### Recurrent Neural Network LMs

- RNNLMs [1] outperform n-grams in many ASR tasks due to the following:
  - RNNLMs allow robust parameter estimation through a continuous-space representation
  - RNNLMs can model longer context dependencies than n-grams
- Recurrent layer can represent full history \( \langle w_{1-}, \ldots, w_{i-1} \rangle \) for word \( w_i \) using concatenation of word \( w_{i-1} \) and remaining context vector \( \langle w_{i-2} \rangle \)

#### Multi-Genre Broadcast Data

- The data were BBC broadcasts and subtitles officially distributed for the MGB challenge [2]
  - Acoustic training data: 2,193 hours with 1,580 hours of audio and lightly supervised transcripts
  - Language training data: 648M words from historical subtitles (LM1) and 104M words from 2,193 hours training shows (LM2)
- Development data: 47 hours with 28 hours of audio
- 8 genres: Advice, children’s, comedy, competition, documentary, drama, events and news

#### Model-Based RNNLM Adaptation

- The following two model adaptation techniques were compared:
  - Genre fine-tuning, which involves further training a RNNLM using genre-specific text data
  - Linear hidden network (LHN) adaptation layer, which introduces a linear multiplicative transform to the hidden layer to adapt to genre-specific text data
- The adaptation layer is cascaded between the hidden and output layers respectively
  - The weights connecting the adaptation and the output layers are initialised using the identity matrix
  - At time of adaptation, only those weights are updated whilst keeping the rest of the network unchanged
- We are the first to apply LHN adaptation layer to RNNLMs

#### Hybrid RNNLM Adaptation

- The following two hybrid adaptation techniques were proposed:
  - Fine-tuning feature-based RNNLM, which involves further training LDA adapted RNNLMS on genre-specific text, thus combining topic and genre domain representations
  - Feature-Based RNNLM with adaptation layer, which involves having an adaptation layer with genre 1-hot features input, together with LDA feature input at the hidden layer
  - Feature-based adaptation layer provides an additive transform through bias adaptation whilst LHN adaptation layer provides a multiplicative transform of the weights at the hidden layer
  - Adaptive transform was shown to be less prone to over-fitting in acoustic domain
  - Overfitting can happen when amount of domain-specific data is small, which is the case for genres such as comedy and drama

#### Feature-Based RNNLM Adaptation

- Append a feature vector \( f \) to the input of the RNNLM
- Two features used in this work:
  - Genre 1-hot auxiliary codes, which represent genre as a \( \mathbf{i} \)-of-\( \mathbf{i} \) vector
  - Latent Dirichlet Allocation (LDA) auxiliary features, obtained by computing Dirichlet posteriors over latent topics after training models by first computing term frequency-inverse document frequency (TF-IDF) vectors on text data

#### Semi-supervised RNNLM Adaptation

- Genre labels are available for LM1 text but not for larger LM2 text
- In order to make the best of the hybrid adaptation techniques, need to generate genre labels for LM1 text
  - LDA features with 100 topics were extracted from LM12 text and a SVM classifier was used to predict genre from the LDA features
  - Classification accuracy obtained on held-out development data was 91.79%
  - Same LDA+SVM model was used to predict genre labels for LM2 text, which was then used for hybrid RNNLM adaptation

### Experiments and Results

- DNN-GMM-HMM Bottleneck acoustic models [3]
- 200k vocabulary used to build baseline n-gram LM on LM1 + LM2 text
  - Trained both LM1 and LM1 + LM2 RNNLMS
- Used a modified version of RNNLM toolkit [4]
- Gained improvements with hybrid RNNLM adaptation compared to previous work [5]
  - LDA topic features and genre 1-hot features were found to be complimentary
  - LHN Adaptation Layer gives improvements over fine-tuning
  - Adaptation layer with additive transform gives better results than with multiplicative transform

### References


