NLP

Introduction to NLP

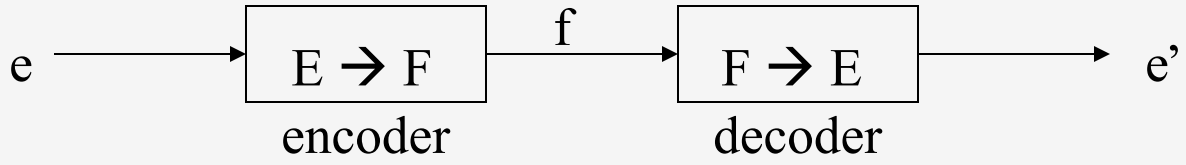
Noisy channel models

The Noisy Channel Model

- Example:
  - Input: Written English (X)
  - Encoder: garbles the input (X->Y)
  - Output: Spoken English (Y)
- More examples:
  - Grammatical English to English with mistakes
  - English to bitmaps (characters)
- $P(X,Y) = P(X)P(Y|X)$

Encoding and Decoding

- Given f, guess e



$$e' = \text{argmax}_e P(e|f) = \text{argmax}_e P(f|e)\, P(e)$$

translation model

language model

Example

- Translate “la maison blanche”

	P(f|e)	P(e)
cat plays piano	-	-
house white the	+	-
the house white	+	-
the red house	-	+
the small cat	-	+
the white house	+	+

Uses of the Noisy Channel Model

- Handwriting recognition
- Text generation
- Text summarization
- Machine translation
- Spelling correction
  - See separate lecture on text similarity and edit distance

Spelling Correction

w	c	$w \mid c$	$P(w \mid c)$	$P(c)$	$10^9 P(w \mid c) P(c)$
thew	the	ew | e	.000007	.02	144.
thew	thew		.95	.00000009	90.
thew	thaw	e | a	.001	.0000007	0.7
thew	threw	h | hr	.000008	.000004	0.03
thew	thwe	ew | we	.000003	.00000004	0.0001

From Peter Norvig: http://**norvig**.com/ngrams/ch14.pdf

NLP