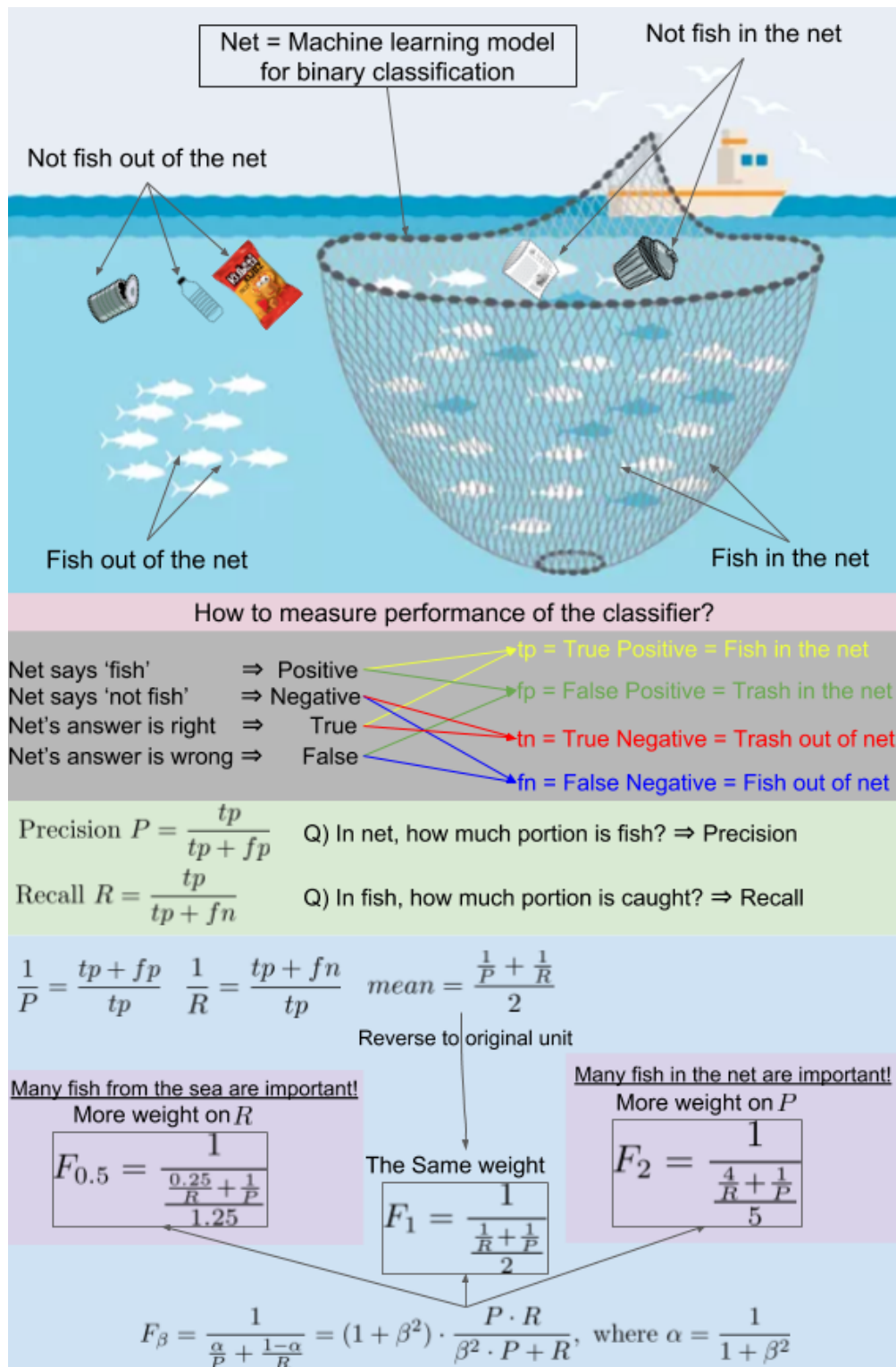


DBlog

laftworld

November 19, 2018

Performance Measure in Binary Classification Problem



Script

Hi, welcome to one page machine learning.

In this video, you'll learn how the performance measure for binary classification works. The formulas look scary, but they are just to say something for the concepts. All you need to do is listen carefully what the formula says. So take it easy.

When you look at the very top of the figure, someone might call it fishing in the sea, or somebody call it a binary classification problem. You'll see a fishing boat. If you are the captain of the boat, probably you wish to get a lot of fish as possible.

So you prepared a huge net. And you put it under the sea to catch fish. As a result, some fish and trashes are caught in the net. Now, how can we measure the performance of the net? This is very important because the amount of money you earn is highly dependent on its performance.

Naively speaking, the performance of net depends on how many fish are inside of the net. It's simple, isn't it? The more fish in the net, the better the net is. All you do for it is just count the number of fishes in the net.

But there is a big flaw in this measure, what if the net contains not only a lot of fishes but a lot of trashes in the sea? Well, that makes a problem if the amount of trash is not negligible. So let's redefine the performance. What you want is to catch a lot of fishes, while excluding trashes as much as possible.

Say that again: the net's performance is how well it includes a lot of fish as possible, while excluding trashes as much as possible.

Now we describe the situation with view of the net. When the net says it is a fish, then we call it 'POSITIVE'. When the net says it is not a fish, we call it 'NEGATIVE'. If the net's answer is right, we call it 'TRUE'. If the net's answer is wrong, we call it 'FALSE'. Then we have four states in total.

The fish in the net is called 'TRUE POSITIVE', which means that the net says it is fish, and the answer is right. The trash in the net is called 'FALSE POSITIVE', which means that the net says it is fish, and the answer is wrong. The trash out of the net is called 'TRUE NEGATIVE', which means that net says it is not fish, and the answer is True. Finally, the fish out of net is called 'FALSE NEGATIVE' which means that the net says it is not fish, but the answer is wrong.

Now it's time to pause this movie clip, and say that again in your own term. Remember, practice makes perfect.

pause...

3, 2, 1, Welcome back. The TP, FP, TN, FN are just a title to call the set of each object. Now we can define a measure which says the net's performance. They are called Precision, recall, and Fscore. Precision P is true positive over true positive plus false positive. See the true POSITIVE and false POSITIVE. Since 'POSITIVE' means the net says yes it is fish, now we restrict our view to things just inside of net. Then how much portion are fish in the net? if the portion is high, we can say that the net's performance is good. For high precision, we need high 'TRUE POSITIVE' and low 'FALSE POSITIVE', 'FALSE POSITIVE' means the number of trash in the net.

Now we look at recall R. The denominator is True Positive plus False Negative, which means that now we consider the whole fish in the sea. For all the fish in the sea, how much portion is inside of the net? this is what recall can answer us.

In some cases, precision can get high but recall gets low. In another cases precision is low while recall is high. Neither cases are wanted.

Thus we have F score. The F score is defined a harmonic mean of precision and recall. 'mean' says the balanced point

between precision and recall. But this is harmonic mean, rather arithmetic mean. You see that both precision and recall has TRUE POSITIVE as their numerator in common? So we can use arithmetic mean with inversed precision and inversed recall. That is why we use harmonic mean, to make their denominators as the same. After having arithmetic mean with these things, F score is derived by inverting it again. This is to return back to the original unit.

Now we see that F1 score is derived by assigning the same weight to both precision and recall. However, in some special situations, precision is much important than recall, or recall is much important than precision. For example, If we feel that having many portion of fish in the net is important for some economical reason, then we weight more on precision. On the other hand, if we feel that catching many fish from the sea is much important, then we weight more on recall.

The scary formula at the very bottom of the figure is for somebody who feels curiosity with it. Try deriving several types of F score.

This is the end of the video. If the concept is not familiar with you, you need to pause video and slowly follow the steps top to bottom. Don't forget, practice makes perfect. See you next time.

APPENDIX

DBlog

디블로그(DBlog)는 분산형 블로그(Distributed Blog)의 줄임말입니다. 여러 편의 문서를 웹에서는 내용을 볼 수 없고 다운로드만 가능한 형태로 배포하고 있습니다. 일종의 실험이지요.

여기에는 몇 가지 이유가 있는데요. 기존의 블로그 서비스를 이용하면서 아무래도 몇몇 점이 저에게는 잘 맞지 않는다는 생각이 들었습니다. 첫째로 저의 창작물 세트가 사기업에 종속된다는 점, 따라서 기업의 서비스 품질이 그대로 저에게 영향을 미친다는 것이 마음에 걸렸습니다. 여러 해 동안 블로그를 운영하면서 국내 블로그 서비스들의 불편한 UI와, 불필요한 광고, 서비스받는 개인 블로그에 기술적인 피해가 가는 경우에 그것을 개인의 치부로 덮는 행태를 보면서, 타사의 블로그에 종속된 채로 있기에는 어딘가 좀 불편한 구석이 있었습니다.

두 번째 이유는, 저는 페이스북에 글을 자주 올리는 편인데, 대부분 일상의 가벼운 내용 중에 어떤 것은 따로 모아서 보관을 해야겠다는 생각이 드는 글도 가끔은 있었습니다. 그런데 페이스북의 즉흥적인 특성이 글을 관리한다는 측면에서는 꽤나 불편했지요. 포스팅은 대부분 유효기간이 하루를 넘기지 않으니 이것을 읽는 독자분들 입장에서도 저의 블로그나 페이스북에 글이 많아질수록 어디서 무엇이 어떻게 연관되는지를 찾는데 혼란스럽기 그지없고, 차라리 검색을 포기해버리는 사례도 있을 것으로 보았기 때문입니다. 페이스북으로는 긴 글을 읽기가 부담스럽다는 단점도 크게 작용했다고 생각했습니다.

그래서 제가 작성하는 문서를 원하는 분이 다운로드를 받을 수 있도록 만들어보면 어떨까 하는 생각을 하게 되었습니다. 글이란 쓰는 사람과 읽는 사람이 시간차를 두고 소통하는 수단인데, 읽는 분들 입장에서도 각자의 서로 다른 필요에 따라서 개별 문서를 손쉽게 보유한다면 좋지 않을까 생각했습니다.

디블로그에는 차례도 목차도 없습니다. 모든 글은 단편으로 작성할 생각이고, 군더더기를 모두 제외하고 오직 문서의 내용으로만 접근하도록 시도합니다. 이것을 읽는 분들이 이 정도 크기의 문서를 저장할 공간이 없어서 불편을 겪지는 않으실 듯 하고, 또 이 문서를 여러분의 컴퓨터에 보유함으로써 느끼는 일종의 충족감도 있으실 테고요. 제가 만든 문서에 대한 피드백을 받는 공간은 개인 페이스북 페이지[Link] 이면 충분하다고 생각했습니다.

디블로그는 오직 저자만이 전체 문서를 보유하고 있습니다. 어느정도 분량이 쌓일 때마다 페이스북에 공개를 하고, 또 그중에 독자 여러분이 읽은 적 없는 문서를 골라가면서 읽는 즉흥적인 재미도 있지 않을까 기대해 봅니다.

컨텐츠

디블로그는 1)배움 (learning)과 2)머신러닝 (machine learning)에 대한 내용을 담고 있습니다.

배움 편에서는 우리가 살면서 무심경하게, 또는 습관적으로 사용해온 여러 단어들을 세심하게 살펴봅니다. 예컨대 주의, 집중, 의지, 노력, 성취 와 같은 것들이죠. 우리는 어릴 적 세월의 대부분을 학교에서 보내면서도 정작 배운다는게 무엇을 어떻게 하는 것인지에는 거의 주의를 기울이지 않고 살아왔습니다. 학교는 공인된 교육기관이니 가만히 출석만 해도 알아서 뭔가 되겠지 라는 안도감에 녹아들었다고 할까요. 그러나 사실은, 우리가 상식적으로 알고 있는 배움의 방식과 전혀 다른 방식의 배움이 세상에는 있으며, 여러분도 저도 어릴 적에는 그것을 알았지만 나이가 들면서 점차 잊게 된 것라고 저는 생각을 합니다. 그것을 다시 떠올려 우리 삶에 적용할 수 있다면, 저 자신이나 혹은 다른 누군가의 삶을 더욱 윤택하게 가꾸는 것이 가능할 거라는 희망도 가지고 있습니다. 배운다는 그것이 대체 어떤 형상을 하고 있는지, 장님 코끼리 만지듯 이곳 저곳을 더듬어보면서 그 느낌에 공감하고 소통하는 것이 이 문서를 작성하는 목적입니다.

머신러닝 편에서는 머신러닝의 다양한 주제들을 추려서 한 페이지의 그림을 곁들여 설명합니다. 머신러닝이 큰 이슈가 되면서 단어를 들어본 분들은 많으실 텐데, 대체로 들어는 봤으나 그게 무엇인지는 잘 모르겠는 정도로 이해하실 거라고 봅니다. ‘4차산업혁명’이라는 애매한 단어와 함께 가면서 그런 애매함이 그저 애매함으로만 지속되는 것을 보는데요. 디블로그에서 담는 머신러닝 컨텐츠는, 머신러닝이 ABC인 줄 알았는데 이거 알고보니 ABD였네? 라는 느낌을 받으실 수 있도록 구성하려고 합니다. 작성된 문서의 일부 한국어 버전은 오픈 튜토리얼스 [Link] 에서 보실 수 있습니다. 그러나 이 분야에 발을 담그면서 한국어의 울타리에 갇혀서는 앞을 헤쳐나가기가 매우 어렵습니다. 그래서 가급적이면 짧은 맥락의 영어와 짧은 길이의 수식으로 작성할 생각입니다. 수식과 더불어 수식을 작성하는 Latex에 대한 설명도 함께 시도할 것이고요. 진도나갈 것에 대한 급함이 없으니 느긋이 함께 해 나가면 좋겠습니다.

DBlog?

This document is a part of DBlog (Distributed Blog) which contains these contents.

- machine learning (English)
 - One-page explanation on general concepts in machine learning.
 - Exercise problems with Latex commands.
 - Research topics in Grammatical Error Correction.
- Issues on Education (Korean)
 - One-page explanation on how we (people) learn.
- Experience (Korean)
 - What have I done and How I am living (because this is a personal blog!).

Roughly speaking, DBlog contains two major topics, **Education** and **Machine Learning**. These two topics has been my top interest for several years so I feel that I have some stories to tell in public. Since machine learning is big and common field, I am going to write machine learning things and education things in English and in Korean, respectively.

In DBlog, Only the author has the entire set of documents. The only way you can read the contents is downloading each file. If you feel some articles are good, then you can have it or re-distribute it without any permission from the author.

Profile

HONG Won Eui (홍원의, 洪原義)

I am a PhD student in CSE (Computer Science and Engineering) in HKUST (Hong Kong University of Science and Technology). Research interest is machine learning in education. In particular, Neural Grammatical Error Correction for language learners.

These are links in which you can find me and press 'like'.

- [Official Website]
- [Github]
- [Facebook]
- [Instagram]
- [LinkedIn]
- [Naver Blog (Deprecated)]
- [T Story Blog (Deprecated)]
- KakaoTalkID: laftworld
- WeChatID: learningengineer
- email: wehong@connect.ust.hk
- [C.V. (Korean)]
- [C.V. (English)]
- [Open Tutorials (Korean)]