CALL FOR PARTICIPATION

The Blizzard Machine Learning Challenge 2017

Evaluating corpus-based speech synthesis on common databases

Keiichi Tokuda, Simon King, Alan W Black, and Kei Sawada

January 2017

In order to better understand and compare research techniques in building corpus-based speech synthesisers on the same data, the annual Blizzard Challenges 2005-2016 were held. We are now pleased to call for participation in the twelfth challenge: Blizzard Challenge 2017.

In the HMM era, by taking a unified view of both Automatic Speech Recognition (ASR) and Text-to-Speech (TTS), it was possible to develop various types of new ASR and TTS techniques, e.g., cross-lingual speaker adaptation, adaptive training for TTS, use of prosody in ASR, etc. We expect that by once again taking a unified view in the current DNN era, it will be possible to develop new types of acoustic modeling techniques that are useful for both ASR and TTS.

The series of Blizzard Challenges has helped us measure progress in TTS. But, to get competitive performance, a lot time has to be spent on skilled tasks such as updating the lexicon, removing inappropriate audio files, segmenting and aligning audio files, detecting alignment errors, etc. This may make the Blizzard Challenge unattractive to Machine Learning (ML) researchers from other fields.

We therefore propose a spin-off challenge that does not involve these speech-specific tasks, and allows participants to concentrate on the acoustic modeling task, framed as a straightforward ML problem, with a fixed data set.

The data that the organizers will provide is in the form of corresponding sequences of linguistic features, speech features and speech waveforms. Participants must train a model to predict speech features from linguistic features (or, to directly predict speech waveforms from linguistic features, as done in WaveNet), and then use that model to make predictions for a test set of previously-unseen linguistic features.

The organizers will conduct a large scale subjective evaluation test, similar to that in Blizzard Challenge (naturalness, intelligibility, and speaker similarity).

The following is an outline description of the challenge – for full details and the rules of participation, please see the web pages.

- **Spoke task 2017-ES1:** Predict speech features from linguistic features
- **Spoke task 2017-ES2:** Directly predict speech waveforms from linguistic features

Blizzard Workshop: The systems will be presented by the participants at the ASRU 2017.

Website: http://synsig.org/index.php/Blizzard_Challenge_2017

Registration: Interested parties should register as soon as possible, by mailing blizzard@festvox.org. They should identify a contact person in their team, and provide email and postal addresses. A registration fee is payable for each system submitted for evaluation – details can be found on the website.

Licenses: The license for the released data can be found via the Blizzard website. Data will be released to each participant once the appropriate license has been agreed to.

Mailing list: It is essential to join the blizzard-discuss mailing list, which is used to send important information to participants. Instructions are on the Blizzard website.

Timeline: http://synsig.org/index.php/Blizzard_Challenge_2017#Timeline

Further Information: For further information please contact blizzard@festvox.org. A description of the challenge was published at Interspeech 2005: http://www.festvox.org/blizzard/bc2005/IS051946.PDF

Please *only* use the address blizzard@festvox.org to contact the organisers of the challenge.