Part I

General discussion

1 Introduction

1.1 Machine learning and speech and language processing

Speech and language processing is one of the most successful examples of applying machine learning techniques to real problems. Current speech and language techniques embody our real-world information processing, automatically including information extraction, question answering, summarization, dialog, conversational agent, and machine translation (Jurafsky & Martin 2000). Among these, one of the most exciting applications of speech and language processing is speech recognition based voice search technologies (by Google, Nuance) and conversational agent technologies (by Apple) (Schalkwyk, Beeferman, Beaufays *et al.* 2010). These successful applications started to make people in general casually use speech interface rather than text interface in mobile devices, and the applications of speech and language processing are widely expanding.

One of the core technologies of speech and language processing is automatic speech recognition (ASR) and related techniques. Surprisingly, these techniques are fully based on statistical approaches by using large amounts of data. The machine learning techniques are applied to utilize these data. For example, the main components of ASR are acoustic and language models. The acoustic model (AM) provides a statistical model of each phoneme/word unit, and it is represented by a hidden Markov model (HMM). The HMM is one of the most typical examples of dealing with sequential data based on machine learning techniques (Bishop 2006), and machine learning techniques provide an efficient method of computing a maximum likelihood value for the HMM and an efficient training algorithm of the HMM parameters. The language model (LM) also provides an n-gram based statistical model for word sequences, which is also trained by using the large amount of data based on machine learning techniques. These statistical models and their variants are used for the other speech and language applications, including speaker verification and information retrieval, and thus, machine learning is a core component of speech and language processing.

Machine learning covers a wide range of applications in addition to speech and language processing, including bioinformatics, data mining, and computer vision. Machine learning also covers various theoretical fields including pattern recognition, information theory, statistics, control theory, and applied mathematics. Therefore, many people are studying and developing machine learning techniques, and the progress of machine learning is rather fast. By following the rapid progress of machine learning, researchers in speech and language processing interact positively with the machine learning community or communities in the machine learning application field by importing (and sometimes exporting) advanced machine learning techniques. For example, the recent great improvement of ASR comes from this interaction for discriminative approaches (recent progress summaries for discriminative speech recognition techniques are found in Gales, Watanabe & Fossler-Lussier (2012), Heigold, Ney, Schluter *et al.* (2012), Saon & Chien (2012*b*), Hinton, Deng, Yu *et al.* (2012). The discriminative training of HMM parameters has been mainly studied in speech recognition research since the 1990s, and became a standard technique around the 2000s. In addition, the deep neural network replaces the emission probability of the HMM from the Gaussian mixture model (GMM) (or is used as feature extraction (Hermansky, Ellis & Sharma 2000, Grézl, Karafiát, Kontár *et al.* 2007) for the GMM) and achieves further improvement on the discriminative training based ASR performance. Actually, current successful applications of speech and language processing are highly supported by these breakthroughs based on the discriminative techniques developed through the interaction with the machine learning community. By following the successful experience, researchers in speech and language processing try to collaborate with the machine learning community further to find new technologies.

1.2 Bayesian approach

This book also follows the trend of tight interaction with the machine learning community, but focuses on another active research topic in machine learning, called the *Bayesian approach*. The Bayesian approach is a major probabilistic theory that represents a causal relationship of data. By dealing with variables introduced in a model as probabilistic variables, we can consider uncertainties included in these variables based on the probabilistic theory.

As a simple example of uncertainty, we think of statistically modeling several data (x_1, $x_2, \cdots, x_N$) by a Gaussian distribution $\mathcal{N}(x|\mu, \Sigma)$ with mean and variance parameters μ and Σ, as shown in Figure 1.1, i.e.,

$$p(x) \approx \mathcal{N}(x|\mu, \Sigma). \tag{1.1}$$

Now, we consider the Bayesian approach, where the mean parameter is uncertain, and is distributed by a probabilistic function $p(\mu)$. Since μ is uncertain, we can consider the several possible μs instead of one fixed μ, and the Bayesian approach considers representing a distribution of x by several Gaussians with possible mean parameters (μ_1, μ_2, and μ_3 in the example of Figure 1.1),

$$p(x) \approx \frac{1}{N} \sum_{\mu=\{\mu_1,\mu_2,\cdots,\mu_N\}} \mathcal{N}(x|\mu, \Sigma), \tag{1.2}$$

where μ_1, μ_2, $\cdots$ are generated from the distribution $p(\mu)$. The extreme case of this uncertainty consideration is to represent the distribution of x, which is represented by