Tools for Rapid Customization of S2S Systems for Emergent Domains

Rohit Kumar, Matthew E. Roy, Sanjika Hewavitharana, Dennis N. Mehay, Nina Zinovieva

Raytheon BBN Technologies
10 Moulton Street, Cambridge, MA 02138, USA
{rkumar, mroy, shewavit, dmehay, nzinovie}@bbn.com

Abstract

Component models of speech-to-speech translation (S2S) systems need to be customized to emerging needs. In this demonstration, we will showcase the technical functionality of BBN's domain customization tools for S2S systems that allow subject matter experts to augment an existing S2S system with new vocabulary items and translation rules using a web-based user interface. To reduce the user effort and time, our tools leverage Wikipedia as a linguistic resource for enrichment of domain profiles by finding lexical items and translations related to the domain. In a recent evaluation of BBN S2S system customized for using these tools, we found 15% (relative) reduction in word error rate as well as 30% (relative) reduction in untranslatable words when used within customized conversational domains.

Index Terms: Speech-to-Speech Translation, Domain customization, Web as a resource

1. Introduction

Over the past decade, considerable progress has been made in developing usable, two-way speech-to-speech (S2S) translation systems that enable real time cross-lingual spoken communication [1]. State-of-the-art S2S systems comprise a number of advanced technology components that perform Automatic Speech Recognition (ASR), Statistical Machine Translation (SMT) and Text-to-Speech Synthesis (TTS). These components are trained on large volumes of data that is collected to represent anticipated domains of use.

On the other hand, unforeseen scenarios for the use of such systems emerge on a regular basis. Examples of emergent domains of use include natural disasters (e.g. landslides due to torrential rains), humanitarian relief (e.g. refugee processing) and medical emergencies (e.g. the recent outbreaks of middle-eastern respiratory syndrome, or MERS). Unfortunately, the performance of system components suffers dramatically when they are used in domains that have limited or no coverage within the training data. In particular, the ASR component will fail to recognize previously unseen (out-of-vocabulary or "OOV") words. Furthermore, novel meanings of known words in new domains may not be translated accurately. These failures cause non-communication and miscommunication of critical concepts in S2S mediated cross-lingual interaction.

Collecting large volumes of speech and text corpora in response to emergent domains may not be feasible, given that a S2S system would need to be deployed in a matter of hours or days, rather than weeks or months. The objective of our current work is to enable rapid extension of the conversational repertoire of S2S systems to new domains.

We achieve this objective by developing workflows, discussed in Section 3, as well as technologies that allow bilingual subject matter experts (SMEs) who are not S2S system developers to customize S2S systems. We have created web-based tools, presented next, that operationalize an efficient customization workflow within which SMEs can specify phrase pairs new to the emergent domain.

2. Domain Customization Tools

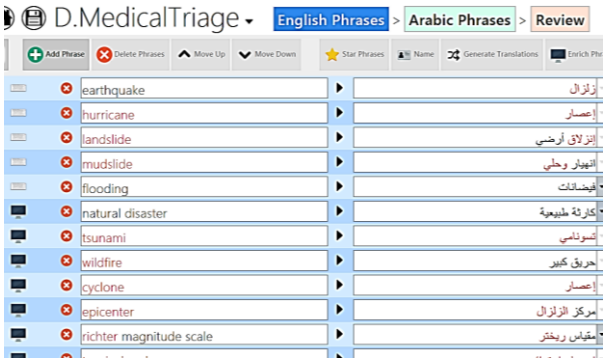

Figure 1. Customization Tool: Editing a domain profile

The domain customization tool, referred to as *Customizer*, is a web based application that allows a customization team to create and maintain domain profiles. Figure 1 shows an example of domain profile for Medical Triage. One of the use-cases for this profile is performing triage following an earthquake. Users can specify a number of phrase pairs in both translation directions. OOV words in the customization are automatically highlighted (e.g. *landslide*) which helps users identify which words are critical to improving ASR performance in the new domain.

Besides functionality for organizing and editing phrase pairs, the *Customizer* includes a number of advanced features to facilitate the customization process. Automatic translations of phrases are made available to reduce the manual effort of the bilingual speaker. Automatic translations can be edited when necessary. Users can enrich manually entered phrases using automatic phrase enrichment which uses Wikipedia to find keywords and phrases in articles about the manually

This work was funded by the DARPA BOLT program under contract number HR0011-12-C-0014. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Approved for Public Release, Distribution Unlimited.

specified phrase being enriched. The automatic enrichment module also uses Wikipedia articles available in multiple languages to find translations of novel phrases.

Once a domain profile is created, a model update corresponding to the profile is generated. Model update involves addition of new words to the ASR vocabulary and extending the ASR language model with n-grams comprising the new words. Finally, translation rules that augment the baseline translation models are generated from the phrase. Model update is fully automated and does not require the user to be an ASR or SMT expert. Updated models instantly become available for download from within BBN's S2S application deployed in the field. This helps in reducing the response time for supporting emergent needs.

Currently the *Customizer* has been developed & configured for the English-Iraqi Arabic language pair to support the evaluation discussed ahead. The tools can be extended to other language pairs.

3. Customization Workflow

Domain customization starts with the description of a domain which includes listing known use-cases as well as anticipable information such as names, terminology, key phrases. This information may be provided to or compiled by the domain expert who will be performing the customization. The domain expert then starts creating a domain profile using the *Customizer* and populates it with phrases directly identified in the description. Iteratively, these phrases are enriched both manually as well as by using the automatic enrichment capability of the *Customizer*. In the third step, the bilingual speaker, in coordination with the domain expert, adds translations to the phrases specified by the expert. New phrases may be added in this step. After this, the bilingual speaker flips the direction of the phrase pairs and copies them over to start the customization process in the other direction. This operation is facilitated by the customization tools. At this time, additional phrases either from the domain description or otherwise may be added in the other direction. Finally, phrase pairs in both directions are reviewed and refined by the customization team before the domain profile is processed and published. Note that the domain expert and bilingual speaker may be the same person in some cases. Each step in this workflow could be iterated over or reviewed by additional domain experts and bilingual speakers depending on the time and personnel available. Furthermore, as the customized profile is used in the field, additions and corrections to the profile may be performed as needed.

4. Evaluation

In January 2015, BBN participated in NIST's evaluation of S2S systems under the DARPA BOLT program. For this evaluation, two customization teams, each comprising a domain expert and a bilingual English-Iraqi Arabic speaker were able to create 11 customized domain profiles in less than 18 working hours (<2 hours per profile). The teams were provided domain descriptions generated by NIST. BBN research team members filled in for the role of domain experts. The bilingual speakers were BBN employees who have several years of experience in manual translation of spoken and text corpora.

Six speaker pairs, each comprised of one native English speaker and one native Iraqi-Arabic speaker, were tasked with communicating with one another using a domain customized S2S system. The speakers were placed in two rooms separated by a see-through glass wall that served as a sound barrier. Each speaker pair interacted with each system over a 90-minute-long sessions. Each session included up to 12 conversations ranging across the 11 customized domains. These domains were grouped into high level categories: Disaster Relief; Humanitarian Aid; General Topics like sports and family. In addition to these, each speaker pair had one conversation about a surprise open-ended domain (e.g. cooking, fashion). The open-ended domain differed from the other domains in that both the speakers and systems did not have any prior knowledge of this domain. Besides the lack of coverage in the customizations, the speakers had highly free-formed conversations in that category.

Table 1. Performance using Domain Customized Models

		Without Customized Models	With Customized Models	Change
ASR Word Error Rate				
English	Disaster Relief	7.7	6.6	14%
	Humanitarian Aid	7.1	5.9	17%
	General Topics	12.5	10.2	18%
	Open-Ended	13.3	11.8	11%
Iraqi	Disaster Relief	8.7	8.0	8%
	Humanitarian Aid	5.9	5.4	8%
	General Topics	7.1	5.8	18%
	Open-Ended	10.2	9.3	9%
Untranslatable Words in ASR Output (words per 1000)				
E2I	Disaster Relief	5.3	2.9	83%
	Humanitarian Aid	2.7	1.8	50%
	General Topics	1.2	9.0	33%
	Open-Ended	6.9	5.2	33%
I2E	Disaster Relief	6.6	5.3	25%
	Humanitarian Aid	8.1	7.3	11%
	General Topics	3.9	3.9	0%
	Open-Ended	4.9	4.9	0%
Translation Edit Rate				
E2I	Disaster Relief	55.5	55.5	0.0%
	Humanitarian Aid	54.2	54.2	0.0%
	General Topics	61.0	60.4	1.0%
	Open-Ended	64.4	62.8	2.5%
I2E	Disaster Relief	51.6	51.5	0.2%
	Humanitarian Aid	49.4	48.8	1.2%
	General Topics	51.0	51.0	0.0%
	Open-Ended	52.7	52.6	0.2%

In total, the six speaker pairs had 69 conversations. User speech files as well as ASR, SMT and TTS outputs for each turn were logged for analysis. Table 1 summarizes the improvement in system component performance using the model updates generated from the customized domain profiles. We find noticeable improvements in ASR and SMT performance using the customized models. In addition to the improvement in translation performance shown above, a subjective evaluation using reference transcriptions of user speech showed a preference for translations generated using customized models.

5. References

[1] R. Kumar, M. Roy, S. Ananthakrishnan, S. Hewavitharana, F. Choi, "Interactive Error Resolution Strategies for Speech-to-Speech Translation Systems", *SigDial 2015, 142-144*