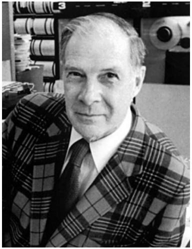
Applications of Statistical Language Modeling

Jon Dehdari

November 9, 2015

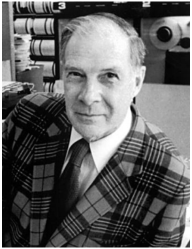
Good Morning!



Richard Hamming

“The purpose of computing is insight, not numbers”

Good Morning!

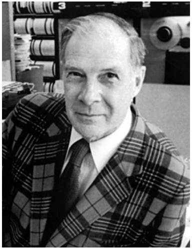


Richard Hamming

“The purpose of computing is insight, not numbers”

“If you expect to continue learning all your life, you will be teaching yourself much of the time. You must learn to learn, especially the difficult topic of mathematics.”

Good Morning!



Richard Hamming

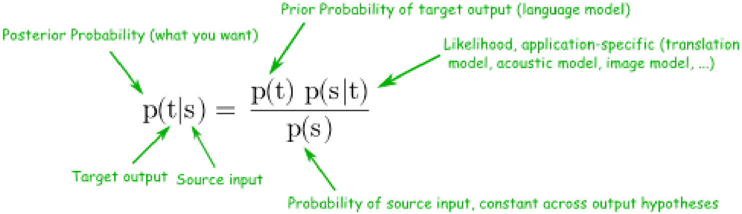
"The purpose of computing is insight, not numbers"

"If you expect to continue learning all your life, you will be teaching yourself much of the time. You must learn to learn, especially the difficult topic of mathematics."

"Any unwillingness to learn mathematics today can greatly restrict your possibilities tomorrow."

Turn That Noise Down!

Bayes' Theorem:



Turn That Noise Down!

Bayes' Theorem:

Prior Probability of target output (language model)

Posterior Probability (what you want)

Likelihood, application-specific (translation model, acoustic model, image model, ...)

$$p(t|s) = \frac{p(t)\ p(s|t)}{p(s)}$$

Target output

Source input

Probability of source input, constant across output hypotheses

Noisy Channel Model (applied to translation):

The Best Translation (probably)

Prior Probability, from Language Model

$$\hat{t} = \arg\max_{t}\ p(t)\ p(s|t)$$

Likelihood, from Translation Model

The Translation having Highest Score

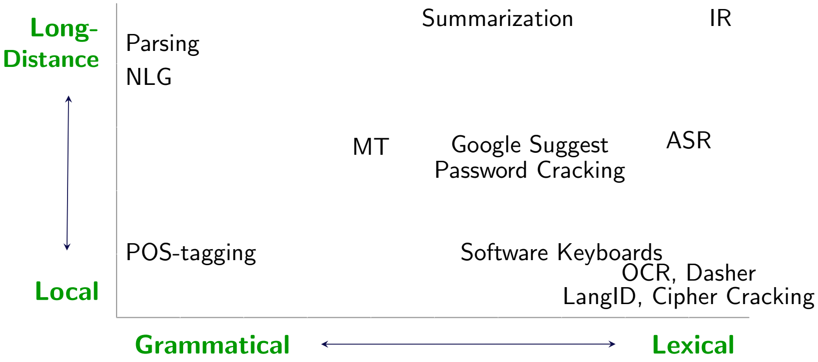
A Few Uses for Language Models

Statistical language models ensure fluency in speech recognition (like Siri), machine translation (like Google Translate), on-screen keyboards (smartphones), etc.

Actually, There's a Lot of Uses!

- Google suggest
- Machine translation
- Assisting people with motor disabilities. For example, Dasher
- Speech Recognition (ASR)
- Optical character recognition (OCR) and handwriting recognition
- Information retrieval / search engines
- Data compression
- Language identification, as well as genre, dialect, and idiolect identification (authorship identification)
- Software keyboards
- Surface realization in natural language generation
- Image caption generation
- Email response generation
- Password cracking
- Cipher cracking

Differences in LM Uses



Actually parsers and POS-taggers aren't applications of language models, but rather they are types of language models.

LM Usage

Typical LM Queries in ...

ASR	: p(recognize speech) vs. p(wreck a nice beach) vs. p(wreck an ice peach), ...
Cipher cracking	: p(attack at dawn) vs. p(uebvmkdvkdbsqk)
Google Suggest	: p(how to cook french fries) vs. p(how to cook french dictionary)
MT & NLG	: lex: p(use the force) vs. p(use the power); ordering: p(ready are you) vs. p(are you ready)
OCR	: p(today is your day) vs. p(+qdav ls y0ur d4ij)
IR	: query(cats and the cradle): doc1(i like cats) vs. doc2(i like dogs)
LangID	: query(a blue watch): lang1(the green witch . . .) vs. lang2(la bruja verde . . .)

A good cipher should obey the principle of diffusion (Shannon, 1949).

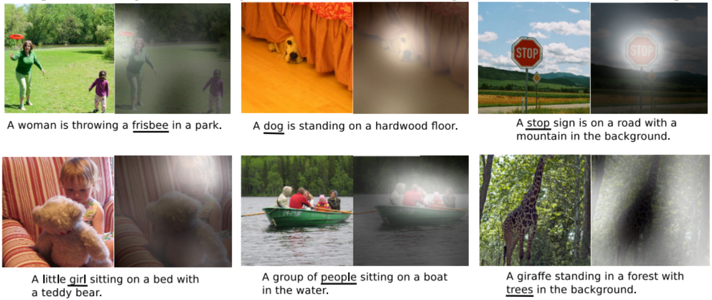
Language Modeling is Interesting!

NLP Task	Avg. Entropy
Language Modeling (=Word Prediction)	7.12
English-Chinese Translation	5.17
English-French Translation	3.92
QA (Open Domain)	3.87
Syntactic Parsing	1.18
QA (Multi-class Classification)	1.08
Text Classification (20 News)	0.70
Sentiment Analysis	0.58
Part-of-Speech Tagging	0.42
Named Entity Recognition	0.31

From Li & Hovy (2015)

Illustration with Image Caption Generation

Figure 4. Examples of attending to the correct object (*white* indicates the attended regions, *underlines* indicated the corresponding word)



From Xu et al (2015; ICML, Fig. 4). This uses the neural attention model, which we'll discuss later in the semester.