US-China Foreign Language, June 2026, Vol. 24, No. 6, 217-221
doi:10.17265/1539-8080/2026.06.002

Research on the Applications of Probability Theory in Artificial Intelligence

HAO Shun-li
Beijing International Studies University, Beijing, China

Artificial intelligence (AI) aims to simulate the thinking process of human perception, learning, and decision-making, but uncertainty problems, such as data noise, information loss, and semantic ambiguity are commonly present in real environments. Traditional deterministic mathematics is difficult to complete modeling and inference in complex scenarios. Probability theory, as a mathematical branch that studies the statistical laws of random phenomena, can quantify the likelihood of events occurring and provide a theoretical basis for uncertainty reasoning, parameter optimization, and distribution modeling for AI. This paper focuses on core probability knowledge, such as conditional probability, Bayesian formula, and probability distribution. It systematically discusses the specific applications of probability theory in traditional machine learning, deep learning, generative AI, natural language processing, computer vision, and reinforcement learning. It summarizes the supporting role of probability theory in the development of AI and looks forward to the research direction of the integration of the two in combination with current technological trends.

Keywords: probability theory, artificial intelligence, machine learning, deep learning

Introduction

In recent years, the continuous maturity of big data, computing power chips, and algorithmic frameworks has propelled the rapid popularization of Artificial intelligence (AI) technology. From intelligent recommendation, facial recognition to generative applications, such as large language models and AI painting, AI has penetrated into various industries. However, the real world is not a completely deterministic system: images captured by cameras are subject to illumination interference and noise, human language exhibits ambiguity due to polysemy, sensor data contain measurement errors, and environmental states are often incompletely observable. Facing with a vast amount of random and uncertain information, it is difficult to achieve effective intelligent judgment relying on deterministic calculations based on fixed formulas (Zhou, 2016; Ian, Yoshua, & Aaron, 2017).

The core value of probability theory lies in quantifying uncertainty, describing the likelihood of event occurrence with probability values, and providing a feasible approach for inducing patterns from data and inferring causal relationships in reverse. It can be said that the process of AI learning data patterns is essentially the process of using probability theory to fit the distributions of real-world data. From the earliest statistical learning models to today's large models with hundreds of billions of parameters, probability theory has permeated the entire process of algorithm design, model training, loss construction, sampling generation, and result evaluation. Organizing the application logic of probability theory in AI not only helps to understand the

HAO Shun-li, Ph.D., associate professor, Department of Basic Sciences, Beijing International Studies University, Beijing, China.

underlying principles of various AI algorithms, but also has significant implications for improving models' performance and enhancing models' interpretability.

Basic Theory of Probability Theory Commonly Used in AI

The design of AI algorithms cannot be separated from the basic knowledge system of probability theory, which mainly includes the following parts (Li, 2019).

Conditional Probability and Bayes' Formula

Conditional probability describes the probability of an event occurring given the occurrence of another event, and is the foundation of causal inference. Bayes' formula facilitates the conversion between forward probability and posterior probability, enabling the use of observational results to infer the cause of an event. It is the core principle of Bayesian learning, probabilistic graphical models, and filtering algorithms. The total probability formula decomposes complex events into several mutually exclusive sub-events, facilitating the modeling of mixed noise and multi-source data.

Random Variables and Probability Distributions

Discrete distributions include Bernoulli distribution, binomial distribution, and multinomial distribution, which correspond to binary classification, repeated trials, and multi-classification tasks respectively. Gaussian distribution is the most commonly used continuous distribution, and most data noise and neural network weight initialization follow Gaussian distribution. Uniform distribution and exponential distribution are mostly used for random sampling and parameter initialization.

Expectation, Variance and Covariance

Expectation reflects the average level of data and serves as the basis for loss function calculation and gradient descent optimization; variance characterizes the degree of data fluctuation and is commonly used in error analysis and overfitting suppression; covariance measures the correlation between different features and provides theoretical support for dimensionality reduction algorithms, such as principal component analysis.

In addition, the law of large numbers and the central limit theorem ensure the rationality of small-batch training, while probabilistic graphical models describe the dependencies between variables through graphical structures and are used for uncertainty reasoning in complex scenarios.

Applications of Probability Theory in Various Fields of AI

Applications in Traditional Machine Learning

Traditional machine learning, also known as statistical learning, the vast majority of classical algorithms are based on probability theory. The naive Bayes classifier, directly framed by Bayesian formula, utilizes the assumption of conditional independence of features to simplify computations, resulting in fast operational speeds. In early stages, it was extensively utilized for tasks such as spam email identification, text sentiment analysis and public opinion classification.

Gaussian Mixture Model (GMM) is a probabilistic clustering algorithm that fits data distribution by superimposing multiple Gaussian distributions and iteratively updates parameters using the EM algorithm. Compared to K-Means distance clustering, it can handle data with irregular shapes and complex data.

The Hidden Markov Model (HMM), which combines Markov chains and state transition probabilities, models time-series data and can infer hidden states from observed sequences. It has been widely used in natural language processing tasks, such as speech recognition, Chinese word segmentation, and part-of-speech tagging, and is a classic probabilistic model for time-series modeling.

Applications in Deep Learning

Deep learning can be understood as the stacking of multiple layers of probabilistic statistical models, with probability theory supporting every aspect of the training process.

In terms of parameter initialization and regularization, neural network weights are usually initialized according to a Gaussian distribution to ensure smooth initial gradients; Dropout introduces randomness to reduce overfitting by randomly dropping neurons with a certain probability.

In terms of optimization algorithms, mini-batch gradient descent relies on the law of large numbers, approximating the expected loss of the entire dataset with the loss expectation of a small batch of samples, achieving a balance between efficiency and accuracy. Bayesian neural networks further set weights as probability distributions rather than fixed numerical values, enabling the output of confidence intervals for prediction results. This compensates for the flaw of ordinary neural networks, which tend to be overly “confident” in their predictions, making them suitable for high-risk scenarios, such as medical diagnosis and autonomous driving.

Applications in Generative Artificial Intelligence (AIGC)

AIGC is the field where probability theory is most intensively applied. Its core idea is to learn the probability distribution of real data and then generate new samples through random sampling. Variational autoencoder (VAE) utilizes the KL divergence to constrain the latent variables to follow a standard Gaussian distribution, achieving the reconstruction and generation of images, text, and audio.

The Generative Adversarial Network (GAN) distinguishes the distribution differences between real samples and generated samples through a discriminator, and continuously narrows the distance between the two distributions in a game-theoretic manner, thereby generating more realistic images. The currently popular diffusion model is entirely based on Markov random processes. The forward process continuously adds Gaussian noise, while the backward process learns the conditional probability distribution to gradually denoise and restore the image, becoming the mainstream algorithm for AI painting and video generation. Top-k, Top-p sampling, and temperature coefficient adjustment in large models all involve modifying the probability distribution of words to control the fluency and diversity of generated content.

Applications in Natural Language Processing

Human language exhibits strong combinatorial randomness, and natural language processing has relied on probabilistic statistical modeling since its inception. Early statistical language models calculated the probability of a sentence occurring through conditional probability, assessing the fluency of sentences and laying the foundation for machine translation and text correction.

In the modern Transformer architecture, the Softmax layer converts the model output into a probability distribution over the vocabulary, enabling the prediction of the next word. The evaluation metric perplexity is a variation of information entropy, used to measure the uncertainty of a language model. A lower value indicates more coherent generated text. Essentially, the attention mechanism also involves assigning probability weights, allowing the model to focus on key contextual information. It can be said that the entire generation logic of large language models is token-by-token prediction based on probability.

Applications in Computer Vision

In the field of computer vision, probability theory is primarily utilized for noise processing, object detection and state tracking. Probabilistic filtering algorithms can be employed to smoothly remove salt-and-pepper noise and Gaussian noise during image acquisition. Object detection networks, such as YOLO and Faster R-CNN, can output class probabilities and object confidence scores. By setting probability thresholds, invalid boxes can be filtered out, thereby enhancing detection accuracy. Semantic segmentation networks make multi-class probability predictions for each pixel and then segment different object regions.

In video tracking tasks, Kalman filtering iteratively predicts the position of an object based on Bayesian estimation, and corrects the trajectory by combining observation probability and state transition probability. It is widely used in scenarios, such as security tracking and autonomous driving visual perception.

Applications in Reinforcement Learning

Reinforcement learning is framed around the Markov Decision Process (MDP) and is fundamentally built upon expectation and probability theory. The agent's policy network outputs a probability distribution for each action, striking a balance between exploring the environment and exploiting existing experience. The Q-value of value function, a mathematical expectation of long-term cumulative reward, is used to evaluate the quality of actions.

The Core Value of Probability Theory in AI

Firstly, uncertainty quantification processing is achieved. Probability theory enables AI to cope with noise, incomplete information, and ambiguous scenarios, breaking free from the limitations of deterministic computation and better adapting to the complex real world.

Secondly, objective criteria for model optimization are provided. Some probabilistic tools constitute a loss function system, which serves as the fundamental basis for model training iterations.

Thirdly, it supports the realization of generative intelligence. Probability distribution sampling and stochastic process deduction are the theoretical foundation of AIGC technology. Without probability theory, the creation of new content cannot be accomplished.

Fourth, it can enhance the model's generalization ability and robustness. Methods, such as regularization and random training suppress overfitting from a probabilistic perspective, making the model more stable on unfamiliar data.

Research Outlook

Although AI has achieved remarkable results, it still faces challenges, such as poor interpretability, difficulties in small-sample learning, and susceptibility to learning spurious correlations. Probability theory holds the key to breaking through these bottlenecks. In the future, Bayesian probabilistic intelligence will become an important research direction in trustworthy AI, relying on prior knowledge and posterior inference to enhance small-sample learning capabilities. Meanwhile, the combination of probability theory and causal inference can guide models to explore causal patterns, rather than merely learning data associations, endowing AI with stronger cognitive abilities. Furthermore, advanced probability theories such as measure probability theory and stochastic processes will also play a greater role in large model training, diffusion model optimization, and general AI research.

Conclusion

Probability theory serves as an indispensable mathematical cornerstone for AI, and it also serves as the "thinking language" through which AI comprehends the world. From simple classification and prediction to complex generation and creation, the entire process of AI learning, reasoning, and decision-making is underpinned by probabilistic thinking. As AI continues to evolve towards generalization, trustworthiness and strong cognition, the integration of probability theory and AI will become more intimate. Deep grasp of the principles of probability theory and the understanding of probabilistic modeling are essential to truly comprehend the inherent logic of AI algorithms and drive continuous optimization and innovation in intelligent technology.

References

Ian, G., Yoshua, B., & Aaron, C. (2017). *Deep learning*. (S. J. Zhao, Trans.). Beijing: Posts & Telecom Press.

Li, H. (2019). *Statistical learning methods*. Beijing: Tsinghua University Press.

Zhou, Z. (2016). *Machine learning*. Beijing: Tsinghua University Press.