음성인식에서 부분공간가우스혼합모형에 기초한 언어모형화방법

리현순, 리정철

위대한 령도자 **김정일**동지께서는 다음과 같이 교시하시였다.

《나라의 과학기술을 세계적수준에 올려세우자면 발전된 과학기술을 받아들이는것과 함께 새로운 과학기술분야를 개척하고 그 성과를 인민경제에 적극 받아들여야 합니다.》 (《**김정일**선집》 증보판 제11권 138～139페지)

선행연구[1, 2]에서 제안한 재귀신경망언어모형(RNNLM)은 언어의 장거리문맥특징정보들을 반영할수 있는 모형이다. 그러나 이 모형에서는 특정한 분야의 적은 학습자료를 리용하여 성능을 더욱 높일수 있는 언어모형의 과제적응방법들이 제시되여있지 않다.

한편 음성인식음향모형화에서 많이 리용되는 가우스혼합모형들은 최대우도(ML)학습, 최대분류오유률(MCE)학습, 최대호상정보(MMI)학습방법과 최대사후확률적응(MAP), 최대우도선형회귀(MLLR)적응 그리고 제한된 최대우도선형회귀특징공간(fMLLR)적응과 같은 우월한 방법들을 가지고있다. 이러한 방법들은 음소를 단어로, 발성자를 주제로 바꾸면 그대로 언어모형화에 적용할수 있다.

론문에서는 장거리문맥정보를 모형화할수 있는 RNN을 리용하여 단어들의 문맥정보를 반영하고 그러한 문맥정보들의 분포를 부분공간가우스혼합모형으로 모형화하는 방법을 제안하였다.

1. 부분공간가우스혼합모형에 기초한 언어모형화(SGMLM)

부분공간가우스혼합모형은 적은 파라메터로 특징공간을 치밀하게 표현할뿐아니라 적은 학습자료를 가지고도 충분히 학습할수 있는 능력으로 하여 이미 음성인식음향모형화에서 가우스혼합모형(GMM)보다 많이 리용되고있다.

련속공간모형의 한 형태로서 부분공간가우스혼합모형에 기초한 언어모형화에서는 매 단어(음향모형에서의 상태에 대응)가 어휘단어공간의 일반배경모형이 만드는 어떤 생성부분공간에 놓이도록 한다.

이 모형화방법에서 어떤 단어 w가 주어진 조건에서 리력 y가 출현할 확률밀도함수는 다음과 같이 표시된다.

$$p(y \mid w) = \sum_{c=1}^{C_w} \rho_{wc} \sum_{i=1}^{I} w_{wci} N(y;\, \mu_{wci},\, \Sigma_i) \tag{1}$$

$$\mu_{wci} = M_i v_{wc} \tag{2}$$

$$w_{wci} = \frac{\exp(w_i^{\mathrm{T}} v_{wc})}{\sum_{k=1}^{I} \exp(w_k^{\mathrm{T}} v_{wc})} \tag{3}$$

$$\sum_{c=1}^{C_w}\rho_{wc}=1 \tag{4}$$

여기서 C_w는 단어 w의 부분상태수(단어 w를 상태공간의 1개 상태로 생각한다.)를, v_{wc}와 $\rho_{wc}(\geq 0)$는 각각 그 단어의 부분상태벡토르와 상태무게를 나타낸다.

어떤 단어 w에 대한 확률밀도함수는 I개의 가우스분포를 가진 GMM이지만 공분산행렬 $\sum_i$는 모든 단어들간에 공유되고 혼합성분무게 w_{wci}와 평균 μ_{wci}는 부분상태벡토르 v_{wc}와 넘기기 M_i, w_i로부터 유도된다. 그러므로 매 단어의 GMM파라메터들이 M_i, w_i가 만드는 전체 파라메터공간의 어떤 부분공간으로 제한된다.

부분공간가우스혼합언어모형의 파라메터학습과 여러가지 적응수법들은 이미 음향모형화에서 충분히 서술되였다. 매 단어를 한 상태로 이루어진 HMM으로 보면 선행연구[4]에 서술된 추정식들을 리용하여 표현할수 있다.

실지 언어모형에서 리용하는 확률 $P(w|h)$는 베이스의 정리에 의하여 다음과 같이 표시된다.

$$P(w|h)=\frac{P(h|w)P(w)}{P(h)}$$

여기서 $P(h)\approx p(y)$, $P(h|w)\approx p(y|w)$, $P(w|h)\approx p(w|y)$가 성립된다.

따라서

$$p(w|y)=\frac{P(w)p(y|w)}{p(y)}=\frac{P(w)p(y|w)}{\sum_{v=1}^{V}P(v)p(y|v)} \tag{5}$$

이다.

2. 파라메터무리짓기

파라메터무리짓기는 음향모형구축에서와 마찬가지로 언어모형학습코퍼스에 출현빈도가 낮은 단어들이 많이 존재하므로 중요한 문제이다. 따라서 음향모형구축에서 대표적으로 쓰이는 파라메터무리짓기방법인 Top－Down방법을 리용한다. 여기서는 왼쪽 문맥에 어떤 단어들이 놓이는가 하는 질문에 따라 단어들을 구분한 다음 한 무리에 속한 파라메터들을 무리짓기한다.

음향모형에서는 기본모형화단위가 음소인것으로 하여 수십개 정도이지만 언어모형에서는 수만개정도이므로 언어학적인 질문들은 단어들의 결합에 대한 사전지식에 기초하여 작성될수 있다. 단어 그자체는 크게 명사, 대명사, 부사, 형용사, 동사, 앞붙이, 뒤붙이, 토 등과 같은 품사정보들로 분류될수 있다.

론문에서는 먼저 음성인식에 합리적으로 설정된 96개의 품사정보들을 리용하여 단어들을 품사별로 분류하였다. 다음 매 부류의 단어들은 그앞에(또는 그앞에) 어떤 품사들이 놓이는가 하는 질문렬에 따라 분류하였다. 이때의 분류는 학습자료의 로그우도를 최대로 하는 질문을 선택하여 진행된다.

품사정보를 반영한 192개의 질문을 가지고 무리짓기를 진행하기로 한다.

3. 성 능 평 가

RNNLM과 SGMLM학습자료로서 44 000 000개의 단어로 구성된 품사표기붙은 신문자료를 리용한다. 생성된 어휘수는 65 000개이다. 언어모형적응자료는 학습에 리용되지 않은 450 000개의 단어들을 포함하고있는 15 000개의 문장으로 되여있다.

평가자료 1은 학습자료와 같은 분야에서 4 200 000개의 단어들로 이루어진 170 000개의 문장들을 포함하고있으며 학습에는 리용되지 않는다.

평가자료 2는 적응자료와 같은 분야에서 취한 130 000개의 문장을 포함하고있다.

비교를 위한 기준언어모형으로서 N－그람언어모형, 가우스혼합에 기초한 언어모형(GMLM), 재귀신경망언어모형(RNNLM)을 설정하였다.

조선어련속음성인식체계 《룡남산》을 복호기로 리용하였다. 음성인식률평가실험에서 N－그람언어모형은 복호기의 첫번째 통과에서, GMLM과 SGMLM은 두번째 통과에서 리용되였다. 음성인식률평가실험은 평가자료 1과 평가자료 2에서 각각 30개의 문장들을 임의로 선택하여 남, 녀발성자(각각 5명씩) 10명이 발성한 음성자료를 가지고 진행하였다.

여러가지 언어모형들의 성능평가를 표에 보여주었다.

표. 여러가지 언어모형들의 성능평가

모형		적응전 test1/test2의 분기수 (음성인식률/%)	적응후 test2의 분기수 (음성인식률/%)		
			MAP	MLLR	CMLLR
N－그람모형		145.2/187.5(97.78/93.51)	148.4(96.70)		
GMLM		156.8/176.1(97.76/94.1)	146.7(96.74)	150.5(96.68)	143.6(96.76)
RNNLM		98.5/153.7(98.30/95.8)			
SGMLM	Top－Dwon	110.6/125.7(98.20/95.30)	119.6(97.44)	115.9(97.52)	111.3(97.62)
	Bottom－Up	102.7/130.8(98.28/95.07)	118.6(97.45)	118.6(97.45)	115.7(97.58)

맺 는 말

RNN의 장거리문맥정보와 부분공간가우스혼합모형의 모형화능력을 결합하기 위한 방법과 부분공간가우스혼합모형의 효률적인 추정을 위하여 Top－Down방법으로 파라메터들을 무리짓기하는 방법을 제안하였다. 실험을 통하여 제안한 방법들이 단어들사이의 장거리문맥정보를 모형화하면서도 새로운 분야에로의 과제적응을 실현할수 있다는것을 확인하였다.

참 고 문 헌

[1] **김일성**종합대학학보(자연과학), 63, 4, 36, 주체106(2017).

[2] L. H-S et al.; ICASSP'11, 5524, 2011.

주체107(2018)년 8월 5일 원고접수

Language Modeling based on Subspace Gaussian Mixture Model in Speech Recognition

Ri Hyon Sun, *Ri Jong Chol*

In this paper, we have proposed the method for combining longer context information of RNN with modeling ability of subspace gaussian mixture model(SGMM) that has been widely used in acoustic modeling for speech recognition.

Key words: language modeling, speech recognition, recurrent neural network, subspace gaussian mixture model