Incorporating Knowledge into Statistical Acoustic Models for SLDS

• Spoken Language Dialogue System (SLDS)

• Fundamental technology:
  Automatic speech recognition (ASR)

• State-of-the-art ASR based on statistical approaches
• Commercially used in controlled conditions
• Not ready for use by any speaker, any language in any environment
  – performance drops if conditions change
  – more representative data required to improve performance,
  – relying solely on statistical approaches with more data is not sufficient
Research Questions

- Enhance robustness of statistical ASR by incorporating various additional knowledge sources whilst keeping training and recognition effort limited

- Scope problems:
  - how to incorporate various knowledge sources from different domains in an efficient and unified way?
  - at which level to incorporate this knowledge?
  - how to solve the increase in model complexity due to incorporating a significant number of additional knowledge sources?
Methods

• Proposed Approach: GFIKS (graphical framework to incorporate additional knowledge sources)

• New efficient general framework designed by utilizing Bayesian network (BN)

• Advantages:

  1. **Probabilistic relationship between any information sources can be learned:** any knowledge may be incorporated at different level of ASR in unified way

  2. **Provides a way to decompose the global probability function:** simplify model complexity, so it can be reliably estimated using limited resources