Paper Reading Template

AI-For-NLP Course Group

|  |  |  |
| --- | --- | --- |
| Date | 2019.01.18 | |
| Title | XGBoost: A Scalable Boosting System | |
| Author | Tianqi Chen, Carlos Guestrin | |
| Question/Task | Description | You Answer |
| 1. Classify this paper | Is this paper pragmatic or theoretical?  Is this paper on science or engineering? | Pragmatic  Both science and engineering |
| 2. Brief Summary | Using the as short as possible to summarize the paper content. | The authors proposed a novel machine learning system for tree boosting, with a sparsity-aware algorithm for sparse data and weighted quantile sketch. Also provide insights on cache access patterns, data compression  and sharding methods. |
| 3. Outline | Outlining the content as multiply parts.  For example, for one paper, you may outline the content as following:  1. Background  2. The other Researcher’s method  …  8. Future Planning  And explain how does these outlines work together to make this article completed. | 1. Background and the great results XGBoost achieved. And a simple description of the paper.  2. Describe details of XGBoost, such as regularized learning objective, gradient tree boost model, two methods shrinkage and column subsampling to avoid overfitting.  3. Introduce the split finding algorithm, exact greedy algotithm and weighted quantile sketch, and sparsity-awre split algorithm.  4. The tree boost learning system design. Using column block for parallel learning, cache-aware access for speed, out-of-core design for saving space consumption.  5. Related works. Such as implement gradient tree boosting, propose a regularized model to prevent overfitting. Other work from greedy forest and sampling technique from random forest, etc.  6. Evaluations. Introduced the dataset and setup. Describes the results on classification and learning to rank problem. And some out-of-core experiment and distributed experiments.  7. Conclusion. |
| 4. Mainly Issue | What is the issue that author want to solve? | A novel tree boosting system with high scalability, performance and accuracy. |
| 5. Find the difficult or important words. | Find what words you are not understood and explain it by yourself.  Find important words in this article. | Weighted quantile sketch  加权分位点概述 |
| 6. Find the difficult sentences confusing you and explain what they mean. | 找出文中你不太懂的句子，试着解释他，最好用另外一种解释方法解释。 不要玩文字游戏。  例如， 《纯理性批判》里有一句话“除了实际存在的事物”，没有任何东西能发生作业。 如果你解释成“如果某个东西不存在，那么它就不能发生作用“，这就属于玩文字游戏。比较合理的解释一个例子是只要可能会下的雨，青草是不会生长的“或者只要可能有的存款，一个人的账号是不可能增加的“。 |  |
| 7. Find the main sentences author written. | Find out sentences which could express the intention of author mostly. | We propose a novel sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning. More importantly, we provide insights on cache access patterns, data compression and sharding to build a scalable tree boosting system. XGBoost scales beyond billions of examples using far fewer resources than existing systems. |
| 8. What have been solved and what not have been solved? | What problems or issues the author have solved?  What problems or issues the author haven’t solved? | Scalable tree boosting system. |
| 9. Rethink of the paper | Can you explain the paper main content to others?  Can you explain the paper to your wife/husband?  Can you explain the paper to a kindergarten pupil?  This answer ***cannot be Yes/No simply.*** *Please write the explanation with integrity.* | Explain the simple decision example in the article. A series of weak models forms a strong model. Measure the accuracy of the model using the regularized objective in the article. Formula derivation using taylor expansion. Algorithms to find the split point. System implementatoion. |
| 10. Which parts do you agree with the author?  Why do you agree with these? | Find out the opinions of author that you agree with.  Give the reason why do you agree with. | All, it’s an article about technique and based on results. |
| 11. Which parts do you not agree with the author? | List the parts or opinions that you do not agree with author. | None |
| 12. Why do you not agree with? | Classify each answer of question 11 as following types:  1. uninformed: 信息不足，必要的信息没有给到；  2. Misinformed: 论点与实事相反或不切合；  3. Logic Error; 逻辑错误，例如马基雅维的《君主论》里边：  所有的政府，不论新或旧，主要的维持基础在法律，如果这个政府没有很好的武装力量，就不会有良好的法律，也就是说，只要政府有很好的武装力量，就会有好的法律。  里边的逻辑错误在于“ 政府有很好的武装力量“ 应该是”有好的法律的“ 必要不重复条件，依照所述的逻辑，如果有好的法律，那么肯定有好的”武装“，但是有好的”武装“并不一定有好的法律。  4. Uncompleted Analysis; | None |
| 13. Is this article helpful to you?  How can you use these knowledges in your life or in future? | 简述这篇文章是否对你有用，对你以后哪些场景下回使用到？ | Absolutely useful.  Most recently, text classification problem. |