Industrial AI Algorithms: Selection and Application

In the process of selecting algorithms, we first consider the following factors: the sources of data, working conditions, clustering characteristics, fault modes, and expertise. As shown in Table 3.1, we suggest the corresponding algorithm selection strategies according to the above factors. When the data source is insufficient—for example if there is only a controller signal with lowsampling frequency—augmented learning can be used. This is when one uses third-party data acquisition and analysis equipment to obtain more signals and carry out targeted analysis. When there are enough data sources, deep learning can be considered to fully mine the implicit information in the data.

Looking at working conditions when the operation conditions of the system are complex, such as the various processing procedures of CNC machine tools or continuous alteration of heating and refrigeration in HVAC (heating, ventilation and air conditioning) systems, we give priority to clustering algorithms in the selection of working conditions. We can also do Fixed Cycle Feature Tests (FCFT) where the system is run repeatedly under fixed conditions during a fixed period. Once the corresponding data is collected, then the performance and decline of the system can be estimated according to the data.

The third consideration is clustering characteristics. If the object of analysis is clustered and there are many identical or similar devices in the monitoring state, then we can consider applying similarity learning or broad learning to analyze the relationship between the device and its adjacent devices, or use the information of these adjacent devices to estimate the status of the device. In addition, when the historical data of the system contains a large amount of fault mode information and corresponding historical data, we can use supervised learning to model the health and fault data of the historical data and analyze the collected signals. However, when the historical data of the system contains less fault mode information, we usually use unsupervised or semi-supervised learning methods only for health data modeling, and enrich the model according to needs in the monitoring process.

Table 3.1 Appropriate algorithms for industrial applications

	Usability	
	Low	High
Data sources	Augmented learning	Deep learning
Working conditions	General AI algorithms	Clustering algorithm, fixed working conditions
Clustering characteristics	General AI algorithms	Similarity learning, board learning
Fault mode history	Unsupervised/semi-supervised learning	Supervised learning
Expert experience	Deep learning, ensemble learning	Fuzzy logic

Expert experience is another factor to consider. Our knowledge of a system often helps us better understand that data and develop targeted data acquisition strategies. When we have a good understanding of the system's operation status, common failure modes, and mechanisms of occurrence, we often use AI algorithms such as fuzzy logic to introduce expert experience into the analysis model. When the expert experience of themonitoring object is scarce, we need to consider using deep learning algorithms to analyze the original data and understand the data patterns with the help of the algorithm. On the other hand, we often use ensemble learning to train several AI models at the same time, and give a comprehensive output based on the results of each model.

After determining the categories of algorithms, we must consider which specific algorithms we want to use in this category. As shown in Table 3.1, we consider the complexity and uncertainty of the system in two axes, and divide the object and data into four quadrants according to these two axes, labeled A, B, C, and D.

In area A, the complexity and uncertainty are low. For example, a single CNC machine tool uses the same spindle speed to produce a single product. In this case, we only need to consider general AI algorithms to use during data analysis to obtain good results.

In area B, the complexity is low while the uncertainty is high. The data in the monitoring process is vulnerable to noise or environmental factors, and the algorithm we choose needs to take this into consideration. If necessary, we can estimate the uncertainty so that the analysis model can adapt to it.

In area C, the uncertainty is low, but the complexity is high. This often happens due to complex working conditions, such as when cutting various materials, the different shapes of tools, and different branding needed on items. Hence, the selected algorithm must have high robustness and self-learning so it can adapt to different working conditions.

In area D, both complexity and uncertainty are high. In this case it is difficult to model the system data. We can use data to describe the state and behavior of the system by estimating its probability, such as by using the naïve Bayesian algorithm or by introducing expert experience to build a relational model.

As we consider the complexity and uncertainty of a system, we also need to consider the computational complexity of the algorithm and accuracy of the final results when selecting an algorithm. These two dimensions are also important in practical engineering applications. After defining the scope of several candidate algorithms, we can narrow down the scope through visualizations like Fig. 3.14. In Fig. 3.15, we also divide the space into four quadrants according to the design requirements of result accuracy and computational speed.
In area A, the accuracy of the results and the rate of computational speed are not very high, and the algorithm has no further requirements.

Fig. 3.14 Four-quadrant diagram of system complexity and uncertainty

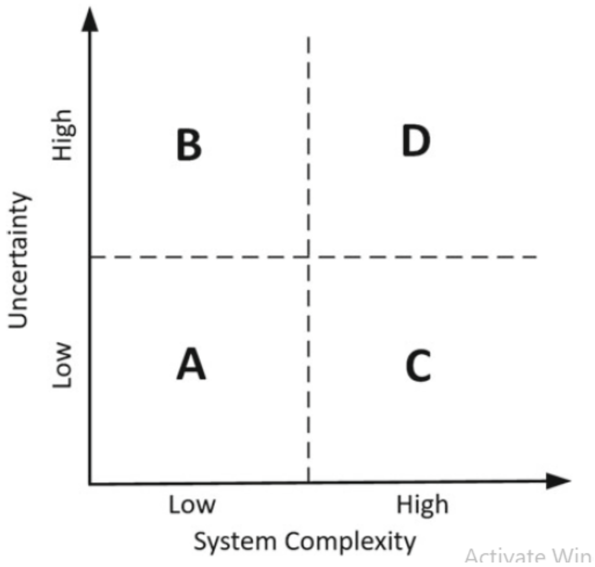
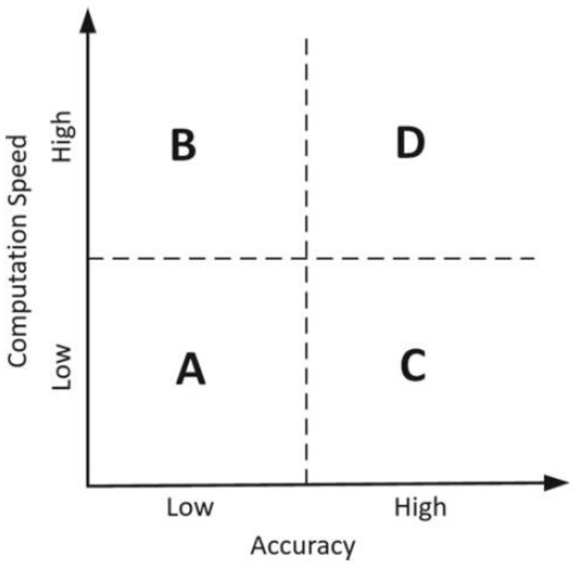


Fig. 3.15 Four-quadrant diagram of accuracy and computational speed



In area B, the accuracy of the results is low, but the computational speed is high. This area usually occurs in scenarios where the analysis results are displayed in real-time. For example, in semiconductor fabrication, the processing results of each wafer need to be evaluated in order to adjust parameters in real-time. Some fast convergent algorithms which are insensitive to the dimensionality of the data can be used including linear regression, logical regression, and so on.

In area C, the accuracy requirements are higher but the computational speed requirements are lower. Most application scenarios can be classified in this area. We hope to know the operation and health information of the system in the form of reports, but we need high result accuracy. Thus, we can choose algorithms like deep learning or ensemble learning to make the best usage of each model and utilize the advantages of each model in the data at the expense of computational speed in order to obtain maximum performance.

Region D indicates that there are high requirements for computational speed and accuracy. This is generally the case for situations where real-time control of a system is needed, and data analysis is quickly utilized. An example is routing multiple ships in real-time according to weather and ocean currents to avoid collisions. In this region, some rule-based algorithms based on expert experience often perform well such as the combination of fuzzy logic and genetic algorithms.