LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

M. Stevens, F. Schiel

Machine Learning of Phonological Probabilistic Pronunciation Rules

with contributions of
F. Contugno and U. Reichel

Motivation and Basic Idea

Speech Production

Linguistic Modelling

- Abstract linguistic concept
- Observations
- Model: Algorithmic Process (e.g. rule system) to predict surface form from given concept
- e.g. Phonolgic Rules, Grammar

Speech Technology

- Speech = *statistical source*
- Observations
- Probabilistic model: conditional probabilities p(s|L) with s = surface form, L = class, factor
- e.g. Hidden Markov Model, *n-gram* Models

Motivation and Basic Idea

Linguistic Modelling	Speech Technology
„theory - driven"	„data - driven"
+ explanatory model based on physiological, neorological etc. facts	- no explanatory model
- rigid model	+ adaptable models
- bad probabilistic prediction	+ good probabilistic prediction
- often predicts un-realistic surface forms	- often 'over-adapted' to observed data

Motivation and Basic Idea

Example: Pronunciation of Contempory Italian

Linguistic Modelling

Phonology:
predict phonetic realisation from phonologic encoding using re-write rules

Speech Technology

Machine learning:
probabilistic model to predict probability of observed phonetic forms

Combined Model
Phonological Probabilistic Pronunciation

Outline

- Data: the CLIPS corpus
- Machine Learning of Probabilistic Rules
- Discussion of resulting rule set
- Technical Applications

CLIPS Corpus

Corpora e Lessici dell'Italiano Parlato e Scritto

- Synchronic survey of contemporary Italian Speech
- Dialectal coverage: *Turin, Genoa, Milan, Bergamo, Parma, Venice, Florence, Rome, Perugia, Naples, Bari, Catanzaro, Lecce, Palermo, Cagliari*
- Broad variety of speech types
- Here: *spontaneous speech from Map Task*
- 30 speakers, 30 dialogues (2 of each recording site)
- phonologic transcription (mainly automatic)
- phonetic labeling and segmentation (manual, 1 pass)
- 3229 turns, 32255 words, 87057 phonetic segments

Savy R, Cutugno F (2009)

Corpora e Lessici dell'Italiano Parlato e Scritto

Play some examples

Savy R, Cutugno F (2009)

1) Extraction
For each dialogue turn we extract:

- from STD labeling: phonologic form (encoded in SAMPA)

```
# s i # s o p r a # a u n # d a d o # p a ss a #
```

- from PHN segmentation: phonetic transcript

```
# s i # s o b r # a n # d a d O # b a ss a #
```

('#' denotes a word boundary)

Savy R, Cutugno F (2009), Wells J C (1997)

2) Alignment
Each pair of phonologic and phonetic form must be aligned to each other using the `PermA` tool:
Levensthein distance is minimized using probabilistic edit costs derived from the data set

```
phonol # s i # s o p r a # a u n # d a d o # p a ss a #
phonet # s i # s o b r _ # a _ n # d a d O # b a ss a #
```

('_' denotes the empty element)

Reichel U (2012)

3) Segmentation

Segment the aligned symbol strings into 'matched' and 'non-matched'

phonol	#	s	i	#	s	o	p	r	a	#	a	u	n	#	d	a	d	o	#	p	a	ss	a	#
phonet	#	s	i	#	s	o	b	r	_	#	a	_	n	#	d	a	d	O	#	b	a	ss	a	#

Schiel F (1999)

4) 'Probabilistic Micro Rules' (PMR)

From each 'non-matched' segment derive a PMR:

$$a,x,b \rightarrow y$$

a : phonol. left context of fixed length `cl > 0`

b : phonol. right context of fixed length `cl > 0`

x : phonologic, arbitrary length > 0

y : phonetic, arbitrary length > 0

phonol	# s i # s o	p	r	a	# a	u	n # d a d	o	#	p	a ss a #
phonet	# s i # s o	b	r	_	# a	_	n # d a d	O	#	b	a ss a #

e.g. cl=2 : so,p,ra → b

Schiel F (1999)

5) Collect, sort and count PMRs: $N(a,x,b \rightarrow y)$

6) Count phonol. environments: $N(a,x,b)$

7) Conditional probability for each PMR:

$$P(y|a,x,b) = \frac{N(a,x,b \rightarrow y)}{N(a,x,b)}$$

8) Pruning threshold: $N(a,x,b \rightarrow y) > T$

Schiel F (1999)

Applied to CLIPS MT data with cl = 1 and T = 4 results in 588 PMRs:

```
a,n,g>a,N,g          0.74961
ja,n,k>ja,N,k        0.73531
SS,E,n>SS,e,n        0.70121
#,S,i>#,SS,i         0.70121
a,dZ,i>a,ddZ,i       0.62998
k,o,d>k,O,d          0.59821
#,o,m>#,m            0.48034
we,s,t,i>we,ss,i     0.46082
o,z,E,g>o,s,e,g      0.46082
u,n,g>u,N,g          0.43987
...
```

Schiel F (1999)

References

Albano Leoni F, Maturi P (1994): Didattica della fonetica e parlato spontaneo. In: Ramat A G, Vedovelli M (Eds.) 'Italiano: lingua seconda/lingua straniera', Pubblicazioni della Societa di Linguistica Italiana Vol 34. Rome, Bulzoni: 153-164.
Anderson A H, Bader M, Bard E G, Boyle E, Doherty G, Garrod S, Isard St, Kowtko J, McAllister J, Miller J, Sotillo C, Thompson H, Weinert R (1991): The HCRC Map Task Corpus. Language and Speech, 34(4): 351-366.
Bertinetto P M, Loporcaro M (2005): The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. Journal of the International Phonetic Association 35(2): 131-151.
Carletta J, Isard A, Isard St, Kowtko J, Doherty-Sneddon G, Anderson A (1996): HCRC Dialogue Structure Coding Manual. Human Communication Research Centre, University of Edinburgh, Technical Report HCRC/TR-82.
Dalcher C V (2008): Consonant weakening in Florentine Italian: a cross-disciplinary approach to gradient and variable sound change. Language Variation and Change, Vol 20: 275-316.
Giannelli L, Cravens T D (1997): Consonantal weakening. In: Maiden M, Parry M (Eds.): The dialects of Italy. London/New York: Routledge. pp 32-40.
Loporcaro M (1997): L'origine del raddoppiamento fonosintattico:saggio di fonologia diachronica romanza. Basel/Tuebingen: Francke Verlag.
Stevens M (2012): A phonetic investigation into "Raddoppiamento Sintattico" in Sienese Italian Speech. Bern: Peter Lang.
Reichel U (2012): PermA and Balloon: Tools for string alignment and text processing, Proc. Interspeech. Portland, Oregon, paper no. 346.
Savy R, Cutugno F (2009): CLIPS Diatopic, diamesic and diaphasic variations in spoken Italian. 5th Corpus Linguistic Conference, Liverpool.
Schiel F (1999): Automatic Phonetic Transcription of Non-Prompted Speech, Proc. of the ICPhS 1999. San Francisco, August 1999. pp 607-610.
Schiel F, Draxler Chr, Harrington J (2011): Phonemic Segmentation and Labeling using the MAUS Technique. Workshop 'New Tools and Methods for Very-Large-Scale Phonetics Research', University of Pennsylvania, January 28-31, 2011.
Tajchman G, Jurafsky D, Fosler E (1995): Learning Phonological Rule Probabilities from Speech Corpora with Exploratory Computational Phonology. In Proceedings of ACL 95, Cambridge, MA, pp 9-15.
Wells J C (1997): SAMPA computer readable phonetic alphabet. In: Gibbon D, Moore R, Winski R (eds.): Handbook of Standards and Resources for Spoken Language Systems. Berlin and New York: Mouton de Gruyter. Part IV, section B.