Paper writeup notes

|  |
| --- |
| Introduction:  Background stuff about open clusters and previous work that has looked at them e.g. Bossini, D. Stello  Potentially talk about the clusters we chose and why they are important and useful/good for study.  Short description about what the paper did.  Put papers into the introduction that we compare our final results to. Hin says maybe talk about 4 papers (but I guess cite more)  Why are clusters useful and thus why are we bothering to study them.  Potentially give brief outline of previous work with neural networks and HBMs e.g. Verma and Hendriks  Maybe find a paper which has used open clusters to do stellar/galactic evolution. (galactic archeology with open clusters)  End of introduction outlines what the following sections will contain, so people no where to go to get the information they want. |
| Neural networks  Background of neural networks  What we did with them and the setup like training inputs and outputs input fundamentals: Age, Mass, Fe/H, Initial Y, MLT  output observables: Teff, radius, delnu  we don’t actually use delnu and we calculate luminosity using radius and temperature through the Stefan-Boltzmann law, and then we determine the error on luminosity by comparing the neural net prediction to the grid  Do not go into too much detail!!! Pass explanation to other papers to save words (eg. back propagation in verma 2016)  We might need to mention that there could be some bias in our grid model because we our only using the initial helium of the star which obviously doesn’t change across the evolution of the star and the neural net training is cut before core helium burning  This is because if MESA doesn’t account for changing helium and calculates Teffs etc. using the initial helium then our predictions may not be very good but if MESA is then we are doing good science and not only that we are also fitting to the fraction of helium when the cluster formed which could be useful for galactic archaeology.  How we trained them  How we chose the architecture  Training methods?  Problems we faced i.e. RGB problem and inbetween track issue of overfitting  How did we try to solve the RGB issue  From that the results we got from D. Stello on just Teff and Delnu  So I guess we can give the results at each stage of the process and how those results determined how we changed our methods.  We need to write down and explain the function used in the neural networks  Figure out architecture and then increasing the regularization  Discuss how the architecture allows for more flexibility as it has more numbers and more numbers allows you to approximate shapes.  Dropout didn’t work  Nadam was good, SGD didn’t work for us  Elu performed better than Relu  We started at 64x6 but consult google sheet to make sure  Batch normalisation improved speed and loss  To high a regularisation damps the process and too low a regularisation allows overfitting.  Show where the neural network does well and where it has limitations.  We can see from the graphs if it’s overfitting between tracks rather than between points along a track.  We should discuss what went wrong with our work if it is helpful and what we did to solve those problems  One the possible problems with our method is that the resolution of the grid is mostly in age variation meaning a lot of the NN flexibility goes into approximating the age which causes some what of a bias in age compared to the other parameters. The [Fe/H] steps are too large and we feel that it’s easier to train them if the steps were finer. We could have achieved this by using data augmentation (,Gaussian processes?), which may have helped our results. Doing data augmentation is much faster than say running more MESA tracks which we may have to talk about how that has issues relating to using particular models in order to generate the data.  Teff scaling, divided by 5000 to reduce the dynamical range…  We use radius instead of luminosity because radius also has a smaller dynamical range  (probably don’t need to talk about radius scaling)  Luminosity changes by a couple of orders of magnitude which is a much greater than the change of magnitude across the Teff range which is possibly why the NNs have so much trouble with the RGB.  The red bump/RGB bump (look up what it’s called), the red bump messes things up and that’s one of the reasons why we removed the RGB from our training, though we ekpt the lower part of the RGB so that the NN would be better at approximating the dip in the Subgiant.  It’s important to remember that how well the neural net performs is based on how well it predicts the points in the data, which gives no bearing on how well it’s off track predictive skill is.  I guess I should talk about how we used batch sizes.  Regularization limits the amount of change that can be made to the weight and biases.  We can always cite someone who talks about how regularization works.  Explain back propagation stuff?  How we cut the RGB data from the grid  Past studies on the topic of using ML on stellar evolution  o Bellinger 2016: random forest and backward modelling with asteroseismic and spectroscopic observables  o Verma 2016: NN to model low mass MS stars, backward modelling with asterseismic observables  o Hendriks & Aerts 2018: NN forward modelling with a genetic algorithm to backwards sample fundamentals. Uses actual mode frequencies as observable output. MS + subgiant + RGB, no rotational stars |
| HBM  Helped inform us that the RGB was still bad as it wasn’t able to sample data properly  We used D. Bossini’s ages and other papers to inform our priors which I want to write as a table  Hyperpriors on age, Fe/H, initial Y, MLT  No hyperprior on mass, but it has a bounding box prior which differs between clusters because the box is to prevent the HBM sampling outside of the training region of the neural net and depending on the hyperpriors, that will effect what masses are allowed.  It is possible that we put on a prior on mass which is influenced by the current spread values in metallicity (i.e. what the current range on the metallicity prior is given the current mean metallicity hyperprior), which we would do because the metallicity is  HBM results: ‘bestfit’ isochrone, with 1sigma and 2sigma error regions  Where bestfit is the mean value of the fundamentals given by the HBM results  Although the shaded regions are potentially overestimates of 1sigma as they are made from beta functions rather than normal distributions. |
| Collecting data – we could put the detailed object selection in the appendix.  How did we link the IDs  How did we choose which stars  Cluster selection  Other papers and sources we used.  Specify that we have a range of ages and metallicities.  And specify the clusters in the Kepler field. |
| Stellar Grid  To give a proper introduction about the grid, you may use the following text or pick what you need. (Please rewrite with your language!)  Theoretical model grid  Stellar models and input physics    We used Modules for Experiments in Stellar Astrophysics  (\textsc{MESA}, version 12115) to establish a grid of stellar models.  \textsc{MESA} is an open-source stellar evolution package which is undergoing active development. Descriptions of input physics and numerical methods  can be found in \citet{2011ApJS..192....3P,2013ApJS..208....4P, 2015ApJS..220...15P}.  We adopted the solar chemical mixture [$(Z/X)\_{\odot}$ = 0.0181]  provided by \citet{2009ARA&A..47..481A}.  We used the \textsc{MESA} $\rho-T$ tables based on the 2005  update of OPAL EOS tables \citep{2002ApJ...576.1064R} and OPAL opacity  supplemented by low-temperature opacity \citep{2005ApJ...623..585F}.  The MESA ‘simple’ photosphere were used as the set of boundary conditions for modelling the atmosphere.  The mixing-length theory of convection was implemented, where  $\alpha\_{\rm MLT} = \ell\_{\rm MLT}/H\_p$ is the mixing-length parameter.  The \textsc{MESA} inlist used for the computation is available on \url{https://github.com/litanda/mesa\_inlist}.    The gird has four independent model inputs: stellar mass (M), initial helium fraction ($Y\_{\rm init}$), initial metallicity ([Fe/H]), and the mixing-length parameter ($\alpha\_{\rm MLT}$). Ranges and grid steps of the four model inputs are summarized in Table \ref{tab:grid}.  The initial chemical composition was calculated by:  \begin{equation}  \log (Z\_{\rm{init}}/X\_{\rm{init}}) = \log (Z/X)\_{\odot} + \rm{[Fe/H]}. \\  \end{equation}      Also, the calculations of observables (Teff and luminosity in our case) are based on a helium fraction that changes during the evolution of the star, correct? Instead of just the initial helium fraction.    The calculation of surface Teff and luminosity in theoretical model is based on a set of structural equations and all stellar parameters, not only the helium fraction. However, I do agree that the surface helium fraction has a big impact on the surface properties. If you are going to discuss this, the surface helium fraction of one evolutionary track are constant during main sequence (same as the initial value), because we do not consider the helium diffusion.      Finally, we are using the delta nu that is available for all stellar models, which means it is calculated directly from scaling relations, correct?    Yes!    Hope these are helpful. And let me know if there are any questions.    Good luck on your final report.    Tanda |
|  |
| We chose are clusters from Bossini using their ages  You can find which papers cited what paper on ADS |
| Section on our problems with the RGB  We can show graphs without it being in a results section |
| “as Table 1 shows, our assumptions for cluster ages are a collection of independent determinations, i.e. they were derived by different authors and using disparate models. Thus our dataset is not homogeneous as far as ages are concerned”  <https://www-aanda-org.ezproxye.bham.ac.uk/articles/aa/full/2005/41/aa3482-05/aa3482-05.html> |
| Selection bias of stars as papers have agendas so the stars they chose to study may not be representative especially when for each cluster we only have 3 or 4 papers.  Show that by changing the stars for NGC 6819 that are sampled you can change the convergence thus more stars isn’t necessarily better.  Leads into a weakness of HBMs |
| @article{Nadam,  title={Incorporating Nesterov Momentum into Adam},  author={Timothy Dozat},  journal={Proc. ICLR Workshop},  year={2016}  } |
| batch norm paper: <https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I/abstract>  @ARTICLE{batch\_norm,  author = {{Ioffe}, Sergey and {Szegedy}, Christian},  title = "{Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift}",  journal = {arXiv e-prints},  keywords = {Computer Science - Machine Learning},  year = 2015,  month = feb,  eid = {arXiv:1502.03167},  pages = {arXiv:1502.03167},  archivePrefix = {arXiv},  eprint = {1502.03167},  primaryClass = {cs.LG},  adsurl = {https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I},  adsnote = {Provided by the SAO/NASA Astrophysics Data System}  },  ada batch: <https://arxiv.org/abs/1712.02029>  @article{pymc3,  doi = {10.7717/peerj-cs.55},  url = {https://doi.org/10.7717/peerj-cs.55},  year = {2016},  month = {apr},  publisher = {{PeerJ}},  volume = {2},  pages = {e55},  author = {John Salvatier and Thomas V. Wiecki and Christopher Fonnesbeck},  title = {Probabilistic programming in Python using {PyMC}3},  journal = {{PeerJ} Computer Science}  },  @ARTICLE{NUTS,  author = {{Hoffman}, Matthew D. and {Gelman}, Andrew},  title = "{The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo}",  journal = {arXiv e-prints},  keywords = {Statistics - Computation, Computer Science - Machine Learning},  year = 2011,  month = nov,  eid = {arXiv:1111.4246},  pages = {arXiv:1111.4246},  archivePrefix = {arXiv},  eprint = {1111.4246},  primaryClass = {stat.CO},  adsurl = {https://ui.adsabs.harvard.edu/abs/2011arXiv1111.4246H},  adsnote = {Provided by the SAO/NASA Astrophysics Data System}  },  ted von hippel's bayesian paper:  https://arxiv.org/abs/1605.02810  HBMs on white dwarfs:  https://arxiv.org/pdf/1703.09164.pdf  https://academic.oup.com/mnras/article/480/1/1300/5056190 |

Acknowledgments at the end?

Maybe write the method and results/conclusion before the introduction then abstract  
assuming we have enough results by that point.

|  |
| --- |
| Is the introduction a more detailed version of the context and aims section of the abstract?  How was the training grid made?  Can I cite Hin’s paper for project work he is covering?  How do we cite work from the other person?  How should MLT be treated in the PGM?  Do we write both of our names under the title? If so both our student IDs?  Can you have citations in the abstract?  Go into details such as MAE? Do we say we’re using keras?  Can we just introduce open cluster stuff in the introduction and keep neural network introduction type stuff till a neural network section?  Question for Tanda:  Is helium treated as changing in the MESA tracks, when calculating Teff etc.?  Is GYRE involved? |
| GET THE RESULTS OF THE HBM from D. Stello on just Teff and Delnu  GET PGM FROM HIN  GET HYPERPRIORS FROM HIN  WHAT VALIDATION FRACTION OUR WE USING?  We have not considered rotating stars?  GET PLOT OF PRIORS FROM HIN  GET APPENDIX TARGET TABLE AND OTHER APPENDIX TABLES |