MK CLASSIFICATION OF SPECTRA USING AN AUTOMATED CLASSIFICATION ALGORITHM

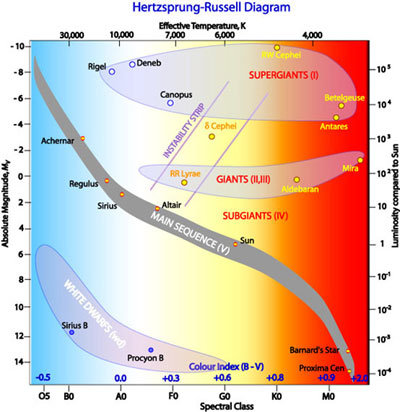
JULY 24, 2018

ROSIE FASULLO, DAVID WHELAN

ABSTRACT

This paper will present the current work being done to create an automated stellar classification algorithm. The program is based off the data reduction and spectral classification tool used at the Adams Observatory to aid in classification. Throughout the summer, optimizations were made to the data reduction algorithms that increased the users spectral typing efficiency by 57 %. The program can now perform automated wavelength calibration and will soon be able to perform automated continuum rectification. The automated spectral typing program will be completed in the next phase of this project during the fall semester of the 2018-2019 academic school year. Several methods of automated spectral classification are discussed including a discussion of the existing spectral classification algorithm MKCLASS by R. Gray and C. Corbally and the stellar label determining algorithm The Cannon by a research group at MIT.

1. INTRODUCTION

 Stars have been studied for years and throughout that study a need for a classification scheme arose. Stars are grouped by two different criteria, spectral type and luminosity class. The spectral type corresponds to the relative color of the light coming from a star. The color of the star is an indication of what the surface temperature of the star is; the bluer the color, the hotter the star (Gray and Corbally 2009). There are 7 different spectral types used by astronomers today and they are, from bluest to reddest, OBAFGKM (Gray and Corbally 2009). For reference, our sun is a spectral type G star, stars above this type, K and M stars, are referred to as late-type stars and stars that are below this type, O, B, A and F stars, are referred to as early- type stars (Gray and Corbally 2009). Within each alphabetic type there is a numeric classification ranging between 0 and 9, with 0 indicating a bluer star and 9 indicating a redder star within a given alphabetic type (Gray and Corbally 2009). Our sun has a spectral type of G2, meaning it is on the hotter end of a mid-range star (Gray and Corbally 2009). Luminosity class corresponds to the brightness of the star which is an indication of the size of the star (Gray and Corbally 2009). Stars range from dwarf stars with luminosity class V to supergiant stars with luminosity class Ia (Gray and Corbally 2009). Our sun is a luminosity class V star which is the most common luminosity class of the stars astronomers have observed (Gray and Corbally 2009). The distribution of spectral types and luminosity classes can be seen in what is called a Hertzsprung-Russell (HR) Diagram in Figure 1 below. The Adams Observatory at Austin College takes spectra of late O, B and early A type stars with a specific focus on B type stars.

*Figure 1: Hertzsprung-Russell Diagram. The Adams observatory currently takes spectra of early-type O, B and A stars which are found in the left portion of the diagram (“Hertzsprung-Russel Diagram”).*

In addition to spectral types and luminosity classes there is a distinction made between normal stars and peculiar stars. Peculiar stars are stars that have spectrum of a normal star but with a noticeable difference indicating a physical peculiarity about the star (Gray and Corbally 2009). The most obvious form of peculiarity is emission in the hydrogen lines. The hydrogen lines may be completely or only partially filled in by the emission. Emission normally indicates there is a shell or disk around the star that is emitting light (Gray and Corbally 2009). Another obvious peculiarity is unusual metallicity. This paper will focus on helium anomalies and general metallic anomalies. Many stars show unusually weak or unusually strong Helium lines, indicating a particular-abundance or lack thereof of Helium in the star (Gray and Corbally 2009). This can make these stars tricky to classify because many astronomers use helium lines as a critical part of their spectral typing process. Additionally, there may also be unusually strong absorption of other metallic lines, also indicating a particularly large abundance of that metal. Another common peculiarity is the spectrum of a binary star system. Depending on the inclination angle of the star system and the relative sizes and brightness’s of the stars, the binary nature of the system, may not be visible from the perspective of the earth. However, taking the spectrum of the system results in a composite spectrum (Gray and Corbally 2009). A composite spectrum is distinguished from a normal spectrum by the presence of characteristics of two separate spectral classifications in the same spectrum. Finally, one of the most difficult kinds of peculiarity to detect is called line broadening. The hydrogen lines appear broader and shorter than the spectral standard of that type, but the strength (or total area of the line) is the same (Gray and Corbally 2009). Line broadening indicates that the star is rotating at an unusually high rate, causing the spectrum to be slightly blurred (Gray and Corbally 2009).

There are several different types of spectral classification systems used to classify stars. One of the most commonly used systems and the one used in this project is the MK classification system developed by W. W. Morgan and P.C. Keenan (Gray and Corbally 2009). The MK classification system classifies stars by comparing the spectrum of the star in question to the spectra of a set of standard stars. The star in question is then classified as whatever spectral standard its spectrum most closely resembles (Gray and Corbally 2009). This comparison is done using only spectra and does not admit the use of any extra information in the determination of the classification of the star Gray and Corbally 2009). In this way, the MK classification system classifies stars according to their most natural groupings. The system does allow for the interpolation of spectral classifications if a star does not precisely match any spectra but appears to lie between two (Gray and Corbally 2009). For a given spectral type there may be multiple spectral standards of three different types: standards anchor points, primary standards and secondary standards (Gray and Corbally 2009). An anchor point is a spectral standard that has not changed classification since the development of the MK classification system began (Gray and Corbally 2009). Anchor points do not by any means cover the entire range of spectral standards and thus the need for primary standards (Gray and Corbally 2009). Primary standards fill in the gaps of the anchor points and provide a reference for the best-known spectrum for each spectral type and luminosity class (Gray and Corbally 2009). Secondary standards are the best-known spectra for stars that are accessible from both the southern and the northern hemisphere (Gray and Corbally 2009).

With the ideology of the MK classification system introduced, the specific methods people use to spectral type stars using the system can now be discussed. We will use ourselves as the reference for the sequence a person goes through in spectral typing a star. When WE begin spectral typing we first look at the broad shape of the spectrum. Depending on the shape of the continuum and the depth of the Hydrogen lines, we make a general guess about whether the star belongs in the O, B or A spectral types. Once we have chosen a suspected general type, we look at the depth of the hydrogen lines to determine if the star is more likely an early-type or a late-type star within the general type and compare the spectra to the spectral standards of the suspected type. At the same time, we also consider the width of the Hydrogen lines beginning to narrow down the luminosity class by comparing with the spectral standard of the suspected luminosity class. Once we believe we have found an approximate match, we look at the Helium lines and compare them to the spectral standard we believe the star fits to confirm my spectral classification. If the spectrum is not a strong match we to compare it to spectral standards around by believed spectra classification and determine if an interpolated spectral classification would be more accurate. This classification process is the process that will be emulated by the automated spectral typing algorithm.

Spectral typing is currently done almost exclusively by human beings. The MK Classification system is a very subjective system based the astronomer’s ideas of what is close enough and making decisions based on incomplete information. Although computers are good at many things, they are very bad at making accurate choices that require decisions to be made based off incomplete information. Thus, it is usually easier and more accurate for spectral typing to be done by humans. However, given the number of stars in the sky and the time it takes to carefully and accurately spectral type a star, there is a need for an accurate and efficient spectral typing program to be developed. In a survey of a thousand stars, what one astronomer or a team of astronomers could do over the course of several days, an accurate and reliable spectral typing program could do in a few minutes. In fact, there has been an attempt at an expert classification by R. Gray and J. Corbally already (Gray and Corbally 2014). They developed a program called MKCLASS that can classify normal stars with a spectral type accuracy within 0.6 of a spectral type and a luminosity class accuracy within 0.5 of a luminosity class (Gray and Corbally 2014). MKCLASS uses a weighted least squares comparison combined with a more detailed look at the metallic lines to determine the spectral type and luminosity class of a given spectrum (Gray and Corbally 2014). MKCLASS is very good at spectral typing stars that are normal or have only slight peculiarities of any spectral type and one can be reasonably confident in the results it produces (Gray and Corbally 2014). However, there are some serious drawbacks to the MKCLASS program. The first and most problematic drawback is that when a spectrum has a peculiarity that is not expected by the program it simply gives up at classifying the spectrum without an explanation as to why the classification failed (Gray and Corbally 2014). The second problematic drawback is that there is no way to numerically measure the expected accuracy of the spectral type of the program (Gray and Corbally 2014). The program provides a file that lists the quality of the classification, but the quality ranges from “Very poor” to “Very good”, which is not helpful for the astronomer in trying to determine an actual accuracy rating (Gray and Corbally 2014). Although there is not a numeric rating on the accuracy of the spectral type the program does provide information on the corresponding signal to noise ratio of the spectra of the star according the quality classification (Gary et al. 2016). While MKCLASS is a strong program that is useful for broad surveys of the sky that do not need extremely accurate classifications of stars, there is definite room for improvement.

Another program that is worth mentioning is called The Cannon and was written by a group of astronomers at MIT trying to find a way to determine spectral labels of large sets of spectra (Ness et al. 2015). It is important to mention that the term “spectral labels” refers to physical characteristics about the star (Ness et al. 2015). The Cannon uses a data-driven approach to determine spectral labels of spectra (Ness et al. 2015). This data-driven approach involves training a classifier on a reference set of spectra and then using the trained classifier to determine the spectral labels of the star (Ness et al. 2015). The classifier learns the characteristics of the spectra of stars with certain spectral labels and looks for those characteristics in the spectra of the stars with unknown spectral labels (Ness et al. 2015). This is a very computationally efficient method of classifying spectra because it does not involve storage of large lists of lines or predetermined line characteristics (Ness et al. 2015). The Cannon is very effective at determining the spectral labels of a large set of spectra with a high degree of accuracy (Ness et al. 2015). The main weakness of this approach is that it requires a sufficiently large data set to train on for it to provide accurate classifications (Ness et al. 2015). Although the program is not immediately applicable to determining spectral types, there is a strong potential for adapting The Cannon for use as a spectral typing program.

So far, stellar classification has been explained as well as the MK classification system and the current work being done on the automatic spectral typing program MKCLASS as well as an introduction to data-driven classifications of stars. In the following section, original spectral typing tool used by astronomers at the Adams Observatory will be introduced as well as the changes made over the course of the summer to improve the program. Next, statistics on the efficiency improvement of the program from the beginning of the program to the current version of the program will be presented in the Results section. Finally, a brief discussion of these statistics and a discussion of the future plans for the spectral typing algorithm and the program as a whole will be detailed in the Discussion section.

2. RESEARCH METHODS

At the beginning of the summer, spectral typing at Austin College was done using a program written by Dr. David Whelan called ao\_redux. The program is written in the programming language Python and was just a tool to aid the astronomers working at the Adams Observatory with classifying stars. The program began by automatically reducing images and then showing the uncalibrated spectrum of the first star in the set of stars to be spectral typed in that run. The user then manually selected the Hydrogen absorption lines for each star. Once the user had completed wavelength calibration for each star, they could begin continuum rectification. Continuum rectification involved the user selecting points representing each section of continuum between the absorption lines manually and solving for an approximate polynomial fit to the curve created by the points. Once the user had gotten the fit to be as smooth as possible, they could view the rectified spectrum. If the spectrum was not rectified to their satisfaction, they would have to undo the rectification and start over. Once all the spectra in the set had been rectified, the user could continue the spectral typing. During the spectral typing phase of the program, the current spectral standard would be plotted over the current spectrum in red so that the user could do a direct visual comparison of the spectra. Overall, the program was a bit cumbersome but reasonably effective as a tool used in spectral typing.

As the summer progressed, several changes were made in preparation for creating the automated spectral typing program in the fall. The original program was weak in several areas and throughout the summer these weaknesses were lessoned or eliminated altogether. The first weakness was the wavelength calibration process. When given only a few stars to analyze the process is not too cumbersome, but when given many spectra, on the order of 50 or greater, the process was extremely time consuming. The user had to zoom in to the tip of each hydrogen point, select the point of the line and then zoom out. To remedy this problem an automatic wavelength calibration program was developed. Using a simple measure of the direction of the slope from one point on the spectrum to another, the program was able to successfully locate 6 out of 6 absorption lines for simple spectra. Once the location of the absorption lines was found for one spectrum, the program then calculated the ratios between the absorption lines and used those as clues to look for absorption lines in all the other spectra. This increased the accuracy for the wavelength calibration algorithm significantly, and the algorithm now always picks Hydrogen absorption lines correctly for spectra that are not affected by emission in the hydrogen lines. The program now has the ability to find the Hydrogen lines automatically and greatly reduce the time the user has to spend doing menial work.

The second weakness was with the continuum rectification process. The continuum rectification process is very important for the accurate classification of the star and the original method of continuum rectification did not aid the user very much in producing the most accurate continuum rectification possible. However, by simply changing the layout of the continuum rectification window the user is able to create an accurately rectified continuum relatively quickly. The unrectified spectrum with the representative point and the continuum rectified spectrum are now displayed side by side, so the user can see how the changes they make affect the rectified spectrum in real time. Additionally, the program now displays a horizontal line on the continuum rectified plot to provide a visual aid for the accuracy of the continuum rectification. There is work being done with the continuum rectification process to produce an algorithm that will automatically provide a rectified continuum for the user to further decrease the time that the user must spend preparing the spectrum for spectral classification. However, at the time that this paper is being written, this algorithm has not been completed and integrated into the overall program.

As of now, there have been no significant changes to the spectral typing process of the program although the plans for how this process will be altered are outlined in the discussion section of this paper.

3. RESULTS

Here we will present our findings for the increased efficiency of the program. This was tested by performing wavelength calibration and continuum rectification on a set of data using the first version of the program and then with the most recent version of the program. The spectral typing algorithm has not been altered as of yet so it is unnecessary to test the time difference in spectral typing between the two programs. The data set was collected at the Adams Observatory on 2018-04-10 and is a set of 16 spectra. It took 38 minutes to perform the data reduction using the original version of the program. It took 2 minutes for the extraction of the spectra, 14 minutes for the wavelength calibration, and 22 minutes for the continuum rectification. Using the most recent version of the program it took 22 minutes to perform the data reduction. It took 1.5 minutes for the extraction of the spectra, 30 seconds for the wavelength calibration and 20 minutes for the continuum rectification. This indicates an overall increase in efficiency of data reduction by 57 % due to changes made in the wavelength calibration and continuum rectification processes.

4. DISCUSSION

The results of the efficiency test clearly show that the changes made to the program dramatically decrease the amount of time the user must spend doing data reduction. Code optimization resulted in a 30 second decrease in the amount of time. The automation of the wavelength calibration means that very little time must be spent on that section. The user must simply check to see if the program has correctly located the hydrogen lines. There was not a significant decrease in the amount of time spent doing continuum rectification, but this is simply because the user must still perform the continuum rectification by hand.

The next step for this project is the development of an effective spectral typing algorithm. We have discussed two possible options for the methodology we will use for spectral typing. The first option is to use a method similar to MKCLASS but specialized for O, B and A type stars. The second option considered is adapting The Cannon for use with spectral classification. Both options have strengths and weaknesses and each option will be discussed in greater detail in the following paragraphs.

Specializing the algorithm employed by MKCLASS would involve two major changes. The first change would be the implementation of a classifying forest used to spectral type the spectra. The forest would consist of unique trees using a variety of criteria including a X2 comparison and the strength of metallic lines to create a spectral type. Each tree would then submit a “vote” for what the spectral classification of the spectrum is and the type with the most votes would be selected as the spectral type for that spectrum. This creates an accessible and quantitative uncertainty value for the spectral classification of the star. The second change would be the development of methods used to determine the spectral classification of spectra with traits that make it difficult to classify them, such as line broadening or a composite spectrum due to a binary system. These methods would be implemented in a separate classifying forest from the forest that would be used to classify the normal type stars. One of the advantages of specializing the MKCLASS methodology is that it still maintains a close connection with the MK classification that humans use to classify spectra. Another advantage is the lack of a need for a large volume of correctly pre-classified spectra for the program to work. Unfortunately, this method requires the maintenance of a large table of line strengths to use by the classifying forests. This table would require both a considerable amount of memory space, it would require the artificial definition of threshold values for line strengths.

Adapting The Cannon for use with spectral classification would require several steps for implementation. A relatively large collection of ideal spectra would have to be complied for use with The Cannon. The collection would have to be created by individually selecting the spectra that most closely resemble the spectral standards of a specific classification. The main classifying algorithm would also have to be redesigned to look for criteria specific to spectral standards. In addition, the program would have to be specialized to detect peculiarities in the spectrum that although they do deviate from the prescribed standard do not disqualify the spectrum from belonging to a specific classification. The major advantage of this method is that it could easily be applied to classifications other than O, B and A. The user would simply have to provide a selection of pre-spectral typed spectra of the type the user wished to include for the classifying algorithm to train on. The other advantage of this method is that it does not require the maintenance of a large table of artificially created thresholds. However, there are disadvantages to using The Cannon. The first being that the Adams Observatory has only collected a set of 653 spectra at the time this paper is being written. Given that the training set used in the paper about The Cannon contained 542 spectra, it is highly possible that the Adams Observatory does not have a sufficient number of spectra to successfully train the classifier employed by The Cannon (Ness et al. 2015). The Cannon does not conform precisely to the MK classification philosophy. While it does generally use the same ideology, it does not use the same methods that humans use when they perform spectral typing and so may produce results inconsistent with human spectral typing.

Regardless of which algorithm is used for the spectral classification, both methods require the creation and maintenance of a database of reference spectra including spectral standards and stars confidently classified. This database will allow the program to grow increasing accurate as it correctly identifies stars because it will have more known spectra to compare unknown spectra to. In addition, this database will allow for easy tracking of the changes of a single star over a period. Multiple spectra of a single star will be able to be easily located and compared to identify any changes or trends in the spectra. Maintaining a database will also allow for spectral standards to be easily changed and for spectra that have been classified by hand to be easily added to the program for use by the classification algorithm.

This project will be continued during the 2018-2019 school year at Austin College. The following is a list of goals to be achieved during the fall semester of the upcoming year.

* Create a database of spectral standards and stars that are very confidently spectral typed by hand
* Create a basic algorithm that can correctly spectral type the stars kept in the database
* Expand the algorithm to be able to interpolate between spectral standards
* Implement functionality for the program to be able to identify stars that are not able to be accurately classified based on their spectra
* Expand the classification algorithm to cover spectra that are chemically peculiar
* Be able to detect and extrapolate individual spectra from a composite spectrum from a binary system

5. REFERENCES

Gray, R. O., and C. J. Corbally. “An Expert Computer Program For Classifying Stars On The Mk Spectral Classification System.” *The Astronomical Journal*, vol. 147, no. 4, 2014, p. 80., doi:10.1088/0004-6256/147/4/80.

Gray, Richard O., and Christopher J. Corbally. *Stellar Spectral Classification*. Princeton University Press, 2009.

“Hertzsprung-Russell Diagram.” *COSMOS - The SAO Encyclopedia of Astronomy | COSMOS*, Swinburne University, 2015, astronomy.swin.edu.au/cosmos/H/Hertzsprung-Russell Diagram.

Ness, M., et al. “Apogee Kinematics. I. Overview Of The Kinematics Of The Galactic Bulge As Mapped By Apogee.” *The Astrophysical Journal*, vol. 819, no. 1, 2016, p. 2., doi:10.3847/0004-637x/819/1/2.

Ness, M., et al. “The Cannon: A Data-Driven Approach To Stellar Label Determination.” *The Astrophysical Journal*, vol. 808, no. 1, 2015, p. 16., doi:10.1088/0004-637x/808/1/16.