IMPROVING SPEECH RECOGNITION ACCURACY OF LOCAL POI USING GEOGRAPHICAL MODELS

Songjun Cao, Yike Zhang, Xiaobing Feng, Long Ma

Tencent Cloud Xiaowei

January, 2021

Overview

Overview

background

POI (Point of Interests) search with voice is prevailing.

Figure: Car navigation

Figure: Mobile phone map

main challenges & our solutions

Main challenges

- Multiple dialects: dialects vary from geographical regions to geographical regions.
- Long-tailed distribution of POI names: it's hard for language model to model infrequent POI names.

Our solutions

- Multi-dialect problem: one single acoustic model with geographic-specific components (Geo-AM).
- Long-tailed POI names: a group of geographic-specific language models (Geo-LMs).

Geo-AM

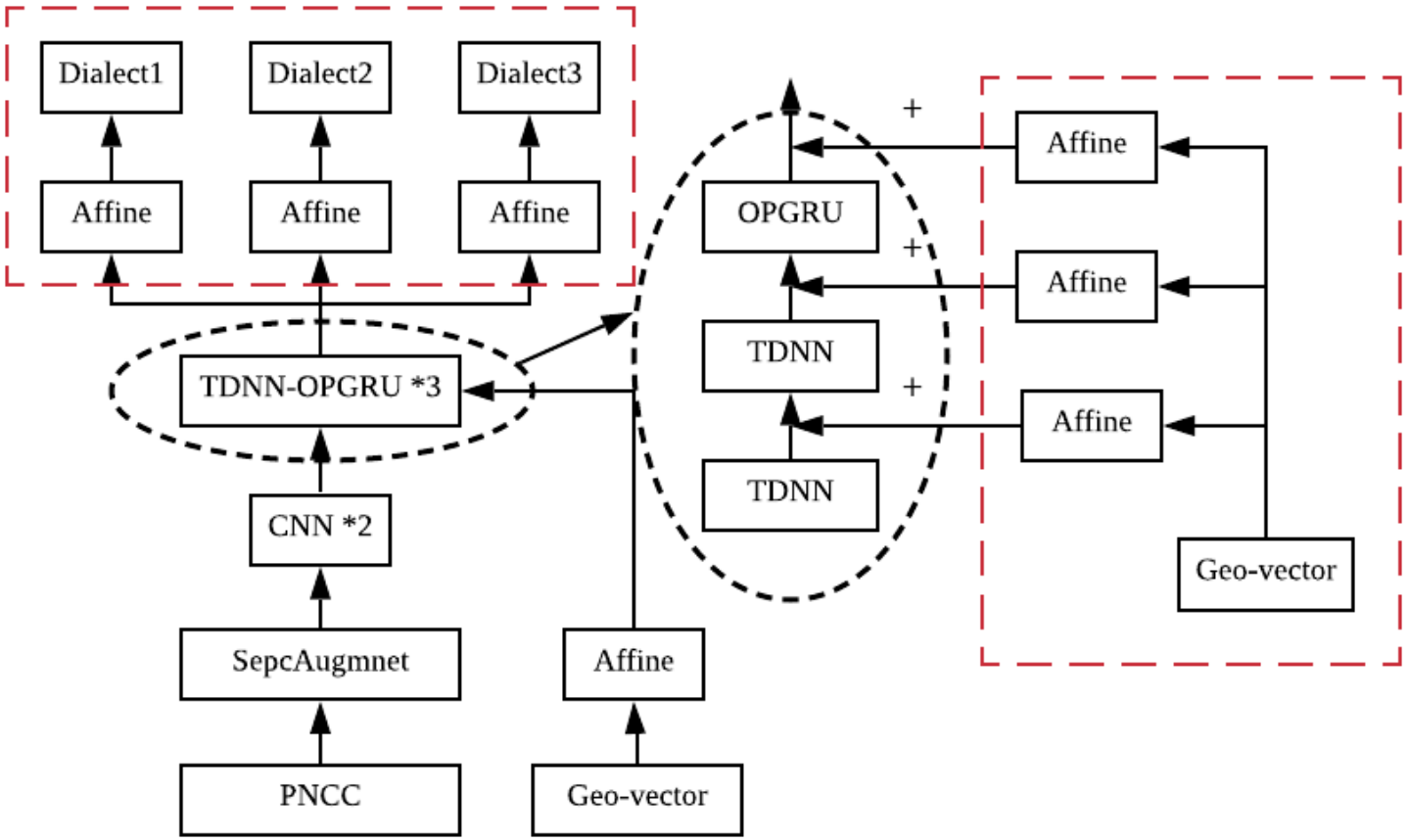

Figure: Geo-AM.

Geo-LMs

$$HCLG_{bi} \circ G_{bi}^{-} \circ G_b \circ G_l \tag{1}$$

$$P^{(1)}(w|h) = \lambda P_b^{(1)}(w|h) + (1-\lambda) P_l^{(1)}(w|h) \tag{2}$$

$$P^{(2)}(w|h) = \alpha P_b^{(2)}(w|h) + \beta P_r^{(2)}(w|h) + (1-\alpha-\beta) P_l^{(2)}(w|h) \tag{3}$$

$$P(w|h) = \gamma P^{(1)}(w|h) + (1-\gamma) P^{(2)}(w|h) \tag{4}$$

- $P_l^{(1)}(w|h)$ is a Geo-LM in first-pass decoding
- $P_l^{(2)}(w|h)$ is Geo-LM in n-best rescoring

Results

Table: CER (%) on the test set of using dialect-specific input (A1) and dialect-specific top layer (A2), integrating Geo-LMs in first-pass decodeing (L1), rescoring n-best lists of the first-pass decoding output without Geo-LMs (L2), integrating Geo-LMs in n-best rescoring (L3).

Province	Baseline	A1	A2	L1	L2	L3
Jiangsu	5.92	6.11	5.9	5.57	5.27	5.06
Zhejiang	4.78	4.73	4.25	4.11	3.94	3.85
Sichuan	3.69	3.44	3.38	3.06	3.2	3.18
Shandong	3.75	3.74	3.6	3.54	3.05	3.06
Henan	5.05	4.87	4.74	4.46	4.19	4.57
Liaoning	4.33	4.02	3.94	3.44	3.36	3.0
Guangdong	5.9	5.54	5.5	5.22	4.86	4.68
Shaanxi	4.98	4.87	4.55	4.45	4.0	3.88
Anhui	4.92	4.82	4.56	4.73	4.26	4.16
Hebei	3.93	3.7	3.65	3.43	3.37	3.11
Total	4.70	4.58	4.38	4.17	3.99	3.82

- The proposed Geo-AM and Geo-LMs totally achieve 18.7% relative CER reduction on the POI voice search task of Tencent Map.

Contributions

- Dialect-specific top layers in Geo-AM make it possible to optimize a certain dialect while maintaining the performance of other dialects.
- Either Geo-AM or Geo-LM for a specific dialect region could be optimized separately, which is more flexible for a production service.
- Geo-LMs are incorporated into the process of both first-pass Viterbi search and N-best rescoring.

The end of highlights

Overview

Geo-AM - dialect-specific input

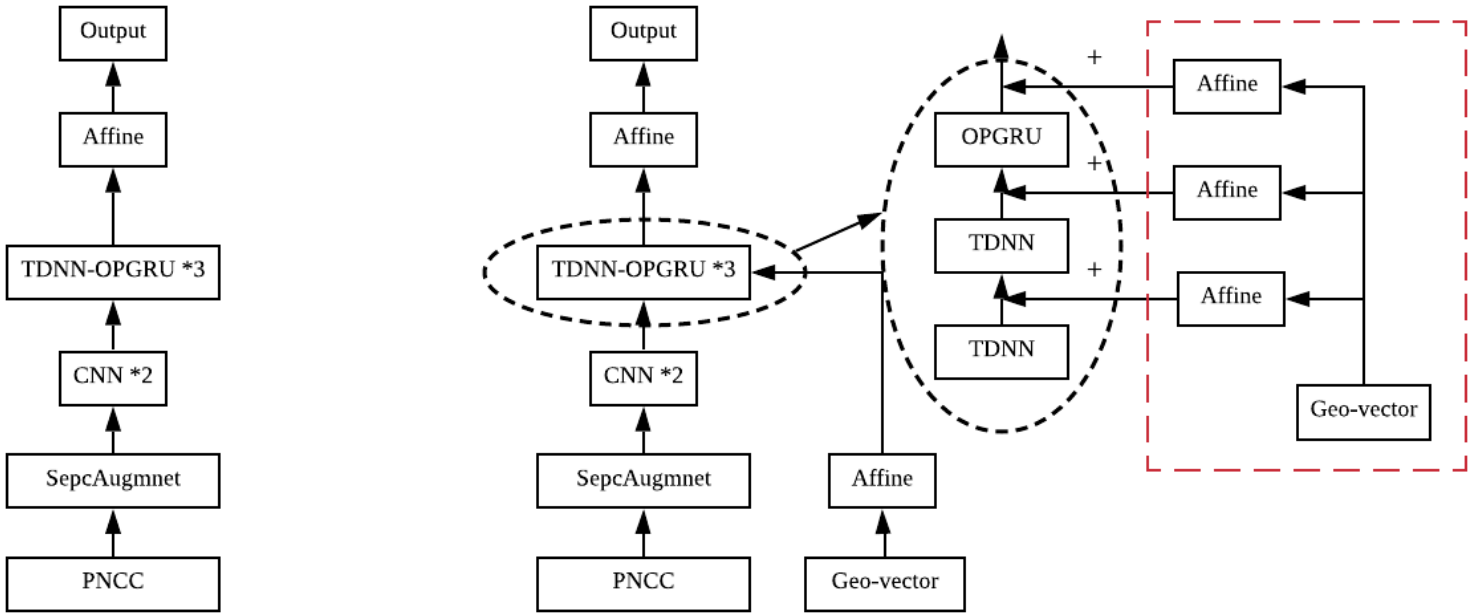

Figure: A0(left) is the baseline AM; A1(right) is A0 with dialect-specific input.

Drawback

It is hard for A1 to improve the accuracy of a specific dialect while maintaining the performance on other dialects.

Geo-AM - dialect-specific top layer

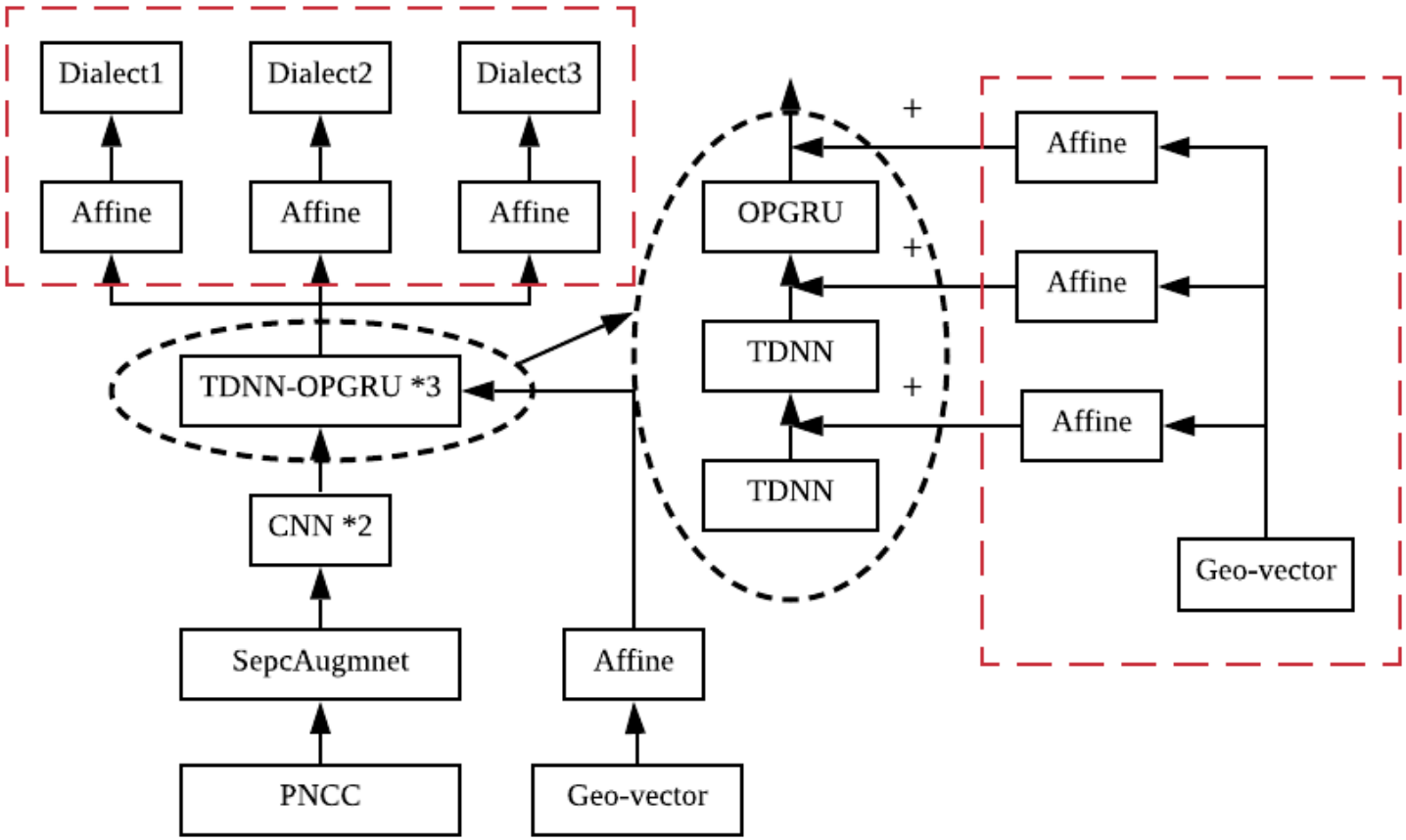

Figure: A2 is A1 with dialect-specific top layer.

Overview

Geo-LMs - first-pass decoding

$$HCLG_{bi} \circ G_{bi}^{-} \circ G_b \circ G_l \tag{5}$$

$$P^{(1)}(w|h) = \lambda P_b^{(1)}(w|h) + (1-\lambda) P_l^{(1)}(w|h) \tag{6}$$

- $\circ$ denotes on-the-fly composition
- G_b is the 5-gram baseline LM
- G_l is a Geo-LM
- G_{bi} consists only uni-grams and bi-grams of G_b
- G_{bi}^{-} is negated score version of G_{bi}
- $P_b^{(1)}(w|h)$ is the probability from G_b
- $P_l^{(1)}(w|h)$ is the probability from G_l
- λ controls the contribution of different LMs

Geo-LMs - n-best rescoring

$$P^{(2)}(w|h) = \alpha P_b^{(2)}(w|h) + \beta P_r^{(2)}(w|h) + (1 - \alpha - \beta) P_l^{(2)}(w|h) \quad (7)$$

- $P_b^{(2)}(w|h)$ is the probability from the character-level baseline LM
- $P_r^{(2)}(w|h)$ is the probability from the QRNN model
- $P_l^{(2)}(w|h)$ is the probability from a character-level Geo-LM
- α and β control the contribution of different LMs

The final language probability is

$$P(w|h) = \gamma P^{(1)}(w|h) + (1 - \gamma) P^{(2)}(w|h) \quad (8)$$

Overview

Experiment setups

- Both training and test data are collected from our POI voice search production, Tencent Map.
- Only one fifth training data (4k hours) have region information.
- Test data are collected from users in top-10 provinces with most traffic.
- Baseline 5-gram models and the QRNN model are trained with 1,200M POI names collected from Tencent Map.
- Geo-LMs are trained with local POI names in corresponding provinces collected from Tencent Map and the Internet.

Results

Table: CER (%) on the test set of using dialect-specific input (A1) and dialect-specific top layer (A2), integrating Geo-LMs in first-pass decodeing (L1), rescoring n-best lists of the first-pass decoding output without Geo-LMs (L2), integrating Geo-LMs in n-best rescoring (L3).

Province	Baseline	A1	A2	L1	L2	L3
Jiangsu	5.92	6.11	5.9	5.57	5.27	5.06
Zhejiang	4.78	4.73	4.25	4.11	3.94	3.85
Sichuan	3.69	3.44	3.38	3.06	3.2	3.18
Shandong	3.75	3.74	3.6	3.54	3.05	3.06
Henan	5.05	4.87	4.74	4.46	4.19	4.57
Liaoning	4.33	4.02	3.94	3.44	3.36	3.0
Guangdong	5.9	5.54	5.5	5.22	4.86	4.68
Shaanxi	4.98	4.87	4.55	4.45	4.0	3.88
Anhui	4.92	4.82	4.56	4.73	4.26	4.16
Hebei	3.93	3.7	3.65	3.43	3.37	3.11
Total	4.70	4.58	4.38	4.17	3.99	3.82

- The proposed Geo-AM and Geo-LMs totally achieve 18.7% relative CER reduction on the POI voice search task of Tencent Map.
- Introducing Geo-LMs into the process of n-best rescoring can achieve better results.

Results

Table: CER(%) and CERR(%) of Geo-AMs on datasets with different level of accent.

Level	**A0**	**A2**	**CERR**
Serious	11.30	10.16	10.1
Medium	9.46	8.67	8.4
Slight	5.21	4.87	6.5
None	3.72	3.54	4.8

- Geo-AM can alleviate the accent problem and achieve 6.5%∼10.1% relative CER reduction on test sets of different accent levels.

Thanks for listening!
Q&A

Appendix A: division of dialect regions for Geo-AM

Table: China is split into 10 dialect regions based on similarity of dialects and distribution of users.

Region ID	**Included provinces**
1	Zhejiang Jiangsu
2	Sichuan Chongqing Guizhou
3	Shandong Henan
4	Heilongjiang Jilin Liaoning
5	Guangdong
6	Shanxi Gansu Shaanxi
7	Hunan Hubei Anhui
8	Yunnan Guangxi Fujian
9	Beijing Tianjin Hebei
10	Others

Appendix B: experimental results of Geo-AM

Table: CER (%) of Geo-AMs on the development set.

Region	**A0**	**A1**	**A2**
1	4.93	4.73	**4.63**
2	6.41	5.96	**5.60**
3	5.23	4.93	**4.80**
4	4.64	3.96	**3.89**
5	4.98	4.91	**4.86**
6	6.07	6.05	**5.60**
7	6.21	6.03	**5.71**
8	6.61	6.69	**6.56**
9	4.01	3.79	**3.73**
10	5.75	**5.73**	5.87
Total	5.37	5.14	**5.02**

Appendix C: experimental results of Geo-LMs

Table: CER (%) on the development set of integrating Geo-LMs in first-pass decodeing (L1), rescoring n-best lists of the first-pass decoding output without Geo-LMs (L2), integrating Geo-LMs in n-best rescoring (L3).

Province	**A2**	**L1**	**L2**	**L3**
Guangdong	4.88	4.56	4.35	4.42
Henan	4.53	4.24	4.08	3.70
Shandong	5.10	4.91	4.74	4.27
Jiangsu	4.64	4.40	4.18	3.73
Zhejiang	4.75	4.18	4.32	3.88
Gansu	7.84	5.56	6.15	5.26
Hainan	7.16	6.97	8.26	6.97
Ningxia	7.19	6.47	6.29	6.12
Xizang	5.88	5.88	5.88	4.24
Qinghai	5.08	3.86	4.47	3.25
Nationwide	5.02	4.51	4.48	3.90