

입말체변환에 기초한 자연발화음성언어 모형화에 대한 연구

리 혁 철

선행연구들에서는 막힘기호의 인위적인 삽입이나 병렬코퍼스를 리용하는 방식의 자연발화음성언어모형화방법들을 제안하였다.

론문에서는 자연발화음성인식언어모형을 구축하기 위한 방법을 제안하였다.

1. 문제 설정

언어모형화는 자연발화음성인식에서 중요한 문제의 하나로서 화제와 함께 자연발화음성에서 나타나는 입말체표현들도 적절히 모형화하여야 한다. 리용가능한 입말체본문 특히 믿음직한 발성표기자료의 양은 일반적으로 발성표기비용문제로 하여 제한을 받는다. 그러므로 일반적으로 일부 자연발화음성코퍼스를 화제특정의 문서자료기지와 결합하는 방법을 리용한다.

선행연구에서는 글말체본문에 막힘기호를 임의로 삽입하여 모방된 입말체본문을 생성하는 방법을 제안하였다. 이러한 합성에 기초한 방법은 간단하면서도 효과적이지만 결과모형에는 ASR성능을 저하시키는 관련없는 언어적표현들도 포함되게 된다.

선행연구[1]에서는 믿음직한 발성표기와 해당한 문서형식의 본문들의 병렬코퍼스를 리용하여 N -그램통계들을 직접 추정하였다. 변환모형이나 번역모형은 보통 작은 코퍼스로 학습되므로 효율적인 모형화가 절실히 요구된다.

론문에서는 통계적기계번역에서 쓰이는 입말체변환방법[2]과 품사정보에 기초한 모형을 적용하여 자연발화음성인식언어모형화를 실현하는 방법을 제안하였다.

2. 언어모형의 통계적변환

통계적기계번역방법의 언어모형변환에서는 원천언어의 문장 X 로부터 목적언어의 문장 Y 의 생성확률 $P(Y|X)$ 를 베이스규칙에 기초하여 최대화하도록 한다.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

문서형식과 발성언어를 각각 X 와 Y 로 표시되는 서로 다른 언어라고 보고 발성언어모형 $P(Y)$ 를 추정하면 식 (1)로부터 다음과 같이 형식화할수 있다.

$$P(Y) = P(X) \frac{P(Y|X)}{P(X|Y)} \quad (2)$$

조건부확률 $P(Y|X)$ 와 $P(X|Y)$ 즉 변환모형은 믿음직한 발성표기와 해당한 문서형식본문으로 이루어진 병렬코퍼스를 리용하여 추정할수 있다.

언어모형변환공정을 그림에 보여주었다.

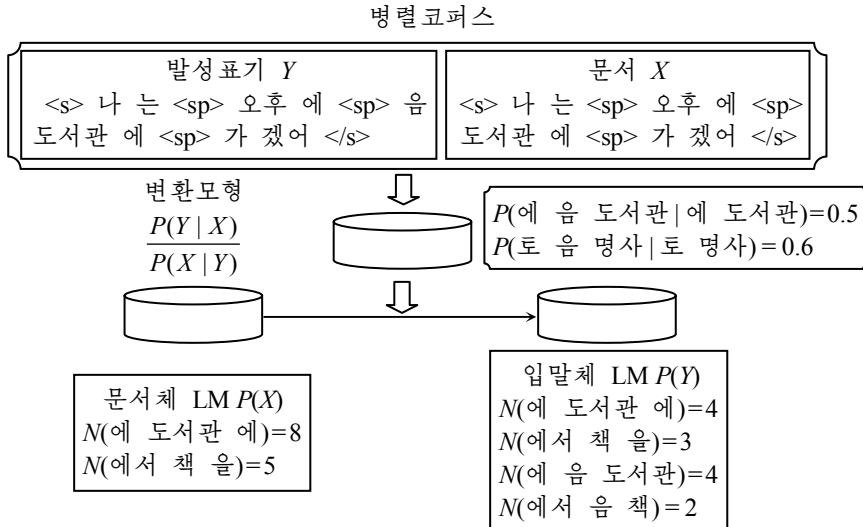


그림. 언어모형변환공정

1) 변환모형학습

단어열들의 출현빈도를 리용하여 단어변환모형을 직접 학습시킨다. 병렬코퍼스를 리용하여 문서형식의 단어열의 빈도 $N(x)$ 와 해당하는 입말체단어열의 빈도 $N(y \leftarrow x)$ 를 계산하고 다음 단어변환확률 $P_{\text{word}}(y|x)$ 를 추정한다.

$$P_{\text{word}}(y|x) = \frac{N(y \leftarrow x)}{N(x)} \quad (3)$$

단어통계들을 리용하면 자료성김문제에 쉽게 부닥치게 되므로 보다 넓은 어휘피복과 믿음직한 추정을 위하여 품사정보에 기초한 모형을 적용한다. 형태부해석기를 리용하여 매 문맥단어에 품사정보를 할당하고 품사패턴들에 해당하는 변환확률 $P_{\text{pos}}(y|x)$ 를 식 (3)과 같은 방식으로 추정한다.

2) 모형추정

모형추정은 다음과 같은 방법들로 진행한다.

① back-off방법

단어모형과 품사모형을 결합하기 위한 가장 간단한 방법으로 back-off방식을 적용하였다.

$$P(y|x) = \begin{cases} P_{\text{word}}(y|x), & x \rightarrow y \text{가 존재하는 경우} \\ P_{\text{pos}}(y|x), & x \rightarrow y \text{가 존재하지 않는 경우} \end{cases} \quad (4)$$

이 방법에서는 매 단어패턴을 차례로 N -그람들에 적용하며 그 패턴이 정합되는 경우 $P_{\text{word}}(y|x)$ 에 의하여 변환을 진행한다. 만일 정합되지 않는 경우에는 품사패턴을 적용한다.

② 선형보간방법

back-off외 다른 방법으로 두 모형의 선형보간방법을 적용한다. 이 방법에서는 단

어확률과 품사확률들의 무게화된 합을 변환확률로 리용하며 단어패턴과 품사패턴중 어느 하나가 본래의 N -그람과 일치되는 경우 변환을 진행한다.

$$P(y|x) = \lambda P_{\text{word}}(y|x) + (1-\lambda)P_{\text{pos}}(y|x) \quad (5)$$

③ 최대엔트로피(ME)방법

back-off방법과 선형보간방법에서는 단어모형과 품사모형이 따로따로 추정되어 후에 결합된다. 어휘정보와 품사정보를 통합방식으로 론의하기 위하여 최대엔트로피방법을 적용하였다.

이 방법에서는 조건부확률 $P(y|x)$ 를 다음과 같이 결정한다.

$$P(y|x) = \frac{1}{z} \exp \left[\sum_i \lambda_i f_i(x, y) \right] \quad (6)$$

여기서 f_i 는 특징함수이고 λ_i 는 특징의 무게이며 z 는 정규화인자이다. 특징 $\{f_i\}$ 들로서 앞단어와 다음단어, 그것들의 품사정보들을 리용하였다. 문서형식모형의 매 N -그람에 ME모형을 적용하며 변환확률이 턱값보다 큰 경우에는 입말형식의 N -그람을 생성한다.

3. 평가 실험

기준 3-그람언어모형(LM0)의 학습용으로 로동신문 3년분(2010-2012)코퍼스를 수집하였다. 변환모형학습용병렬코퍼스를 작성하기 위하여 크기가 520K인 영화문학교퍼스를 준비하였다.

검사모임으로서 단어사전크기가 43K인 극문학교퍼스를 선정하고 비교를 위해 영화문학교퍼스와 기준모형을 결합한 언어모형(LM1)을 준비하였다.

론문에서 제안한 입말체변환모형을 리용하여 back-off와 선형보간, 최대엔트로피모형들(LM2, LM3, LM4)을 생성하고 분기수와 단어오유률의 견지에서 평가하였다.

분기수와 단어오유률평가결과를 표에 보여주었다.

표. 분기수와 단어오유률평가결과

모형	분기수	단어오유률/%
LM0	82	14.4
LM1	77.9	13.8
제안모형	LM2	54.7
	LM3	55.0
	LM4	57.4

표에서는 각이한 모형들의 분기수와 단어오유률을 보여주고있다. 변환모형을 리용한 모든 모형들은 분기수를 크게 감소시킨다. 기준모형(LM0)과 비교모형(LM1)에 비한 분기수감소는 LM2에서 각각 35.0%와 30.6%이다. back-off모형과 선형보간모형의 성능은 거의 같으며 최대엔트로피모형의 성능이 약간 낮아진다. 그것은 변환확률들이 일반적으로 최대엔트로피추정에 의하여 보다 작아지기때문이다.

음성인식실험결과 표에서와 같이 비교모형(LM1)과 비교해볼 때 변환된 모형들(LM2, LM3)은 단어오유률을 상대적으로 2.0%만큼 개선하였다.

맺 는 말

자연발화음성인식용언어모형을 효과적으로 생성하는 새로운 통계적변환방법을 제안하였으며 피복률과 모형의 민음성을 높이기 위하여 품사정보를 리용하였다. 그리고 평가 실험을 통해 제안된 방법에 기초한 back-off와 선형보간모형들의 분기수감소와 단어오류를 개선하였다.

참 고 문 헌

- [1] Y. Akita, T. Kawahara; Proc. ICASSP, 1, 1049, 2005.
- [2] Y. Akita, T. Kawahara; IEEE Transactions on Audio, Speech, and Language Processing, 18, 6, 1539, 2010.

주제107(2018)년 8월 5일 원고접수

Study on Language Modeling of Spontaneous Speech Based on Transform to the Spoken Style

Ri Hyok Chol

For language modeling of spontaneous speech, we apply the approach based on the statistical machine translation framework, which transforms a document-style model to the spoken style. We investigated several methods that combined POS-based model or integrated POS information in the maximum entropy scheme, and verified the improvement in perplexity and WER through experimental evaluation.

Key words: spontaneous speech, language model, automatic speech recognition