Review of: "Mining Double-Line Spectroscopic Candidates in the LAMOST Medium-Resolution Spectroscopic Survey Using a Human-AI Hybrid Method"

Guillem Anglada-Escudé[1]

1 Institute of Space Sciences, Barcelona, Spain

Potential competing interests: No potential competing interests to declare.

The research is of high quality. Both the general astrophysical introduction and the methodology sections are very clear and well-structured. The results read well and are clearly presented. The work is well-framed within the state-of-the-art global efforts (good literature review, and cross-matching with other large area surveys such as Gaia/ESO, etc.). The article also makes a very nice point on addressing one of the main objections to using ML learning methodologies. That is, many researchers still perceive the algorithms as black boxes. The authors show that with a clear methodological approach where the training data and the algorithms are trained on well-known sets (synthetic in this case), this objection becomes meaningless and the full potential of ML methods can be exploited reliably. This is high-quality work that deserves publication.

I only have a few comments/suggestions.

The effect of Signal-to-Noise. When building the training dataset, the authors generated a good random grid on the parameters such as Teff, flux ratios, and RV differences. However, they didn't produce training samples with different signal-to-noise ratios (S/N). That is, training your algorithms with noisy data. The S/N is a very important feature of the 'data'. While the presented results are good, the fact that the S/N dimension is missing makes the final results a bit weaker.

The inclusion of the S/N as a dimension in the parameter space of the simulations would have enabled diagnostics such as precision/accuracy and other statistics as a function of S/N, so one could have an additional quality indicator in the final classification table. Also, this would mitigate false positives (multiple peaked CCFs) in the low S/N regime, as the algorithm would learn to assign lower credibility to classifications in that regime.

While re-running everything with the injection of noise in the training sample would be too much work (and the study contains many other virtues), I suggest adding a paragraph discussing the impact of signal-to-noise, and the benefits it would have, especially in the low S/N regime. I would also then add comments in the conclusions stating the importance of making training sets with synthetic noise injection as an important improvement for future work.

About future strategies to inject noise, note that it is not only a matter of adding Gaussian noise to the spectral elements. One needs to use some noise model (noise amplitude as a function of wavelengths, possibly downweighting regions affected by telluric absorption, and other non-trivial instrumental effects). At the end of the day, one would like to generate synthetic data elements as close as possible to those generated by the instrument. There are ML methods that can enable that. For example, using well-known and well-behaved single-lined stars to train a noise injection deep NN, among others.