J. Comp. Int. Sci. (2015) 6(2):81-90
http://epacis.net/jcis/PDF_JCIS/JCIS-0097.pdf
jcis@epacis.net

Pan-american Associacion of Computational Interdisciplinary Sciences

An expert supernova spectral classification using artificial neural networks

Marcelo Módolo[a,c1], Lamartine N. F. Guimarães[b] and Reinaldo R. Rosa[a]

[a]Ph. D. Program in Applied Computer Science, National Institute for Space Research (INPE), São José dos Campos, SP, Brazil
[b]Nuclear Energy Division, Institute for Advanced Studies, Department of Aerospace Science and Technology (IEAv/CTA), São José dos Campos, SP, Brazil
[c]College of Engineerings, Technology and Information, Methodist University of São Paulo (UMESP), São Bernardo do Campo, SP, Brazil

Received on june 14, 2015 / accepted on july 30, 2015.

Abstract

The supernovae (SN) classification is an important scientific issue in astrophysics and cosmology. Usually, it is done by analyzing the spectra observed near the peak of the correspondent light curve. Because this task is difficult and usually made by an expert astronomer, it is important the study of computational techniques that allow the automatic classification of these spectra. In this paper we perform SN automatic classification method based on computational intelligence that simulates the human analytical expertise, making it a more formal classification and less prone to subjectivity of human analysis. Our classifier was developed using *Multilayer Perceptron Neural Network* to identify the usual SN types: Ia, Ib, Ic and II. The classifier was trained and tested on a database with 331 spectra of 56 different SN. The results are promising and indicate viability of this methodology for automatic SN classification in larger data sets.

Keywords: *Supernovae automatic classification, spectrum analysis, artificial neural network, computational data analysis.*

1. Introduction

The supernovae (SN) classification technique has been developed since 1941, recognizing two types of SN: Type I, characterized by hydrogen absence in its composition; and type II, with hydrogen presence [3]. The initial classification derived supernovae subtypes and the main current classification schemes consider nine different types of supernovae: Ia, Ib, Ic, IIb, IIL, IIP, IIF, IIn and IIpec. The subtypes determined by spectral properties are written in lower case and the subtypes determined by light curve properties are written in upper case.

Usually, the SN spectrum analysis should be done soon after the explosion arises, close to its maximum light emission, which occurs about fifteen days after the star explosion. Figure 1 shows the basic SN GK classification scheme (after Giunt and Kim) [1] which identifies that type Ia are associated with thermonuclear explosion of white dwarf stars, while other types of SN are associated with core collapse of massive stars.

[1]E-mail Corresponding Author: marmodolo@hotmail.com

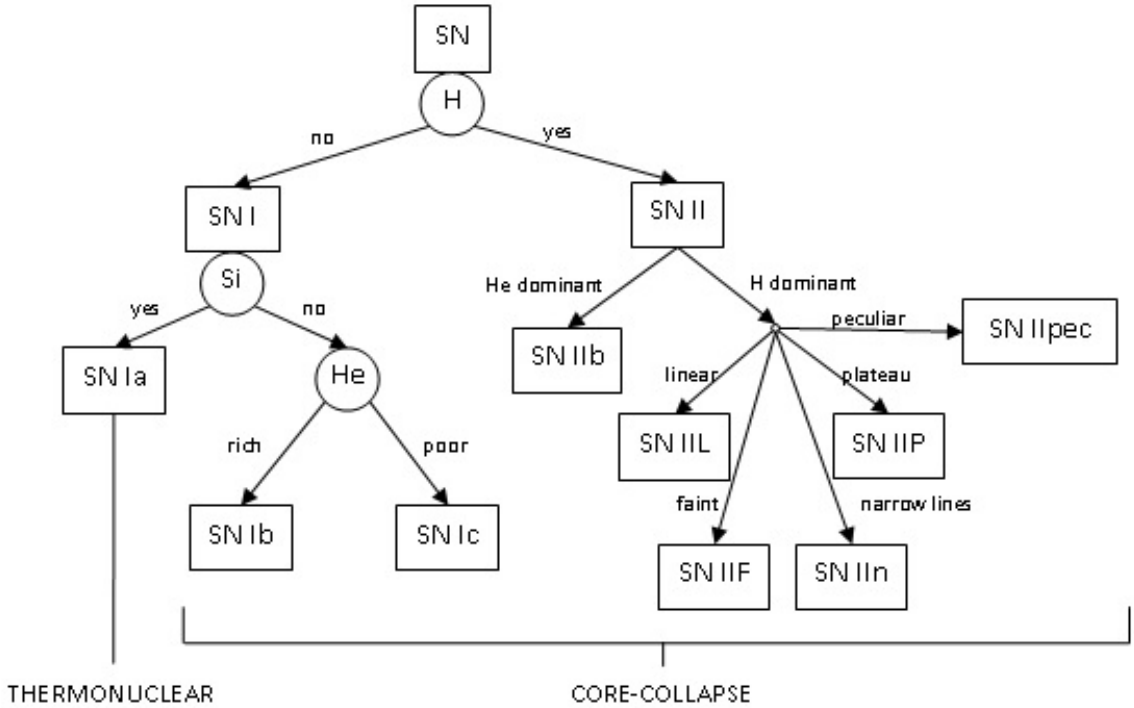

Figure 1 - The GK Supernovae Classification Scheme [1].

The classification from spectrum analysis considers presence or absence of hydrogen (H), helium (He) and silicon (Si) to determine "classic" types of supernovae: Ia, Ib, Ic and II.

- Type Ia: hydrogen absence and the silicon presence;
- Type Ib: hydrogen and silicon absence and Helium presence;
- Type Ic: hydrogen, silicon and helium absence;
- Type II: hydrogen presence.

According to the classification scheme shown in Figure 1, supernovae type IIb have He predominating over H. Among type II subtypes, wherein H is predominant over He, type IIn is the one in which emission spectrum shows narrow lines. Type IIpec has unique characteristics that do not fit the others.

There are also in type II supernovae, subtypes that are only differentiated by light curve shape: type IIL has almost linear luminance in time. Type IIP has plateau in their light temporal variation in time. And the type IIF supernova is considered weak.

Figure 2 shows examples of "classic" types of supernovae spectra captured at three different occasions: at full brightness, three weeks elapsed and a year later [3]. It also shows presence or absence of elements H, He and SI for different types of supernova spectra on maximum brightness. In the spectra captured, three weeks and one year after the maximum brightness, it is not possible to verify the presence of all these elements, making it impossible to use spectra to identify supernova type. In He specific case [4] these lines are visible only in the range of ten days prior to fifteen days after maximum brightness. This restriction limits spectra which can be used in identification of supernovae types and shows requirements to detect supernovae near its maximum light period.

The usage of classification scheme by spectrum analysis to identify supernova type is not trivial and only few expert astronomers are able to do so. Harutyunyan [5] considers classification made by experts unsatisfactory and based on subjective convictions. This subjectivity is due to this analysis being done visually in the spectrum chart, and may lead to different classifications of the same SN since there are no exact parameters. The difference in classification can be seen in [4], which proposes corrections of existing classification for 26 supernovae.

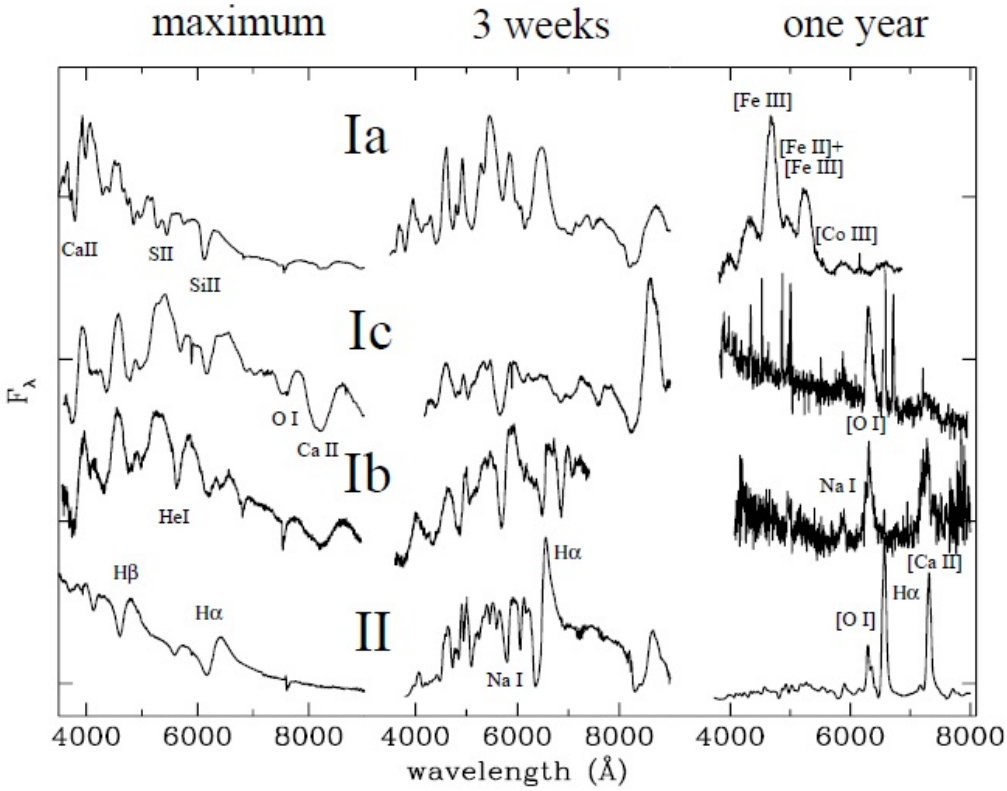

Figure 2 - Supernovae spectra examples at maximum brightness, three weeks later and a year later [3].

Based on these considerations, the automatic classification of supernovae considering spectrum characteristics has become the focus of the current research. This work includes one computational intelligence paradigm that simulates the classification process made by humans. It emerges as a way to minimize the subjectivity and allows any astronomer to make sorting.

Only two researches were found, which resulted in automatic classifiers of supernovae using spectrum characteristics. One developed in United States of America (USA) and the other in Italy. The USA method was developed in both the Harvard-Smithsonian Center for Astrophysics and the University of Hawaii Institute for Astronomy. It was implemented in a tool called SuperNova IDentification (SNID) [6]. This tool, besides supernova type, also identifies the redshift and the age of the supernova. Initially, this algorithm was developed to determine supernova redshift, but it was adapted to identify also the type. It takes advantage of correlation techniques commonly used for galaxies spectra comparison.

The Dipartimento di Universit Degli Studi Astronomy di Padova in Italy developed a supernova classification method in a doctoral thesis and implemented the PAdova Supernova Spectra comPARison TOOl (PASSPARTOO) [5]. The main algorithm was implemented in GEneric cLAssification TOol (GELATO) software and makes preprocessing all spectra for redshift correction, graphics smoothing and spectrum division in bins. It makes comparison of similarity between bins and considers the spectrum with lowest average between the relative distances the most similar. Figure 3 shows example of similarity comparison of supernovae spectra.

Computational intelligence paradigms are not used in these works, but conventional methods that use mathematical correlations to find similarity between spectra. A few researches used computational intelligence paradigms to develop supernovae automatic classifiers [7] [8] [9], but these researches use the light curve analysis and not spectrum analysis.

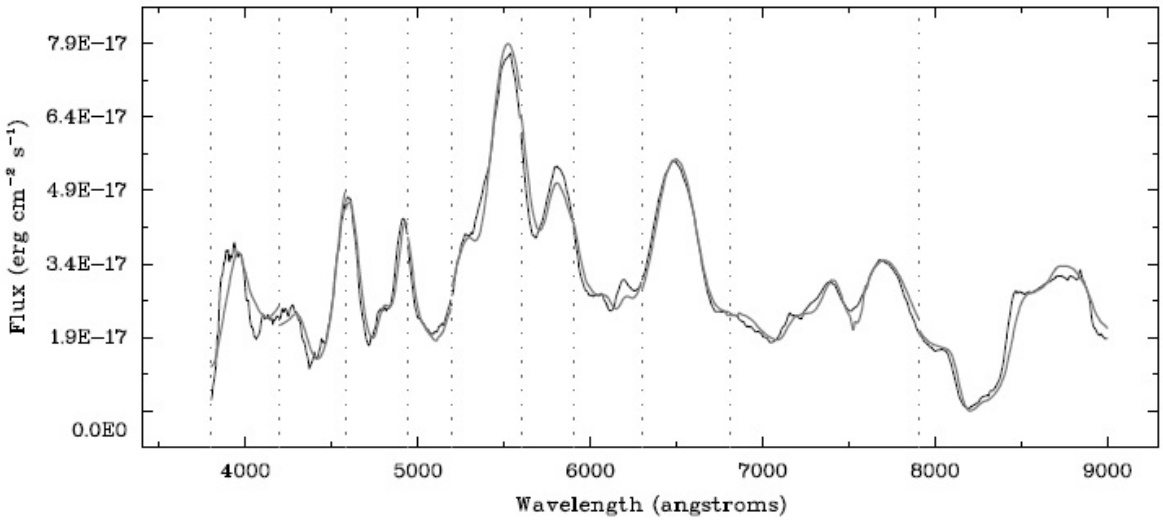

Figure 3 - Example of similarity comparison of supernova spectra 2002an and 2005bs [5].

The SNID [6] and GELATO [5] classifiers do not consider the criteria used by specialists and do not question them, simply analyze the similarity. These two works still have in common the spectrum comparison of newly discovered supernova, with all other supernovae spectra that were previously classified and assigning the type whose spectrum has greater similarity with newfound supernova. Thus, the classification made in these methods does not consider existing set of spectral characteristics, but only the features of one spectrum that is most similar.

The supernovae classification automatic method developed in this work uses the characteristics of a spectrum set, and not only one spectrum, thereby minimizing the impact that can happen when a supernova, used in the system spectra base, has its classification modified, as the cases presented in[4]. The other systems that uses only a spectrum, considered the most similar, to define the supernova type can have a big impact in case of classification change in one supernova because all the supernovae that used those spectra as more similar, also will have their classification modified.

This automatic system of supernova classification simulates the analysis of the expert, to look for presence or absence of certain elements in supernova spectrum and uses Computational Intelligence paradigms, in particular, Artificial Neural Networks, to identify supernova type.

The purpose of this paper is to automate supernova classification aiming at the reducing the existing subjectivity in human classification and making homogeneous analysis from training artificial neural network. Simulating the human expert analysis is the basis for creation of this classifier, but this work also seeks to systematize this analysis.

This systematization is relevant because subtypes have been introduced, excluded and subdivided[3], but the existing taxonomies remain ambiguous and not exhaustive, resulting in many supernovae classified as "peculiar"[5].

The classification made here, as in the other two automatic classifiers surveyed, only analyzes spectrum of supernova and not its light curve, and more specifically the types called "classics": Ia, Ib, Ic and II. The types IIb, IIn and IIpec are also determined by spectrum analysis, but they do not belong to the classic types, and were not explicitly separated in this work. Another reason for not separating them is that, there are few spectra of these types available and the subdivision of type II would not produce meaningful results because other subtypes have much more spectra than those. In addition, there are no supernovae spectra of types IIn and IIpec available in the spectra base used in this work.

Three versions of the classifier are developed and are described here in their chronological developed order to show the path followed to reach the current version. All versions were developed using Multi Layer Perceptron Neural Network, but perform analysis of different parts of the spectrum.

The first version uses the entire supernova spectra for the neural network training. The second version divides the spectrum into bins [5], and the third version uses intervals near the elements lines

The current version tends to use the same spectrum intervals which are used by human experts. It identifies separately if the supernova may or not be classified in one of the "classic" types

In the next section, the development of classifier three versions are described. Section 3 shows the results of the three versions and Section 4 discusses these results. Section 5 presents the conclusions.

2. Classifier Three Versions Development

During the development of the classifier, three different versions were built. They are shown here in its temporal development sequence, because results obtained in a previous version were used to improve results of next version.

The spectra database used in the three versions of the classifier was obtained on The Online Supernova Spectrum Archive (SUSPECT), hosted at the Department of Physics and Astronomy at The University of Oklahoma [10]. For these versions, spectra were selected from supernovae following some restrictions. The preprocessing performed in the three versions was the same used by Harutyunyan [**?**].

The spectral classification is an appropriate task to the Multilayer Perceptron (MLP) Neural Networks because it is a classification problem that can be solved using error correction learning, since it has an expected output [11]. Thus, MLP Neural Networks were built to do the classification of supernovae types on the three versions of the classifier. The spectra selection, the preprocessing performed and classifier three versions development are showed in the following.

2.1. Spectra Selection

Spectra database The Online Supernova Spectrum Archive (SUSPECT), provided by Department of Physics and Astronomy at The University of Oklahoma, consists of 1741 spectra of 185 captured between 1989 and 2006 [10].

The spectra selection used in this work for training and testing was done following criteria:

- Wavelength: 3800 to 6800 angstroms;
- Days of maximum light: -14 to +14 days;

Using these features were selected 331 spectra from 56 supernovae for training and testing. Of these, 244 are spectra from 28 supernovae of type Ia and 87 are spectra from 28 supernovae of other types. These types are shown in Table 1.

Table 1 - Types of all spectra selected for Neural Networks Training and Testing.

Type	Supernovae No.	Spectra No.
Ia	28	244
Ib	5	11
Ic	9	42
II	14	34
TOTAL	56	331

Spectra selected were divided into two sets: training (80%) and testing (20%). The set division followed the restriction: spectra of supernova that is in a set may not be in another one. Following this restriction, 265 spectra from 37 supernova were selected for training and 66 spectra from 19 supernova were selected for testing.

Table 2 shows the types of the 265 supernovae spectra (80% of total) of 37 supernovae selected for training.

Table 2 - Type of selected spectra for Neural Networks Training.

Type	Supernovae No.	Spectra No.
Ia	18	195
Ib	4	9
Ic	6	34
II	9	27
TOTAL	37	265

Table 3 shows the types of the 66 supernovae spectra (20% of total) of 19 supernovae selected for testing.

Table 3 - Type of selected spectra for Neural Networks Testing.

Type	Supernovae No.	Spectra No.
Ia	10	49
Ib	1	2
Ic	3	8
II	5	7
TOTAL	19	66

All spectra selected for training and testing were subjected to the same preprocessing.

2.2. Preprocessing

Preprocessing was performed in the spectra following steps of [5]. Here are the parameters used in the preprocessing:

- Make redshift (z) correction; - Graph smoothing with 70 angstroms parameter;
- Interpolating linearly at fixed points every eight;
- Vector normalization to magnitude one.

The same preprocessing was performed for the three versions of the classifier.

2.3. Version One

The first version of the classifier used only one MLP Neural Network in order to identify supernovae types "classic" using whole spectrum between 3800 and 6800 angstroms. Since this interval is interpolated every eight angstroms, the neural network had 376 inputs.

Some MLP Neural Network architectures with different numbers of layers and different amounts of neurons in layers were tested and one that had the best result was one hidden layer with 30 neurons.

2.4. Version Two

This version is based on the proposal of dividing the supernova spectrum in eleven bins of different sizes presented by Harutyunyan (HARUTYUNYAN, 2008). Table 4 shows the bins with the wavelength bands and spectral features of supernovae. Each bin presents peaks or valleys of spectrum, which mean presence, or absence of certain elements used for supernova classification.

In this version, an adaptation of the division into bins was made and only a few of them was used in each classification step, according to the presence or the absence of the component to be determined. The spectrum is divided into eleven bins, but only eight bins (2 to 9) are used. Bins 1, 10 and 11 are not used because they contain no elements required for classification, as well as because some spectra do not include spectral extreme values.

Table 4 - Bins with the wavelength bands and spectral features of supernovae [5].

Bins	Wavelength range	Ia features	Ib/c features	II features
1	3504 - 3792	Ca II	Ca II	Ca II
2	3800 - 4192	Si II, Ca II	Ca II	Ca II, H
3	4200 - 4576	Mg II, Fe II	Fe II	Mg II, Fe II, H
4	4584 - 4936	Fe II	Fe II	Fe II, H
5	4944 - 5192	Fe II	Fe II	Fe II
6	5200 - 5592	S II	S II, O I	S II
7	5600 - 5896	Si II, Na I	Na I, He I	Si II, Na I
8	5904 - 6296	Si II	He I	Si II
9	6304 - 6800	Fe II	Si II, He	I O I, H
10	6808 - 7904	O I	O I	O I
11	7912 - 9000	Ca II	Ca II	Ca II

Three MLP Neural Network have been created to verify the presence of hydrogen, silicon and helium that determine the type of supernovae according to the scheme of Giunt and Kin [**?**].

The first neural network was created to separate into two classes supernovae types I and type II by identifying the absence (type I) or presence (type II) of hydrogen. This neural network has 206 inputs corresponding to spectral intervals of boxes 2, 3, 4 and 9. Many different architectures of neural networks were tested and some of them obtained the best result. The neural network with the least amount of neurons that achieved this result was with only one hidden layer with five neurons. All the 265 spectra were used in the training and all the 66 spectra were used in the tests.

Another neural network was created to separate type Ia supernovae of types Ib and Ic supernovae, called here type Ibc, by identifying the presence (type Ia) or absence (type Ibc) of silicon. This neural network has 113 inputs corresponding to spectral intervals of boxes 8 and 9. The best architecture of the tested neural networks was also with only one hidden layer with five neurons. The type II supernova spectra were excluded from the training and test sets of this neural network because are analyzed only type I supernovae spectra. Thus, 238 spectra were used in the training and 43 spectra were used in the tests.

The third neural network was created to separate type Ib supernovae of tupe Ic supernovae by identifying the presence (type Ib) or absence (type Ic) of helium. This neural network has 202 inputs corresponding to spectral intervals of boxes 2, 7, 8 and 9. The neural network with one hidden layer with five neurons was the one that had the best result. The type II and the type Ia supernova spectra were excluded from the training and test sets of this neural network. Thus, only 43 spectra were used in the training and only 10 spectra were used in the tests.

2.5. VersionThree

Three MLP Neural Network have been also created to verify the presence of hydrogen, silicon and helium that determine the type of supernovae according to the scheme of Giunt and Kin [**?**] One MLP Neural Network was created to identify each element using an interval of [-100 +100] angstroms from spectrum element line. Identification of the elements frequency values lines were performed:

- Hydrogen: Balmer series are shown in Table 5.

Table 5 - Balmer series.

Color	Name	Wavelength
Red	H	6563
Green	H	4861
Blue	H	4341
Viiolet	H	4102

- Silicon: Peak in 6150 angstroms; Valley in 6355 angstroms.

- Helium: Distinction between Ib and Ic using beyond of He I 5876 line, also He I 6678 and He I 7065 lines that are visible between phases -10 and +15 [4].

The first neural network was created to separate into two classes supernovae types I and type II by identifying the absence (type I) or presence (type II) of hydrogen. This neural network has 104 inputs corresponding to spectral intervals of Balmer series. Many different architectures of neural networks were tested and some of them obtained the best result. The neural network with the least amount of neurons that achieved this result was with only one hidden layer with five neurons. All the 265 spectra were used in the training and all the 66 spectra were used in the tests.

After that, a neural network was created to separate type Ia supernovae of types Ib and Ic supernovae, called here type Ibc, by identifying the presence (type Ia) or absence (type Ibc) of silicon. This neural network has 113 inputs corresponding to spectral intervals of peak and valley of silicon, which are on 6150 and 6355 angstroms, respectively. The best architecture of the tested neural networks was with only one hidden layer with 15 neurons. Here also the type II supernova spectra were excluded from the training and test sets of this neural network because are analyzed only type I supernovae spectra. Thus, 238 spectra were used in the training and 43 spectra were used in the tests.

The third neural network was created to separate type Ib supernovae of tupe Ic supernovae by identifying the presence (type Ib) or absence (type Ic) of helium. This neural network has 202 inputs corresponding to spectral intervals of the helium lines. The neural network with one hidden layer with five neurons was the one that had the best result. The type II and the type Ia supernova spectra were excluded from the training and test sets of this neural network. Thus, here also only 43 spectra were used in the training and 10 spectra were used in the tests.

The results of tests performed on these three classifier versions are shown in the next section.

3. Results

Tests of the three versions of the classifier were made using the same spectra selected of the SUSPECT database, which were subjected to the same preprocessing. The three versions were developed using MLP Neural Networks in training and testing, but with different architectures. The ratio of the spectrum number correctly classified by the total number of spectra is used to calculate the correct percentage in the three versions.

3.1. Version One Results

The only MLP Neural Network of the first version of the classifier used 265 patterns for training and 66 patterns for testing. Were observed 16 errors that result in 75.8% accuracy. Table 6 shows a summary result obtained with all errors of the neural networks for classifier version one.

Table 6 - Summary results of Classifier Version One MLP Neural Networks (NN).

Spectra	SN Type	SN Name	Phase	NN Class
53	Ic	1995F	10	Type-Ia
50	Ib	1999dn	0	Type-Ic
49	Ia	2003hv	9	Type-Ic
52	Ic	2004aw	1	Type-Ia
55	Ic	2004aw	5	Type-Ia
56	Ic	2004aw	6	Type-Ia
13	Ia	2005bl	12	Type-II
22	Ia	2005bl	-3	Type-Ic
26	Ia	2005bl	4	Type-Ic
30	Ia	2005bl	-5	Type-II
36	Ia	2005bl	6	Type-Ic
15	Ia	2005hk	13	Type-Ic
27	Ia	2005hk	-4	Type-II
31	Ia	2005hk	-5	Type-II
37	Ia	2005hk	-6	Type-II
44	Ia	2005hk	-8	Type-II

As can be observed in Table 6, the type Ia supernovae spectra were responsible for 11 errors and other types and spectra of other types identified incorrectly as type Ia were responsible for more four errors. These 15 errors represent 93.8% of all 16 errors. In relation to the 49 spectra of type Ia supernovae it means that 30.6% of spectra were identified incorrectly.

3.2. Version Two Results

Three MLP Neural Networks have been used to identify the elements hydrogen, silicon and helium. Each neural network was responsible for identifying an element.

In the neural network that identified the hydrogen, all the 265 spectra were used in the training and all the 66 spectra were used in the tests. The results were:

- Error amount = 4
- Percentage of correct answers = 93.9%

The neural network that identified the silicon used 238 spectra in the training and 43 spectra in the tests. The results were:

- Error amount = 12
- Percentage of correct answers = 79.7%

Only 43 spectra were used by neural network that identified the helium in the training and only 10 spectra were used in the tests. The results were:

- Error amount = 5
- Percentage of correct answers = 50.0%

Table 7 shows a summary result obtained with all errors of the three neural networks for classifier version two.

Table 7 - Summary results of Classifier Version Two MLP Neural Networks (NN).

Neural Network	Errors No.	% Correct	Spectra	SN Type	SN Name	Phase	NN Class
NN H	4	93.9%	54	Ic	SN 1991N	5	Type_II
			64	II	SN 1999gi	-3	Type_I
			13	Ia	SN 2005bl	12	Type_II
			15	Ia	SN 2005hk	13	Type_II
NN Si	12	79.7%	18	Ia	SN 2000E	-2	Type_Ibc
			20	Ia	SN 2000E	-3	Type_Ibc
			3	Ia	SN 2003cg	1	Type_Ibc
			13	Ia	SN 2005bl	12	Type_Ibc
			22	Ia	SN 2005bl	-3	Type_Ibc
			26	Ia	SN 2005bl	4	Type_Ibc
			30	Ia	SN 2005bl	-5	Type_Ibc
			15	Ia	SN 2005hk	13	Type_Ibc
			27	Ia	SN 2005hk	-4	Type_Ibc
			31	Ia	SN 2005hk	-5	Type_Ibc
			37	Ia	SN 2005hk	-6	Type_Ibc
			44	Ia	SN 2005hk	-8	Type_Ibc
NN He	5	50%	4	Ic	SN 1995F	10	Type_Ib
			1	Ib	SN 1999dn	0	Type_Ic
			6	Ic	SN 2004aw	5	Type_Ib
			7	Ic	SN 2004aw	6	Type_Ib
			10	Ic	SN 2004aw	8	Type_Ib

As can be observed in Table 7, the type Ia supernovae spectra were responsible for 14 of the 21 errors that occurred, which represents 66.7% of all errors. In relation to the 49 spectra of type Ia supernovae it means that 28.6% of spectra were identified incorrectly.

3.3. Version Three Results

The third version of the classifier also used a MLP Neural Network to identify each of the elements hydrogen, silicon and helium. The number of patterns used in the training and test each neural network was also the same.

The results to the neural network that identified the hydrogen, obtained for all 265 spectra used in the training and all the 66 spectra used in the tests, were:

- Error amount = 2
- Percentage of correct answers = 97.0%

The results to the neural network that identified the silicon, obtained using 238 spectra in the training and 43 spectra in the tests, were:

- Error amount = 8
- Percentage of correct answers = 84.6%

The results to the neural network that identified the silicon, obtained using only 43 spectra in the training and 10 spectra in the tests, were:

- Error amount = 4
- Percentage of correct answers = 60.0%

Table 8 shows a summary result obtained with all errors of the neural networks for classifier version three.

Table 8 - Summary results of Version Three MLP Neural Networks (NN).

Neural Network	**Errors No.**	**% Correct**	**Spectra**	**SN Type**	**SN Name**	**Phase**	**NN Class**
NN H	2	97.0%	2	Ic	SN 1991N	5	Type_II
			53	Ia	SN 2005bl	12	Type_II
NN Si	8	84.6%	3	Ia	SN 2003cg	1	Type_Ibc
			8	Ia	SN 2003cg	10	Type_Ibc
			13	Ia	SN 2005bl	12	Type_Ibc
			15	Ia	SN 2005hk	13	Type_Ibc
			18	Ia	SN 2000E	-2	Type_Ibc
			20	Ia	SN 2000E	-3	Type_Ibc
			26	Ia	SN 2005bl	4	Type_Ibc
			30	Ia	SN 2005bl	-5	Type_Ibc
NN He	4	60.0%	1	Ib	SN 1999dn	0	Type_Ic
			4	Ic	SN 1995F	10	Type_Ib
			5	Ic	SN 1991N	5	Type_Ib
			7	Ic	SN 2004aw	6	Type_Ib

Table 8 show that the type Ia supernovae spectra were responsible for 9 of the 14 errors that occurred, which represents 64.3% of all errors. In relation to the 49 spectra of type Ia supernovae it means that 18.4% of spectra were identified incorrectly. This result was considered unsatisfactory for the purposes of this study and a new version of the classifier is in development. All these results are discussed in the next section.

4. Discussion

In analyzing the results of the three versions of the classifier, it is important to note that the first version of the classifier had a unique percentage of correct identification of the four classic types. The other two versions presented three percentage of correct answers, one for each of the three neural networks, which identified the elements hydrogen, silicon and helium. Making a superficial analysis of the percentage of correct answers of the three versions of the classifier, it seems that there was no progress against the first version, mainly due to lower results obtained by the neural network that identifies the helium to separate type Ib spectra of type Ic spectra. However, this neural network used the small amount of spectra, only 43 spectra in training and 10 spectra in test, while the other two neural networks used a much larger amount of

spectrum because they include the spectra of supernovae type Ia which are the most in SUSPECT database. In the neural network that identifies the hydrogen 265 spectra were used for training and 66 spectra for testing to separate the type I of type II. In the neural network that identifies the silicon were used 238 spectra for training and 43 spectra for test to separate type Ia of types Ib and type Ic.

Thereby, to make the comparison of neural networks results that used almost the same amount of spectra in the training and testing, the percentage of correct answers was calculated using the results of the neural networks of the last two versions that separate type Ia of types Ib and type Ic. In this calculation, for the last two versions of the classifier, have been used only the results of the neural network that verified the presence of silicon, which are 79.7% for the first version and 84.6% for the second version. For the first version of the classifier were used the 15 errors occurring to separate type Ia of other types, thus the correct percentage was 77.3%. Figure 4 shows a comparison between the percentage of correct answers of the classifier three versions in the separation of type Ia supernovae spectra.

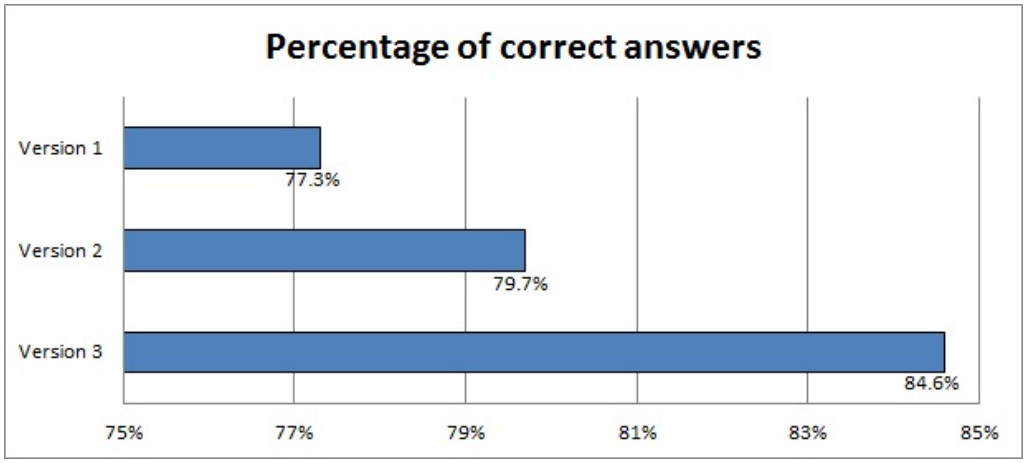

Figure 4 - Comparing the percentages of correct answers from the three versions of the classifier on the separation of supernovae spectra type Ia.

In Figure 4 you can verify that the results of versions one and two are very close, since increase was only 2.4% for the second version. The third version has already obtained a more significant improvement of 4.9% over the second version and 7.3% over the first version. Thus, although not very large, for each new version there was an increase in the correct percentage in the identification of type Ia supernovae.

The three versions of the classifier showed the greatest number of errors in identifying the type Ia supernovae, but this is due to the greater quantity of this type of supernova present in SUSPECT database and that were used in the tests. Table 9 shows the amount of type Ia supernovae spectra compared to the other types and the number of errors in identifying the type Ia supernovae spectra compared to other types.

Table 9 - Ratio between the quantity of error in supernovae spectra of Ia type and other types.

Version	Spectra No. Type Ia	Spectra No. Other types	% Spectra Type Ia	Error No. Type Ia	Error No. Other types	% Errors Type Ia
1	49	17	74.2%	11	4	73.3%
2	49	10	83.1%	12	0	100.0%
3	49	10	83.1%	8	0	100.0%

As can be viewed in Table 9, in the classifier first version the percentage of classification errors of type Ia supernovae spectra is very close to the percentage of spectra of this type that were used in the tests. In the classifier versions two and three, the errors were 100% in type Ia supernovae, but this may be due to the percentage of type Ia spectra be even greater because the seven type II spectra were excluded in the network testing that separating type Ia from types Ib and Ic.

Comparing the results of the three versions of the classifier with the results of the two researched classifiers (GELATO [5] and SNID [**?**]) cannot be made because the percentage of correct answers of these classifiers are not been disclosed in the researched work.

About the computational cost, it can be said that the training time of the neural network is very small. All the neural networks had only one hidden layer of no more than 30 neurons and all of them converged in a few seconds.

5. Conclusions

This paper proposes creating a supernovae automatic classifier as from the spectrum analysis in order to reduce the existing subjectivity in human analysis, making the classification more homogeneous due to systematization that is done during the automation process. Three versions of supernova classifier was successfully constructed using Artificial Neural Networks to analyze the same spectrum intervals to identify the types of supernova. The results achieved are promising and indicate viability of continuing this research.

The number of type Ia supernovae spectra used in this study is much higher than all the other types together among selected spectra, 73.7% are of the type Ia spectra and only 26.3% of other spectra types (Ib, Ic and II). The inclusion of more supernova spectra these other types can result in better results of classifier.

The selection of spectra considering an interval of -14 to +14 days relative to the maximum light. The reduction in this interval results in the selection spectra captured closer to the maximum light can also improve the performance of the classifier, because it reduces the spectra quantity and diversity of a single supernova, causing the selected spectra have characteristics more similar. This reduction depends on the inclusion of supernovae spectra all types.

The classifier versions presented increasing results for the percentage of correct answers in identifying the types of supernovae. In the first version, entire spectrum was analyzed and the worst result (77.3%) was obtained. In the last two versions use only certain parts of the spectrum proved to be a better strategy. Spectrum division in boxes of second version, obtained a better result (79.7%) than the first version, and use spectrum intervals with the lines of the elements in the center, had the best result (84.6%), showing was the best strategy. Thus, continuation of this work should following the strategy of using some intervals of the spectrum, but try to define these intervals as near as possible to those used by human expert.

Acknowledgments. RRR gratefully thanks the support given by FAPESP under the grant number 2014/11156-4.

References

[1] Giunti, C. and Kim, C. W. Fundamentals of Neutrino Physics and Astrophysics, Nova York: Oxford University Press, 2007.

[2] Turatto M., Benetti S. e Pastorello A. Supernova classes and subclasses, American Institute of Physics Conference Series, vol. 937, pp. 187-197, 06 2007.

[3] Turatto M. "Classification of Supernovae", Supernovae and Gamma-Ray Bursters, vol. 598 de Lecture Notes in Physics, pp. 21-36, 01 2003.

[4] Modjaz M., Blondin S., Kirshner R. P., Matheson T., Berlind P., Bianco F. B., Calkins M. L., Challis P., Garnavich P., Hicken M., Jha S., Liu Y. Q. and Marion G. H. "Optical Spectra of 73 Stripped-envelope Core-collapse Supernovae", The Astronomical Journal, vol. 147, pp. 99-115, 05 2014.

[5] Harutyunyan A. "Automatic Objective Classification of Supernovae", Padova, 2008.

[6] Blondin S. and Tonry J. L. "Determining the Type, Redshift, and Age of a Supernova Spectrum", The Astrophysical Journal, vol. 666, pp. 1024-1047, 09 2007.

[7] Poznanski D., Maoz D. e Gal-Yam A. Bayesian Single-Epoch Photometric Classification of Supernovae, The Astronomical Journal, vol. 134, pp. 1285-1297, 2007.

[8] Rodney S. A. and Tonry J. L. Fuzzy Supernova Templates. I. Classification, The Astrophysical Journal, vol. 707, pp. 1064-1079, 2009.

[9] Pascale M. D. Automatic Classification of Supernovae using machine learning methods, Padova, 2011.

[10] Hogan M., Parrent J. and Feldt A. "SUSPECT", 2010. [Online]. Available: http://suspect.nhn.ou.edu/ suspect/. [Accessed 10 06 2013].

[11] Haykin S. Redes Neurais - Princpios e Prticas, 2 ed., Porto Alegre: Bookman, 2001.