Highlights

Multi-sensor Planetary gearbox fault diagnosis using Multi-layer Convolutional Gated Recurrent Unit Networks

CV Radhakrishnan,Han Theh Thanh,CV Rajagopal,Rishi T.

- Research highlights item 1
- Research highlights item 2
- Research highlights item 3

Multi-sensor Planetary gearbox fault diagnosis using Multi-layer Convolutional Gated Recurrent Unit Networks[⋆,⋆⋆]

Sir CV Radhakrishnan[a,c,*,1] (Researcher), Han Theh Thanh[b,d], CV Rajagopal Jr[b,c,2] (Co-ordinator) and Rishi T.[a,c,**,1,3]

[a]*Elsevier B.V., Radarweg 29, Amsterdam, 1043 NX, The Netherlands*
[b]*Sayahna Foundation, Jagathy, 695014, Trivandrum, India*
[c]*STM Document Engineering Pvt Ltd., Mepukada, Malayinkil, 695571, Trivandrum, India*

ARTICLE INFO

Keywords:
quadrupole exciton
polariton
WGM
BEC

ABSTRACT

abstract
`\beginabstract ...\endabstract` and `\begin{keyword} ... \end{keyword}` which contain the abstract and keywords respectively.
Each keyword shall be separated by a `\sep` command.

1. Introduction

With the increase in the service life of urban rail trains, key components such as gears, wheelsets, and bearings gradually enter the wear and tear failure period, and the failure rate is increasing. To meet the important requirements for the safe operation of urban rail transit, it is urgent to monitor and diagnose the health status of key components. Due to the safety requirements of practical application, there are problems with small fault data and a few types of faults in the operation data of urban rail trains. This unbalanced small sample set is difficult to train the fault diagnosis model effectively, which makes it difficult to extract fault features. Therefore, carrying out intelligent, accurate, and fast fault diagnosis research on the gearbox of urban rail trains has become an urgent problem to be solved in the safe operation of urban rail transit.

The existing fault diagnosis methods for gearboxes can be divided into four categories: physical model-based method, signal processing-based method, machine learning-based method, and hybrid method. Physical model-based detection methods require a very high level of knowledge in the field of inspectors and require a very deep understanding of machine composition. As a result, it is difficult to build a relatively accurate physical model for complex mechanical equipment in a dynamic and noisy complex environment. Solid fault representation theory and mathematical basis are based on the premise of the signal processing detection method, which aims to explore advanced signal denoising and filtering technology to extract fault features effectively. As a typical representative of data-driven methods, machine learning detection methods are flourishing with the development of the modern industry. Traditional shallow machine learning models often require professionals to manually extract features related to selection tasks, which are very limited in large data analysis. Secondly, feature extraction and decision-making are separate from each other. The asynchronous optimization process takes a lot of time, and because of the shallow structure, it is difficult to effectively mine high-dimensional features. As the complexity of the mechanical structure increases, the number and type of

[⋆] This document is the results of the research project funded by the National Science Foundation.

[⋆⋆] The second title footnote which is a longer text matter to fill through the whole text width and overflow into another line in the footnotes area of the first page.

This note has no numbers. In this work we demonstrate a_b the formation Y_1 of a new type of polariton on the interface between a cuprous oxide slab and a polystyrene micro-sphere placed on the slab.

[*] Corresponding author
[**] Principal corresponding author
`cvr_1@tug.org.in` (C. Radhakrishnan); `cvr3@sayahna.org` (C. Rajagopal); `rishi@stmdocs.in` (R. T.)
`www.cvr.cc,cvr@sayahna.org` (C. Radhakrishnan); `www.sayahna.org` (C. Rajagopal); `www.stmdocs.in` (R. T.)
ORCID(s): `0000-0001-7511-2910` (C. Radhakrishnan)

[1] This is the first author footnote. but is common to third author as well.

[2] Another author footnote, this is a very long footnote and it should be a really long footnote. But this footnote is not yet sufficiently long enough to make two lines of footnote text.

sensors often need to be increased to achieve better diagnostic performance, increasing data dimension and dynamics. In summary, it is difficult for traditional fault detection methods to obtain satisfactory diagnostic results.

The most popular branch of machine learning, in-depth learning, has witnessed the flourishing development of computer images, natural language processing, and other fields. In addition to the intrinsic factors such as powerful data processing, feature extraction ability, and network architecture innovation, the explosive growth of extensive industrial data, hardware breakthroughs and other stimuli are also external factors that cannot be ignored. Naturally, in-depth learning has also triggered a wave of intelligent fault diagnoses in the past five years.

Among them, deep learning methods such as convolution neural networks and automatic encoders take the leading position in the current gearbox fault diagnosis because they can automatically extract spatial features. [Citation]. In summary, CNNs can extract spatial characteristics from the variation of amplitude and different levels of spectrum and can fuse data sources from multiple sensors by identifying the correlation between measurements. However, time features that contain fault-sensitive information, such as the order or interval of time steps corresponding to the exception, cannot be extracted.

Deep learning methods such as cyclic neural networks and LSTM are gradually applied to extract time features automatically for gearbox fault diagnosis. [Citation]. In summary, most existing methods capture time-varying changes in time series data by capturing the time interval between the start and end of an anomalous response caused by a failure and the normal response. However, due to the limitations of the network structure, these in-depth learning methods cannot determine the short-term spatial correlation among multiple observations in a single time step.

Due to the complex structure of the gearbox, a variety of sensor devices, such as accelerometers [Citation], shaft encoders [Citation], and acoustic sensors [Citation], have been used to monitor the health of the gearbox and classify the failures to improve the accuracy of classification. Different types of sensors have different measurable frequency ranges. Generally speaking, in the contact sensor, the speed sensor is suitable for measuring low-frequency vibration caused by unbalance, misalignment, loose contact, etc. It can measure the vibration displacement and obtain stable data. The acceleration sensor is suitable for measuring medium and high-frequency vibration signals caused by faults of gears and bearings, but its measurement of vibration displacement is often unstable. Therefore, the accelerometer is generally only used to measure the vibration speed. Its advantage is that it can detect the high-frequency vibration signal. With the increase in sensor types and numbers, gearbox fault diagnosis is gradually combined with multi-sensor data fusion technology. Traditional data fusion methods extract features from a single observation and then extract features based on physical and statistical analysis. [Citation]. However, different sensor signals often exhibit different physical characteristics. For example, there is little correlation between accelerometers and acoustic sensors. As a result, most existing studies can only extract the relationship between different sensors of the same type. Traditional methods need to extract manual features from that sensor source, which results in lower efficiency.

To solve these problems, capturing the space-time characteristics and fusing different kinds of sensor data, and extracting the features at the same time, we propose a new in-depth learning method and apply it to gearbox fault diagnosis under different working conditions. First, the sampling signal is divided into several time steps, and the signal in each time step is divided into frequency band information by wavelet packet decomposition. Based on the Heisenberg uncertainty principle, the time-domain resolution in the decomposed time-frequency domain information decreases, and the frequency-domain resolution increases. In this way, one-dimensional time domain signals can be developed into a feature matrix containing time-frequency information. By splicing the feature matrices derived from the decomposition of multiple sensors in the depth dimension, a three-dimensional matrix with time domain resolution, frequency domain resolution, and several sensors as long, wide and high can be obtained. Secondly, the gated unit in the GRU network is improved by using convolution instead of matrix multiplication to get the ConvGRU unit. Convolution operations in ConvGRU units are designed to extract short-term signal characteristics over the same time step, while forward propagation between ConvGRU units is designed to capture anomalous responses through long-term time-series dependencies of characteristics between different time steps.

The main contributions in this paper are summarized as follows.

1. With little reliance on expert knowledge and the use of specific signal processing methods for feature selection, the proposed method uses the original time series signal as input to perform gearbox fault diagnosis end-to-end.
2. The proposed MCGRU network model fuses multi-sensor data effectively by convolution operation, which makes the model scalable for the number and category of sensors, and improves the diagnostic accuracy when using multi-sensor data. As the number of layers of the cyclic neural network increases, the number of parameters of the convolution and cyclic neural networks decreases dramatically compared with the traditional convolution and cyclic neural networks by pooling and other operations.

3. By combining the advantages of convolution and cyclic neural networks, the MCGRU network model proposed in this paper calculates the three-dimensional feature matrix as the input information and the intermediate feature in the cyclic neural network, capturing the long-term time series dependence hidden in the sequence data while preserving the spatial and frequency domain characteristic information of the signal. The dependencies derived from the proposed network model are applied to the fault classification tasks under different working conditions. The experimental results show that the proposed method is superior to or comparable to the existing intelligent diagnostic methods.

The rest of this paper is organized as follows. Section 2 introduces the technical background used in the proposed method. Section 3 illustrates the framework of the proposed MCGRU method in detail. Section 4 presents a case study to demonstrate the effectiveness of the proposed framework. Finally, we conclude the paper and future work in Section 5.

2. Background and preliminaries

2.1. Wavelet packet decomposition

Fourier transform is suitable for processing stationary signals. It can only obtain the components of which frequencies a signal generally contains, but it does not know the time when each component appears. Therefore, two signals with large time-domain differences may have the same spectrum. Wavelet analysis and wavelet packet decomposition are suitable for non-stationary signal analysis. Compared with wavelet analysis, wavelet packet decomposition can be used to analyze the signal more finely, divide the time-frequency plane more finely, and adaptively select the best wavelet basis function according to the characteristics of the signal, to better analyze the signal, so wavelet packet decomposition is more widely used.

Wavelet analysis only further decomposes the low-frequency part of the signal and does not continue to decompose the high-frequency part, that is, the detail part of the signal. Therefore, wavelet transform can well represent a large class of signals with low-frequency information as the main component, and cannot well decompose and represent signals containing a large amount of detail information, small edges or textures, such as Non-Stationary Mechanical vibration signals, remote sensing images, seismic signals, and biomedical signals.

Wavelet packet decomposition can decompose both low-frequency and high-frequency signals, and this decomposition has neither redundancy nor omission, so it can better analyze the time-frequency localization of signals containing a large amount of medium and high-frequency information. The frequency band is divided into multiple levels, and the high-frequency part of multi-resolution analysis without subdivision is further decomposed. According to the characteristics of the analyzed signal, the corresponding frequency band can be adaptively selected to match the signal spectrum, to improve the time-frequency resolution.

Wavelet denoising is to classify the decomposed wavelet coefficients by wavelet transform. After wavelet decomposition, the wavelet decomposition coefficient of useful signal is larger and that of noise is smaller. Therefore, by setting the threshold function for denoising, the decomposition coefficient greater than the threshold is retained, while the decomposition coefficient less than the threshold is eliminated by setting zero. Based on wavelet transform, wavelet packet decomposition refines and decomposes the untreated high-frequency components in wavelet transform again. Compared with wavelet denoising, wavelet packet decomposition has a higher frequency resolution, which can further eliminate the noise margin in the high-frequency part and improve the denoising accuracy.

According to Parseval's theorem, the total energy of a signal in its time domain is equal to that in the frequency domain. For wavelet packet decomposition, only the high-frequency and low-frequency components of the signal are separated, and the signal form has changed, but the total energy before and after decomposition remains equal.

2.2. Convolution, pooling, activation operation, batch normalization

2.2.1. convolution

Usually we are touching the convolution of one-dimensional signals, that is

$$\mathrm{y}[n] = x[n] * h[n] = \sum_{k} x[k] * h * [n-k], \tag{1}$$

In signal processing, $x[n]$ is the input signal and $h[n]$ is the unit response. The output signal $y[n]$ is then the delayed overlay of the input signal $x[n]$ response. This is also the nature of one-dimensional convolution: weighted

overlay/integral. For two-dimensional signals, such as images, the convolution formula is as follows:

$$y[m,n] = x[m,n] * h[m,n] = \sum_j \sum_i x[i,j] * h * [m-i, n-j], \tag{2}$$

It can be seen that two-dimensional convolution is also weighted overlay/integral. It should be noted that Convolution Kernel flips horizontally and vertically, and two-dimensional convolution is often used to extract spatial features.

2.2.2. pooling

The pooled layer derives new element values by sampling or aggregating information from a locally related set of elements based on the idea of local correlation. Pooling effectively reduces the parameters required by subsequent layers by reducing the dimension of the Feature Map. Moreover, the pooling operation can introduce translation invariance, which indicates that the output of the pooling layer is almost invariant when the elements in the input slightly shift in the neighborhood, thus enhancing the robustness of the network and having some anti-disturbance effect.

2.2.3. activation operation

The input and output of each layer in the neural network are a linear summation process. The output of the next layer only takes on the linear transformation of the input function of the previous layer. So if there is no activation function, no matter how complex the neural network is, the final output is a linear combination of inputs and cannot solve more complex problems. By introducing the non-linear activation function, the neural network can approximate any other non-linear function, thus enhancing the representation and learning ability of the network. The activation functions of neural networks usually have the following properties:

1. Continuous and differentiable in the definition domain, allowing a few points to be non-differentiable, the differentiable activation function can directly use the numerical optimization method to learn network parameters.
2. The activation function and its derivatives should be as simple as possible, too complex will reduce the efficiency of network calculation;
3. The derivative range of the activation function should be in an appropriate range, otherwise, it will affect the efficiency and stability of training.

2.2.4. batch normalization

The active input values of the deep neural network before making the non-linear transformation gradually shift or change with the depth of the network or during the training process. Generally, the training convergence is slow because the overall distribution gradually approaches both ends of the upper and lower limits of the value range of the non-linear function. Therefore, the gradient of the low-level neural network disappears when the backpropagation occurs. This is the essential reason for the slower and slower convergence of the deep training neural network. For each hidden layer neuron, batch normalization forces the input distribution that gradually maps to the non-linear function and closes to the limit saturation of the value interval back to the standard normal distribution with a mean of 0 and a variance of 1, so that the input values of the non-linear transformation function fall into the sensitive input area, thus avoiding the problem of gradient disappearance. In this way, a small change in input will result in a large change in the loss function, resulting in a larger gradient, avoiding the problem of gradient disappearance, and a larger gradient means faster learning convergence and faster training speed.

2.3. GRU for sequence modeling

Classic neural networks have difficulty in capturing the dependence of long time intervals in time series because of gradient disappearance or more common gradient explosion. Although gradient clipping can effectively solve the problem of gradient explosion, it cannot handle the disappearance of the gradient. Gated-loop neural network is proposed to better capture long-term time series dependencies in time series. It controls the flow of information by introducing reset gates and update gates, where reset gates determine how new input information is combined with previous memory, which defines how much previous memory is saved to the current time step memory. The reset and update doors are calculated as follows:

$$\begin{aligned} r_t &= \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \\ z_t &= \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right), \end{aligned} \tag{3}$$

and the gated neuron unit then assists in the calculation of the hidden state by calculating the candidate activation.

$$\begin{aligned} \tilde{h}_t &= \tanh\left(W \cdot \left[r_t * h_{t-1}, x_t\right]\right) \\ h_t &= \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t, \end{aligned} \tag{4}$$

where $\tilde{h}_t$ represents the candidate hidden state. The reset door controls how the last hidden state flows into the current candidate hidden state. The update door controls how the candidate hidden state containing the current time step information updates the current hidden state.

3. Diagnosis framework with Multi-layer Convolutional Gated Recurrent Unit Networks (MCGRU)

The overall structure of the proposed method is shown in figure **??**. First, the original sensor signal is divided into several time segments, each representing a time step in the cyclic neural network. The original time series signal is analyzed by wavelet packet to resolve the frequency domain close to that of the time domain, where the resolution of the frequency domain increases and the resolution of the time domain decreases. By improving the matrix multiplication operation in the original GRU cell to a convolution operation, the spatial relationship between the time-frequency information in one-time step and different sensors is preserved, and the forward propagation of the GRU network determines the temporal dependence of the above characteristics in two adjacent time steps. Both short-term spatial and long-term temporal dependencies derived from anomaly correspondence are updated to the hidden state in the MCGRU network unit. Feature extraction and learning are improved by updating the weights and offsets of network elements in different layers. The hidden state of the end time step in the final layer of the network is fully connected with the diagnostic results. The cross-entropy loss function is used to determine the error of fault diagnosis in each forward propagation process. The weight and bias values in the network are optimized by backpropagation.

3.1. Input layer

In the input layer, the length of the time series signal obtained by the sensor is determined by the sampling frequency and the monitoring time. Because the sensor sampling frequency is usually high, there is a lot of redundant information in the time domain. First, the signals obtained by each sensor are sliced using a time window of the same size, and the signals inside the window are converted to a two-dimensional time-frequency feature matrix using wavelet packet analysis. This method reduces the time domain resolution and enhances the frequency domain resolution, reduces the redundant information in the information domain, and increases the frequency domain information, making the features more diverse. By splicing the feature matrices of different sensors in the same time step in the depth dimension, a three-dimensional feature matrix with $T \times F \times D$ dimension is obtained, where T, F, and D are time domain resolution, frequency domain resolution, and several sensors respectively. This three-dimensional feature matrix represents the current time step T and fuses the time-frequency domain signal characteristics of different sensors. Because sampling frequency and monitoring time may differ between different sensors, to ensure the consistency of information dimensions in the time-frequency domain of different sensors, use 0 to fill in the missing part and assume that there is no prior knowledge about the filling part. Each feature matrix is then used as input to MCGRU to learn the short-term spatial correlation in each time step and the long-term temporal correlation between different time steps. The number of cells in the MCGRU layer is equal to $L/(T \times F)$.

3.2. Convolutional gated recurrent unit cell

FC-GRU uses full joins to stitch data into one-dimensional vectors in input-to-state and state-to-state transitions, resulting in the loss of spatial information when processing multidimensional data. To solve the above problem, data input per time step $X_1 \cdots X_t$, cell unit candidate hidden state $\tilde{h}_1, ..., \tilde{h}_t$, hidden state $H_1, ..., H_t$ and gated units r_i, z_i of the MCGRU is a three-dimensional matrix rather than a one-dimensional vector, where the first two dimensions represent the time-frequency characteristics of the time-domain resolution and the frequency-domain resolution, respectively. Convolution is used in the calculation of gated cells instead of direct connection. By neighborhood, the convolution in each gate fuses the high-dimensional data and updates the high-dimensional features to the hidden state in the MCGRU cells. The longer each time step is divided, the more information with short-term correlation can be stored in the three-dimensional feature matrix, as shown in the figure. Reset gates and update gates in MCGRU determine the current candidate hidden state through their input and past states of local neighborhood units and use

this candidate state to update the current hidden state. The key equations for the calculation process in MCGRU are as follows:

$$\begin{aligned} r_t &= \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \\ z_t &= \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \\ \tilde{h}_t &= \tanh\left(W \cdot \left[r_t * h_{t-1}, x_t\right]\right) \\ h_t &= \left(1 - z_t\right) * h_{t-1} + z_t * \tilde{h}_t, \end{aligned} \tag{5}$$

where $*$ denotes the convolution operator, [] denotes the Hadamard product, W_{xr}, W_{hr}, W_{xz}, W_{hz}, W_{xh}, W_{hh}, are all weight matrices, and the weights and offsets in the matrices are updated in each iteration.

If each hidden state is considered as a representation to reveal time series signal anomalies, MCGRU with a larger convolution core should be able to capture features with larger time-frequency variation, while networks with smaller convolution cores can capture features with smaller time-frequency variation. As the number of layers of the network increases, when the underlying hidden state is propagated vertically as the input of the upper layer, the pooling operation is used to extract the features at a higher level and reduce the dimensions. In the higher level of convolution operation, the size of the convolution kernel adapts to the dimension of the feature, so different network levels can obtain different levels of feature information. The whole process can be seen as the laterally propagating features of the same dimension in the circular neural network, and a vertical convolution neural network exists at each time step. The convolution neural network uses the hidden state of the output of the low-level circular neural network as input and the intermediate feature after the convolution operation as the input of the circular neural network in the next time step. The pooled and activated features are used as inputs to the next convolution neural network, and the whole process is shown in the figure.

3.3. Feature space

In the fault diagnosis model of many traditional convolution neural networks combined with the cyclic neural network, the spatial feature relationship is extracted using CNN as a feature extractor, or a matrix composed of images or signals. However, since the traditional circular neural network can only handle 1D vectors, it is necessary to stitch the extracted 2D spatial feature matrix into a 1D spatial feature vector. Compared with the traditional method of combining the CNN-RNN network, the MCGRU network is obtained by integrating CNN into the GRU network, which can create a 3D feature space, including 2D time-frequency feature and D sensor spatial feature. As shown in the diagram, the MCGRU unit can simultaneously extract the time-frequency characteristics of the fused multisensor from the 3D input and capture the long-term temporal dependence of the time-frequency characteristics.

3.4. Hidden state and classification layers

The hidden state of the end time step in the highest layer of the MCGRU network contains the time-frequency information after the multi-sensor fusion in all time steps. The final state is connected to the label of the training data through two layers of full connection. The output of the fully connected layer is predicted by a soft Max operation. The cross-entropy is used to measure the difference between the true distribution p(x) and the predicted distribution q(x) for the same random variable x, and the smaller the cross-entropy value, the closer the two distributions are. The cross-entropy formula is as follows:

$$H(p, q) = -\sum_{i=1}^{n} p\left(x_i\right) \log\left(q\left(x_i\right)\right), \tag{6}$$

where n represents the total number of possible events. Obviously, using q(x) to describe samples is not as accurate as p(x), and q(x) needs to be continuously learned to fit the exact distribution p(x).

From the formula, cross-entropy focuses on the prediction probability of the correct category. In classification problems, the output space of the model is a probability distribution, but the target output space is the category of the sample, that is, our ultimate goal is to get the correct category. Cross-entropy is the negative conditional logarithmic likelihood of the model distribution assuming that the model distribution is polynomial.

4. Experimental evaluation

Spatial features learnt by different network: t-Distributed Stochastic Neighbor Embedding (t-SNE) [] is applied to visualize the features in different levels of the models.

		Level 1	Level 2	Level 3	Level 4
Motor vibration					
Motor torque					
Gearbox vibration direction	planetary	X	Y	Z	N/A
	parallel	X	Y	Z	N/A
Fault type		Chipped tooth	Missing tooth	Root fault	Surface fault
Rotating speed - load		20-0	30-2	N/A	N/A

Table 1

Number of groups(Fault categories)	Number of tests in each group	Length of each test	Number of Sensors
10			8

Table 2
Data description.

MCGRU	Multi-layer convolutional gated recurrent unit networks (Proposed method)
ConvGRU	Convolutional structures fused into gated recurrent unit structure
CNN-GRU	CNN output to gated recurrent unit structure for sequence modeling input
GRU	Gated recurrent unit structure for sequence modeling
1D-CNN	Bidirectional gated recurrent unit structure for sequence modeling
CNN	Convolutional neural networks

Table 3
Baseline deep learning methods.

Activation heatmap of spatiotemporal features learnt by MCGRU: The heatmap is obtained via CAM to signify the important zones that are used for classification, where larger value (red in the colormap) indicates that the region is activated and used for classifying the corresponding condition (listed below each subfigure).

performance under noise environment: Add additive white Gaussian noise to the original signals to composite signals with different SNR, and the definition of SNR is shown as follows:

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{P_{\mathrm{signal}}}{P_{\mathrm{noise}}}\right), \tag{7}$$

where P_signal and P_noise are the power of signal and the noise respectively.

Necessity of BN, kernel dropout and ensemble learning: Experiments are conducted to compare the performances of networks trained with BN, without BN, and network that is trained with BN and has dropout in first convolutional kernels.

Performance across different load domain: When the target domain is becoming more different from the source domain, the result is supposed to be lower, because the target domain becomes harder to adapt to.

Networks visualizations: Try to explore the inner operating process of the proposed MCGRU model by visualize the activations in this neural network. 1.fist-layer convolutional kernels 2.all convolutional neuron activations 3.Feature visualization via t-SNE

Method	Window length	Kernel size	output channels	Hidden units	Dropout	number of layers
ConvGRU						
CNN-GRU						
GRU						
1D-CNN						
CNN						

Table 4
Hyperparameters for proposed and baseline methods.

4.1. Data description and Data preprocessing
4.2. Experimental environment
4.3. Case study
4.3.1. Experimental setup
4.3.2. Baseline
4.3.3. Performance on diagnosing untrained fault
4.3.4. Discussions
5. Conclusion
6. Bibliography

Two bibliographic style files (`*.bst`) are provided — model1-num-names.bst and model2-names.bst — the first one can be used for the numbered scheme. This can also be used for the numbered with new options of natbib.sty. The second one is for the author year scheme. When you use model2-names.bst, the citation commands will be like `\citep`, `\citet`, `\citealt` etc. However when you use model1-num-names.bst, you may use only `\cite` command.

`thebibliography` environment. Each reference is a `\bibitem` and each `\bibitem` is identified by a label, by which it can be cited in the text:
In connection with cross-referencing and possible future hyperlinking it is not a good idea to collect more that one literature item in one `\bibitem`. The so-called Harvard or author-year style of referencing is enabled by the LaTeX package natbib. With this package the literature can be cited as follows:

- Parenthetical: `\citep{WB96}` produces (Wettig & Brown, 1996).
- Textual: `\citet{ESG96}` produces Elson et al. (1996).
- An affix and part of a reference: `\citep[e.g.][Ch. 2]{Gea97}` produces (e.g. Governato et al., 1997, Ch. 2).

In the numbered scheme of citation, `\cite{<label>}` is used, since `\citep` or `\citet` has no relevance in the numbered scheme. natbib package is loaded by cas-sc with `numbers` as default option. You can change this to author-year or harvard scheme by adding option `authoryear` in the class loading command. If you want to use more options of the natbib package, you can do so with the `\biboptions` command. For details of various options of the natbib package, please take a look at the natbib documentation, which is part of any standard LaTeX installation.

A. My Appendix

Appendix sections are coded under `\appendix`.

`\printcredits` command is used after appendix sections to list author credit taxonomy contribution roles tagged using `\credit` in frontmatter.

CRediT authorship contribution statement

CV Radhakrishnan: Conceptualization of this study, Methodology, Software. **CV Rajagopal:** Data curation, Writing - Original draft preparation.

References

Author biography without author photo.