UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Semantic Analysis of Spoken Input Using Markov Logic Networks

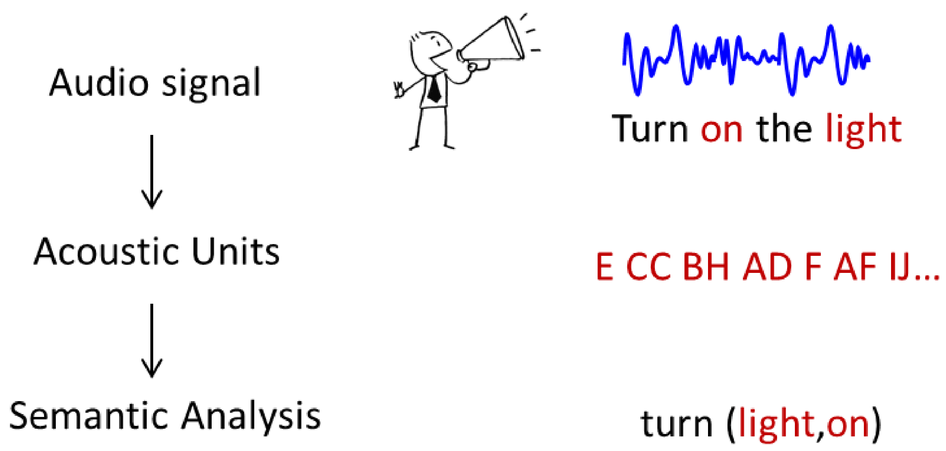
Vladimir Despotovic[1], Oliver Walter[2] and Reinhold Haeb-Umbach[2]

[1]University of Belgrade, Serbia
vdespotovic@tf.bor.ac.rs

[2]University of Paderborn, Germany
{walter,haeb}@nt.uni-paderborn.de

Introduction

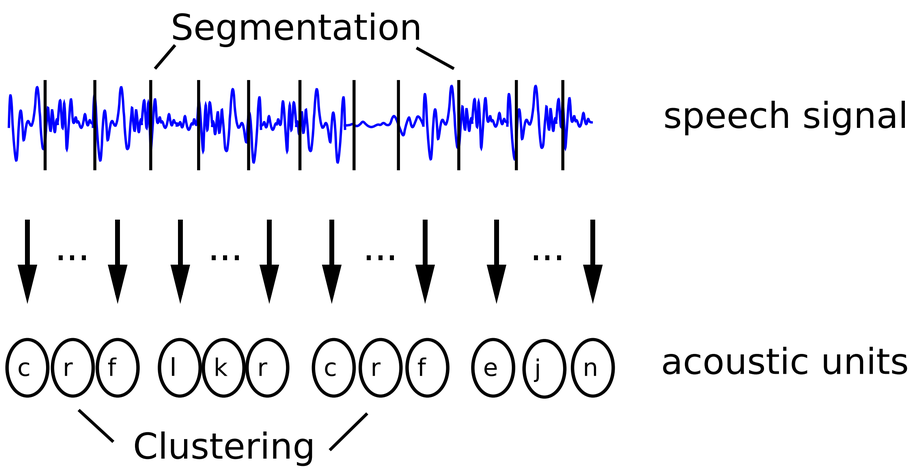
- **Objective:** Semantic analysis of spoken language
  - Decode spoken command into a subword unit sequence
  - Learn subword unit models in an unsupervised fashion
  - Map the subword unit sequence to semantics



- **Challenge:** Learning in the presence of noisy and inconsistent input data

Acoustic representation

- Segmentation of the speech signal into audio segments
- Clustering of similar segments into segment labels (acoustic units - AUDs)
- Iterative training of HMM models for AUDs



Markov logic networks

- **Markov Logic Networks:**
  - Combine first-order logic and Markov networks in a single representation
  - Weighted first-order formulas
- **Probability distribution:**

$$P(X=x) = \frac{1}{Z}\exp\left(\sum_{i=1}^{F}\omega_i \sum_{g\in G_i} g(x)\right) = \frac{1}{Z}\exp\left(\sum_{i=1}^{F}\omega_i n_i(x)\right)$$

- **Example:** a set of first-order clauses for mapping of AUD sequences to semantic frames (PATCOR dataset)

$$\begin{aligned}
&HasAUD(+a,u) \Rightarrow FromSuit(+s,u)\\
&HasAUD(+a,u) \Rightarrow FromValue(+v,u)\\
&HasAUD(+a,u) \Rightarrow TargetSuit(+s,u)\\
&HasAUD(+a,u) \Rightarrow TargetValue(+v,u)\\
&HasAUD(+a,u) \Rightarrow FromFoundation(+f,u)\\
&HasAUD(+a,u) \Rightarrow TargetFoundation(+f,u)\\
&HasAUD(+a,u) \Rightarrow DealCard(u)
\end{aligned}$$

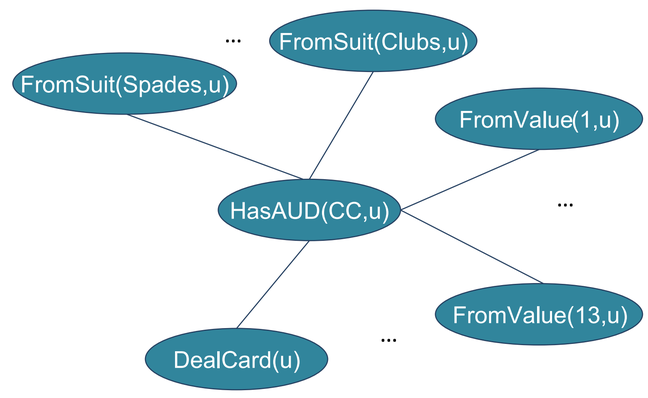
Experimental setup

Frame	Slot	Value	Prob.
MoveCard	FromSuit	s	0.000
		d	0.020
		h	0.014
		c	0.942
	FromValue	1	0.005
		2	0.850
		...	
		13	0.004
	TargetSuit	s	0.004
		d	0.005
		h	0.005
		c	0.990
	TargetValue	1	0.871
		2	0.006
		...	
		13	0.003
DealCard	{}	{}	0.006

Utterance
Klaveren twee op klaveren aas

AUD transcription
u ∈ {CC BH AD AAI G GH FI AC DA E BF CE AB AJI E CC BH AD F AF IJ AC CD AAH H A}

Mapping
MoveCard(clubs,2,clubs,1)



- **Setup 1:**
  - slot value with the highest probability is inferred for every slot only if it is higher than a predefined threshold
  - not all the slots need to be inferred for a semantic frame
- **Setup 2:**
  - null slot value is added to each slot - assigned when there exists no mapping to a particular slot
  - all slots are inferred for each semantic frame (no treshold)
  - slots mapped to null slot value are dropped
- **Setup 3: Hierarchical**
  - 1st step - mappings of AUD sequences to slots
  - 2nd step - mappings of AUD sequences to slot values of the particular slot (separate MLN for each slot)

Experimental results

- **PATCOR dataset:** Card game, 8 speakers, ≈ 3.3 h speech, >2000 commands

Speaker	1	2	3	4	5	6	7	8	Average
# Utterances	274	169	260	278	221	247	223	240	249
Baseline: NMF	66.1	69.3	76.2	55.9	90.9	54.7	77	48.5	67.3
Baseline: MNB	61.8	81.6	74.7	62	86.5	56.7	72.7	48.3	68
Baseline: SVM	59.7	80	77.8	57.5	89.4	49.9	65.9	45.1	65.7
MLN: Setup 1	66.3	82.9	80.6	65.1	91.4	54.8	79.1	56.5	72.1
MLN: Setup 2	68.4	85.6	83.6	67.4	94.5	65.1	81.4	56.1	75.3
MLN: Setup 3	68.4	83	83.2	67.6	93.5	66.1	82.2	56.2	75
MLN: Transcriptions	77.6	91.6	94.6	83.2	97.6	78	96.5	66.6	85.7

- **Domotica 3 dataset:** Home automation task, 9 speakers (7 dysarthric), ≈ 4 h speech, 26 distinct commands

Speaker	17	28	29	30	31	34	35	41	44	Average
# Utterances	347	204	174	198	225	331	268	144	164	228
Baseline NMF	96.3	82	94.5	87.6	74.3	90.2	94.6	86.8	93.3	88.8
Baseline MNB	94.1	79.3	88.8	85.2	73.8	91.4	95.3	86.2	96	87.8
Baseline SVM	97.6	76.6	93.5	86.1	73.3	93	97.4	85.2	95.8	88.7
MLN: Setup 1	98.5	90.6	97	92.1	83.1	96.2	98.8	91.6	98.8	94.1
MLN: Setup 3	98.2	90	96.6	90	81.5	96.7	98.8	91.8	99	93.6
MLN: Transcriptions	100	100	100	100	100	100	100	100	100	100

Conclusions

- MLNs perform well in the presence of noisy input data
- Especially applicable to impaired speech where standard ASR is not appropriate

DFG Ha3455/9-1, SPP1527. V. Despotovic: EUROWEB scholarship