Review and matlab implementation of the paper:

Using neural network ensembles for bankruptcy prediction and credit scoring (Tsai, Wu)

Filippo Tolin

Ca' Foscari University of Venice

March 28, 2024

Summary

1. Paper review
   - Goals
   - Tools
   - Advancements vs previous works
   - Study 1
   - Study 2
   - Study 3
   - Conclusions
   - Conclusions
2. Matlab implementation
   - A subsection
3. Second Section

Paper review

Goals

- Observe the performance differences between different ANN ensemble approaches, namely single classifiers, multiple classifiers and diversified multiple classifiers, with regards to based on a set credit scoring and bankruptcy detection. The study is based on three of heterogeneous datasets;
- Evaluate the three classifier architectures performance with regards to Type 1 error and Type 2 error.

Tools

- Multilayer Perceptron feedforward Artificial Neural Network
  - One **hidden** layer
  - Five different values for the **hidden nodes** hyperparameter
  - Four different values for the **training epochs** hyperparameter
- Multiple classifiers with two techniques to compute them:
  1. Best n classifiers for every epoch
  2. Best n classifiers among all the epochs

Note

Technique 1 is only applicable to $n = 3 \wedge n = 5$, with $n \in [3, 5, 7, 9, 11, 13, 15]$. Multiple classifiers are based on **majority voting**.

Advancements vs previous works

- Employment of **multiple datasets** for system validation;
- Usage of **Type 1 and Type 2 errors** and not only average accuracy measures;
- Testing the classifiers performance on multiple classification tasks rather than a single one, specifically on **credit scoring** and **bankruptcy prediction**.

Brief remark

- **Type 1** error is associated with **false positives**;
- **Type 2** error is associated with **false negatives**.

Examples

- *Type 1*: the model classifies a credit-worthy client as a credit-risky one;
- *Type 2*: the model classifiers a credit-risky client as a credit-worthy one.

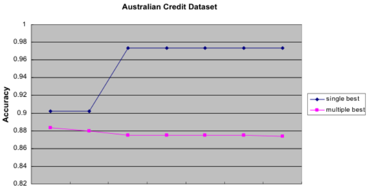
Study 1: single vs multiple classifiers

- Datasets are split into training (70%) and test (10%);
- For single classifiers the number of nodes is $nn \in [8, 12, 16, 24, 32]$ and learning epochs $[50, 100, 200, 300]$.
- Multiple classifers are build with the voting strategy combining the results of the top n classifers, with $n \in [3, 5, 7, 9, 11, 13, 15]$.

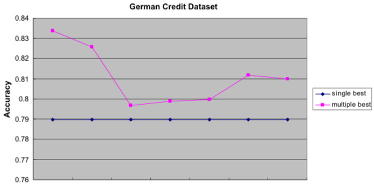
Takeout

On average, the single best classifer outperforms multiple classifers.

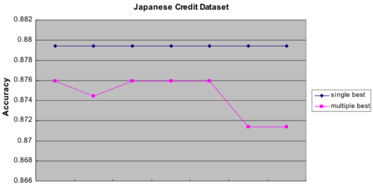
Study 1: single vs multiple classifiers



(a) Australian



(b) German



(c) Japanese

Figure: Comparison between single classifiers and multiple classifiers.

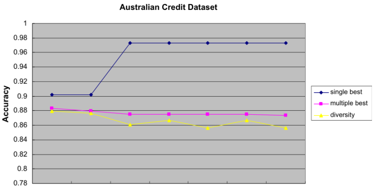
Study 2: single vs multiple vs diversified classifiers

- Train-test dataset generation is different for diversified multiple classifiers. Specifically, every model composing the classifier is trained on a fraction of the observations from the same dataset, then the majority voting is executed using a test dataset;
- The procedure aims at ensuring **diversity** between classifiers.

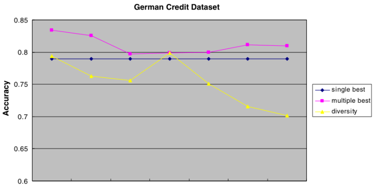
Takeout

The best single classifier is still, on average, a better classifier than the diversified multiple classifier (and the multiple classifier, as seen before).

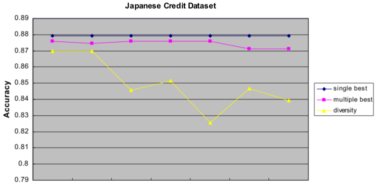
Study 2: single vs multiple vs diversified classifiers



(a) Australian



(b) German



(c) Japanese

Figure: Comparison between single, multiple and diversified classifiers.

Study 3: Type 1 and Type 2 errors

- In Study 1 and Study 2 the results of classifiers are compared based on the **accuracy** of the classifiers. Study 3 compares the models performance with regards to Type 1 and Type 2 errors.

Takeout

This study highlights how single classifiers do not totally outperform multiple or diversified classifiers.

Study 3: Type 1 and Type 2 errors

Average error rate of Type I and Type II errors

	Australian credit			German credit			Japanese credit		
	S	M	D	S	M	D	S	M	D
Type I error	*12.16*	12.85	14.28	*44.27*	45.25	59.82	15.02	15.07	*14.42*
Type II error	12.97	12.14	*11.55*	9.48	*8.46*	8.67	10.79	*10.00*	14.06

Figure: Type 1 and Type 2 error across datasets and classifier architectures.

Conclusions

- If the performance is measured with **accuracy**, **single best neural network classifier is more suitable** for bankruptcy prediction and credit scoring tasks, if compared with multiple or diversified multiple neural network classfiers;
- If performance is measured with **type 1 and 2 errors**, there seems to be no **clear winner** among the model architectures analysed.

Matlab implementation

Blocks of Highlighted Text

In this slide, some important text will be highlighted because it's important. Please, don't abuse it[1].

Block

Sample text

Alertblock

Sample text in red box

Examples

Sample text in green box. The title of the block is “Examples”.

[1] This is a footnote

Multiple Columns

Heading

1. Statement
2. Explanation
3. Example

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

Second Section

Table

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296

Table: Table caption

Theorem

Theorem (Mass–energy equivalence)

$E = mc^2$

Figure

Uncomment the code on this slide to include your own image from the same directory as the template .TeX file.

Citation

An example of the `\cite` command to cite within the presentation:

This statement requires citation [Smith, 2012].

References

John Smith (2012)
Title of the publication
Journal Name 12(3), 45 – 678.

The End