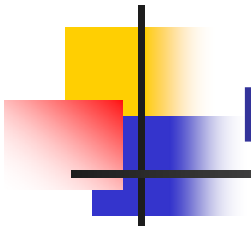
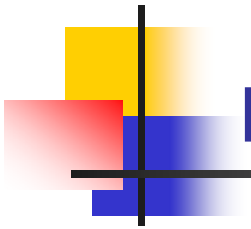


Machine Learning for Adaptive Spoken Control

Bryan McEleney and Gregory O'Hare

Speaking Autonomous Intelligent Devices Group (SAID)

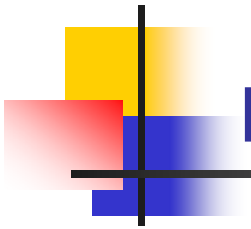
Department of Computer Science,
University College Dublin
Republic of Ireland.



Machine Learning for Adaptive Spoken Control

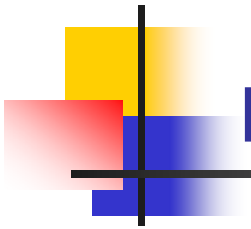
- Spoken control is becoming necessary for mobile computing
- What we have today is inflexible
  - VoiceXML?
  - Interface Development is straightforward, but:
    - Too much bias toward System-Initiative
    - Fixed, user-learned commands

Machine Learning for Adaptive Spoken Control



- What about User-Inititative?
  - 1: Change a contact in the list.
  - 2: Edit a contact.
  - 3: Change a name in the list.
  - 4: Change a contact.
  - 5: Change the details of a contact.

- ASR is still poor - 5% per word, or 18% per 4-word sentence

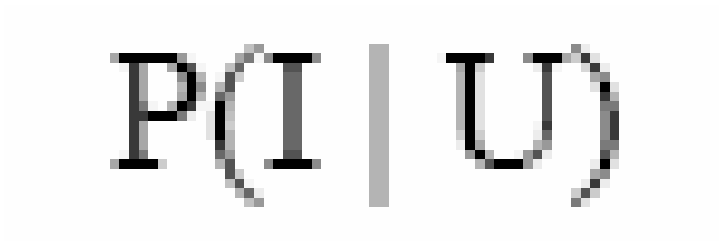


Machine Learning for Adaptive Spoken Control

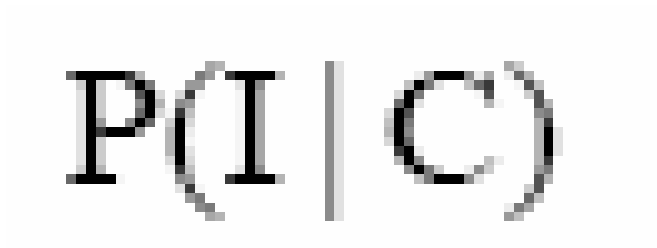
- How about Machine Learning?
- Understanding as Classification
- 20 dialogue act classes
- Named Entity Extraction:
  - “Send an email to Bryan” ->
    “Send an email to <name>” + “Bryan” ->
    send_email + entity(Bryan)

Machine Learning for Adaptive Spoken Control

Utterance Evidence



Context Evidence

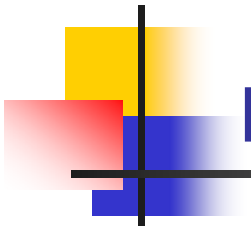


Machine Learning for Adaptive Spoken Control

- Using Bayes rule + conditional independence

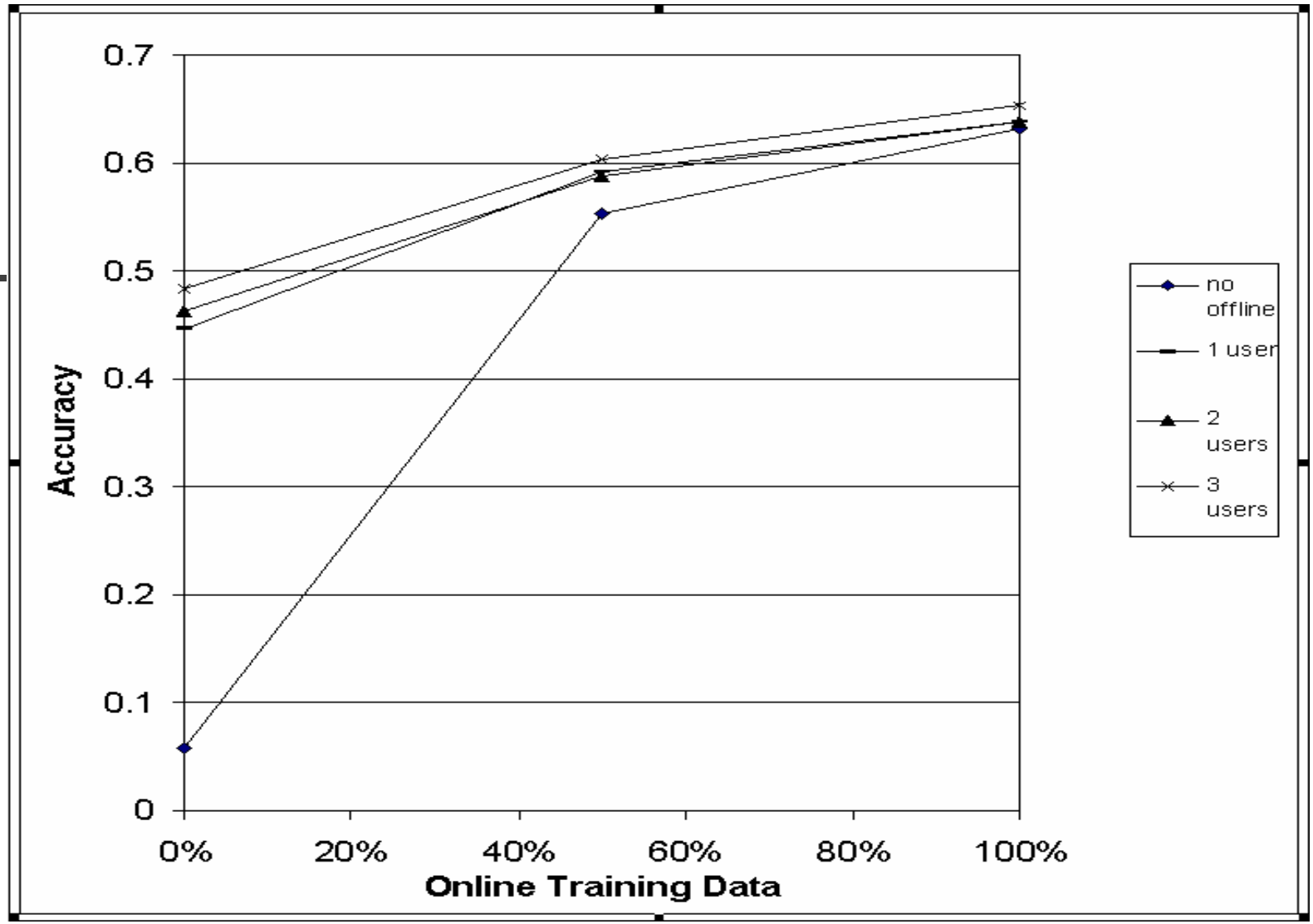
$$\hat{I} = \underset{I}{\operatorname{argmax}} \frac{P(I|U).P(I|C)}{P(I)}$$

- This is the same as is used in speech recognition



Machine Learning for Adaptive Spoken Control

- Utterance Classifier
- A type of “keyword spotting” using Nearest Neighbour Classification
- Computes Edit Distance against a training set

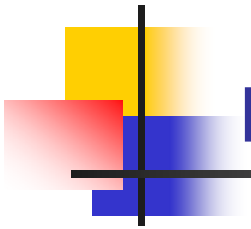


- 20 classes, perplexity 15.1
- 95% confidence interval: [0.601, 0.705]

Machine Learning for Adaptive Spoken Control

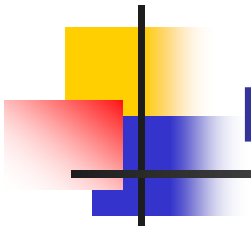
- Context Interpreter
- Dialogue Act N-grams

Act N-2 (user turn)	Act N-1 (system turn)	Act N (user turn)
a	c	P(a) = 0.5 P(b) = 0.5
a	d	P(a) = 0.1 P(b) = 0.9
b	c	P(a) = 0.6 P(b) = 0.4
b	d	P(a) = 0.3 P(b) = 0.7



Machine Learning for Adaptive Spoken Control

- Development Lifecycle
  - Domain Independent
  - Simple, efficient algorithms
  - Obtaining Training Data
    - Offline - supervised
    - Online - unsupervised
      - Perplexity Heuristic
      - Failed task heuristic
    - Pooling



Machine Learning for Adaptive Spoken Control

- Still to do
  - Evaluate against traditional approaches.
    - 82% accuracy bound for parsing approach + hand coding problem
    - Static keyword spotting needs extensive hand coding to achieve acceptable coverage
    - Backoff methods might be useful
  - Evaluate context interpreter