

Forum:[Hack Forums](#)

Thread: [Machine Learning Integration with HF](#)

+Haze - January 20, 2019, 11:56 am

This thread is aimed at Omni, but at the same time, it's also aimed at anyone else who could contribute to the discussion of a better forum.

I've taken a course on artificial intelligence for the past 6 months and it has completely changed my outlook of how we can use data to create artificially intelligent systems. Machine learning doesn't have this high entry barrier people think it does, anyone can play around with it, and for machine learning to be effective, you need lots of data and I know HF has a lot of that.

In terms of the forum, I wonder how the application of ML could be used to improve certain aspects of the forum? I don't necessarily mean the appearance or the features of the forum, but I think it can be used to reduce the amount of work done by an admin.

I haven't been very active on HF for the past 2 years so I am unsure of the current challenges faced by admins so I'm not sure where ML can be best applied, but when I was active I remember fraud being an issue and ML is good for things like that, and it's straightforward easy to implement and use.

Xerotic - January 20, 2019, 12:28 pm

We use Stripe as a payment processor and they already implement RADAR which is a machine learning tool to help prevent fraud. We don't have enough members actively posting about AI/ML yet to warrant a section IMO.

I myself have dabbled in machine learning, particularly different neural network architectures in a couple languages. I have thought of potential use cases for ML on HF but at this time we really don't have a strong need to develop a ML tool. Moderator work at this point doesn't require it. I have some fun ideas and some security ideas in mind but they aren't quite needed here I think.

I would like to see some more discussion about it though. It is a very important facet of the future (and really the present) of computing so I would be very happy if you would start some discussion about it. Start some threads. Write some tutorials. Get members interested and show them that starting ML work really isn't as intense as it seems. You don't need high-level mathematics knowledge to employ AI, you mainly need it to develop new algorithms and have a fine understanding of the theory of ML. Many libraries (in many languages - see Python primarily) provide easy access to developing ML tools.

Start some conversation about it and teach some members.

+Haze - January 20, 2019, 12:44 pm

(Jan 20, 2019 01:28 PM)Xerotic Wrote: [\[CITE\]](#) x

Quote:We use Stripe as a payment processor and they already implement RADAR which is a machine learning tool to help prevent fraud. We don't have enough members actively posting about AI/ML yet to warrant a section IMO.

Right so when I talked about fraud, I didn't mean payments, I meant more like fishy accounts. I remember a while back Omni had an issue where there were a high number of bought accounts and

they were used to scam people. I was thinking how a neural network could be used to identify high-risk accounts and report them to admin who can manually inspect them? Or perhaps a neural network that can automatically report posts or threads. I got the idea from a game called LoL where they use ML to identify people who are cheating, and they found that the players reported by the AI had a conviction rate of 85% versus when other players reported each other the conviction rate was like 20%. I thought something similar could be applied on HF, but I don't really have an up to date understanding of this forum because of my long period of inactivity, so there may not be a need for an ML tool right away.

Quote:I would like to see some more discussion about it though. It is a very important facet of the future (and really the present) of computing so I would be very happy if you would start some discussion about it. Start some threads. Write some tutorials. Get members interested and show them that starting ML work really isn't as intense as it seems. You don't need high-level mathematics knowledge to employ AI, you mainly need it to develop new algorithms and have a fine understanding of the theory of ML. Many libraries (in many languages - see Python primarily) provide easy access to developing ML tools.

Start some conversation about it and teach some members.

I would be keen on starting a discussion on machine learning on here, and I think it's something that could change the interaction of technology here. I wanted to start a few tutorials or projects on here on ML, I wanted to show how neural networks can be used to classification and how genetic algorithms can be applied.

Dragosthethird - January 20, 2019, 12:52 pm

(Jan 20, 2019 01:44 PM)+Haze Wrote: [\[CITE\]](#) Right so when I talked about fraud, I didn't mean payments, I meant more like fishy accounts. I remember a while back Omni had an issue where there were a high number of bought accounts and they were used to scam people. I was thinking how a neural network could be used to identify high-risk accounts and report them to admin who can manually inspect them? Or perhaps a neural network that can automatically report posts or threads. I got the idea from a game called LoL where they use ML to identify people who are cheating, and they found that the players reported by the AI had a conviction rate of 85% versus when other players reported each other the conviction rate was like 20%. I thought something similar could be applied on HF, but I don't really have an up to date understanding of this forum because of my long period of inactivity, so there may not be a need for an ML tool right away.

You are assuming here that there aren't already alerts in place. I'm sure Omni gets alerts about certain things like big changes in IP addresses. (Different countries/states) We also have trust scan and im sure certain things in the trust scan so more details to Admin and Staff and may even send alerts in extreme cases.

+Haze - January 20, 2019, 1:05 pm

(Jan 20, 2019 01:52 PM)Dragosthethird Wrote: [\[CITE\]](#) You are assuming here that there aren't already alerts in place. I'm sure Omni gets alerts about certain things like big changes in IP addresses. (Different countries/states) We also have trust scan and im sure certain things in the trust scan so more details to Admin and Staff and may even send alerts in extreme cases.

That's my point exactly, I'm sure those alerts are in place, but those alerts are basic in the world of machine learning. Those alerts are just basic checks, machine learning takes it to a whole another level, for example, all those data inputs being collected can be used to predict outcomes of certain events/actions.

The whole point of machine learning is that the system learns for itself, I can imagine with things like the Trust scan, there is probably a condition statement that goes along the lines of

```
Code if (account.age > 1000) {  
    account.status = good;  
}
```

But that is a very one dimensional way of looking at the problem, a machine learning algorithm can work in multi-dimensional space that our brain cannot comprehend and better predict than we can.

Omniscient - January 20, 2019, 1:16 pm

I have triggers for things but not necessarily ML. In my view something like ML would be useful at the http service level (nginx or apache) for DDoS prevention.

It could be used for other things too on HF itself but we're not at a place to really be deeply involved in coding such features. And I think it anything it would be more up to the MyBB team to develop smarter ML features. I know they won't ever do it though, just saying it should be them that integrates those types of options.

Nice suggestion thread.

+Haze - January 20, 2019, 1:20 pm

(Jan 20, 2019 02:16 PM)Omniscient Wrote: [\[CITE\]](#) I have triggers for things but not necessarily ML. In my view something like ML would be useful at the http service level (nginx or apache) for DDoS prevention.

It could be used for other things too on HF itself but we're not at a place to really be deeply involved in coding such features. And I think it anything it would be more up to the MyBB team to develop smarter ML features. I know they won't ever do it though, just saying it should be them that integrates those types of options.

Nice suggestion thread.

I just thought of something. I have been posting for a bit here and I have noticed that sometimes I am given a captcha to prove that I am not a bot. I assume there must be something I am doing that is causing the system to think I am suspicious. Now, we all know I am not a bot, and it's very annoying and frustrating to do these captchas every now and then for no reason. Providing an innocent user with a captcha every now is annoying and can really affect their user experience and that can lead to me not wanting to use the forum as much (which is not what any website owner would want). And for someone like me, I post a lot of code as I am a programmer, so I would always get flagged and when I am presented with a captcha, I almost always have to do like 5-6 before I pass.

In this scenario, a neural network can easily classify if I am a threat or not and get better over time. I think this can help reduce the number of captchas being served to members who are innocent and it's a better approach than using static threat analysis.

If the current threat analysis is something along the lines of

```
Code if (user post contains suspicious text) {  
    show captcha challenge  
}
```

```
if (user IP has changed) {  
    show captcha challenge  
}
```

then an ML model that can learn over time would be a better approach. The thing is, it wouldn't be difficult to implement either, you just need enough data. I don't know how MyBB works, but I don't think you need them to add this quality of life changes.

This is just an example of one of the many use cases. I think Microsoft said that ML is going to be the future of tech, and businesses that are not already integrating it with their services will lose out. At first, I thought they were just talking rubbish, but I now understand what they mean.

Athr - January 20, 2019, 1:22 pm

Interesting thread.

The easiest usecase of ML I can think of would be reporting LQ posts with text like "25chars" or excessive "..." and its variants.

+Haze - January 20, 2019, 1:32 pm

(Jan 20, 2019 02:28 PM)xadamxk Wrote: [\[CITE\]](#) Why not make tutorials about it if you think it has potential ?

I haven't been that active for the past 2 years so I didn't really know the forum as well so I didn't know how much of ML exists on HF.

Plus tutorials aren't really my thing, I don't really like write-ups, I do enough of it at uni and I hate it. I am a programmer and I prefer to show things by implementing it.

(Jan 20, 2019 02:22 PM)Athr Wrote: [\[CITE\]](#) Interesting thread.

The easiest usecase of ML I can think of would be reporting LQ posts with text like "25chars" or excessive "..." and its variants.

I used to run a freelancing service where I'd offer my programming services on HF. I had multiple ways to get customers/clients, and one of them was to go through threads in the buyers bay, service requests page etc.

A few years ago, I wrote a JavaScript userscript that would check through those forums and look for certain keywords like "C#", coding, programming, code" but the problem is that that list can go on and on, and users vocabulary can evolve over time. Yesterday I thought to myself how I could improve that userscript, and I thought if I had a neural network to analyse threads and flag them to me on which one it thinks I may be interested in, then that could in theory statistically improve the number of clients I got any money. The NN can learn over time, what type of threads I usually respond to and I am willing to do, therefore saving me the effort and gives me time to focus on more important aspects of my service.

I know this will work because this is the same technology that is used in SPAM for emails.

+Haze - January 20, 2019, 2:28 pm

(Jan 20, 2019 03:27 PM)xadamxk Wrote: [\[CITE\]](#) Then make a project for the community to learn from.

Open source it and document it.

Yes, that's the plan for the near future.

Elitas - January 20, 2019, 2:32 pm

(Jan 20, 2019 02:20 PM)+Haze Wrote: [\[CITE\]](#) I just thought of something. I have been posting for a bit here and I have noticed that sometimes I am given a captcha to prove that I am not a bot. I assume there must be something I am doing that is causing the system to think I am suspicious. Now, we all know I am not a bot, and it's very annoying and frustrating to do these captchas every now and then for no reason. Providing an innocent user with a captcha every now is annoying and can really affect their user experience and that can lead to me not wanting to use the forum as much (which is not what any website owner would want). And for someone like me, I post a lot of code as I am a programmer, so I would always get flagged and when I am presented with a captcha, I almost always have to do like 5-6 before I pass.

In this scenario, a neural network can easily classify if I am a threat or not and get better over time. I think this can help reduce the number of captchas being served to members who are innocent and it's a better approach than using static threat analysis.

```
If the current threat analysis is something along the lines of
Code if (user post contains suspicious text) {
    show captcha challenge
}
```

```
if (user IP has changed) {
    show captcha challenge
}
```

then an ML model that can learn over time would be a better approach. The thing is, it wouldn't be difficult to implement either, you just need enough data. I don't know how MyBB works, but I don't think you need them to add this quality of life changes.

This is just an example of one of the many use cases. I think Microsoft said that ML is going to be the future of tech, and businesses that are not already integrating it with their services will lose out. At first, I thought they were just talking rubbish, but I now understand what they mean.

I believe that captchas are served at a cloudflare level. Not sure if CF would allow such customizations, but maybe they do.

+Haze - January 20, 2019, 2:37 pm

(Jan 20, 2019 03:32 PM)Elitas Wrote: [\[CITE\]](#) I believe that captchas are served at a cloudflare level. Not sure if CF would allow such customizations, but maybe they do.

Oh okay, right, so maybe Cloudflare judgement doesn't take into account factors from my profile, it only sees me as a network request.

Elitas - January 20, 2019, 2:51 pm

(Jan 20, 2019 03:37 PM)+Haze Wrote: [\[CITE\]](#) Oh okay, right, so maybe Cloudflare judgement doesn't take into account factors from my profile, it only sees me as a network request.

Yes, indeed. It will see that you have hit the firewall by mentioning specific things.

For example: /etc/passwd

javaPro - January 20, 2019, 2:55 pm

I actually work in the Data Science field, machine learning is very powerful. For instance I made a holiday recommendation engine. You were able to freely type what type of holiday you wanted e.g. I want to go somewhere sunny that has activities for the kids, near a beach etc. I would then use a mixture of techniques including Natural Language Processing to understand and process the data, and display recommendations based on what they wanted.

Doing something like spam detection is pretty simple, you can use existing databases to recognise the patterns of spam and then grade them accordingly. There are many things you could do with machine learning.

If this was my website however, I would initially use it to assist in the making of business decisions. I assume analytics is used, but this only provides analytical data. Using historical data, you could produce much more meaningful insights. For instance, you could produce a Sankey diagram to demonstrate the month to month flow between new user, churned, active and passive. You can then implement a new feature, competition, etc and see the impact this has on these different groups.

Historical data is so powerful, that is why the industry has exploded over the past few years.

+Haze - January 20, 2019, 2:59 pm

(Jan 20, 2019 03:55 PM)javaPro Wrote: [\[CITE\]](#) If this was my website however, I would initially use it to assist in the making of business decisions. I assume analytics is used, but this only provides analytical data. Using historical data, you could produce much more meaningful insights. For instance, you could produce a Sankey diagram to demonstrate the month to month flow between new user, churned, active and passive. You can then implement a new feature, competition, etc and see the impact this has on these different groups.

Historical data is so powerful, that is why the industry has exploded over the past few years.

This is exactly what I am trying to convey, I am just not that good with words sometimes. The use of analytics to aid business decisions which can have a big impact on the forum in a positive way.