024

CSE 584: Machine Learning Homework 1

Solution/Approach

The paper introduces a new method called Batch Active Learning by Diverse Gradient Embeddings (BADGE) to address these challenges. BADGE selects batches of data points based on their gradient embeddings, which reflect both the uncertainty and diversity of the samples. It calculates the gradient embeddings for each sample and then uses the k-MEANS++ algorithm to choose a diverse set of high-magnitude gradient samples. This approach ensures that each batch contains a mix of informative and varied examples, which helps improve learning efficiency. By avoiding the need for manual hyperparameter tuning and being adaptable to different model architectures and batch sizes, BADGE offers a practical solution for active learning in real-world scenarios.

Novelties/Contributions

1. BADGE introduces a hybrid approach that combines uncertainty sampling with diversity sampling. By using gradient embeddings from the final layer, it captures both model uncertainty and the impact of labeling each example, enhancing the active learning process without needing manual hyperparameter tuning.
2. The algorithm employs k-MEANS++ for selecting diverse samples based on their gradient embeddings. This method strikes a balance between choosing examples with high gradient magnitudes (indicating high uncertainty) and ensuring diversity within the batch, offering an improvement over traditional methods that focus only on one aspect.
3. BADGE shows robust performance across various neural network architectures, batch sizes, and datasets, making it a versatile tool that performs well in different experimental settings.
4. The use of k-MEANS++ instead of the more computationally demanding k-DPP allows BADGE to efficiently select diverse samples, making it more practical for real-world applications where computational resources and time are limited.

Downsides

1. Although BADGE is designed to work across different architectures, its effectiveness can still vary based on the specifics of the neural network used. It may not perform as well with very different or non-standard model configurations.
2. Even with the efficiency of k-MEANS++, the overall computational burden can be significant for very large datasets or high-dimensional embeddings, potentially making it resource-intensive.
3. The reliance on gradient magnitudes to assess uncertainty may not always accurately reflect the true informativeness of a sample. Samples with small gradients might still hold valuable information that the model has not yet recognized.
4. BADGE is tailored for supervised learning scenarios and may not be directly applicable to unsupervised learning or reinforcement learning tasks, limiting its general applicability.