Intent Classification of Voice Queries on Mobile Devices
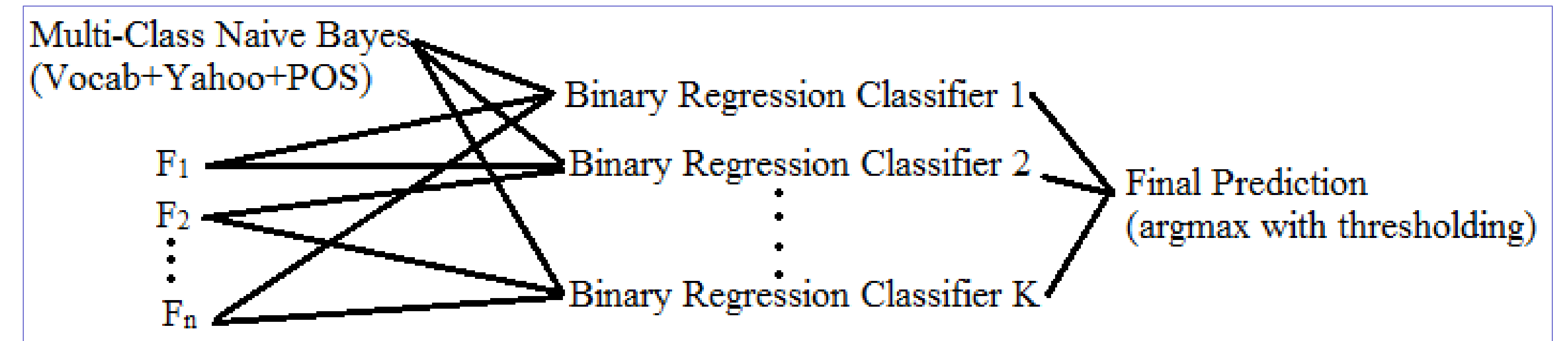
Subhabrata Mukherjee[†], Ashish Verma[†], Kenneth W. Church[‡]

[†]IBM India Research Lab, [‡] IBM T.J. Watson Research Center

Introduction

- Mobile query classification is made difficult by short, noisy queries and the presence of inter-active and personalized queries like map *(where is the closest restaurant)*, command-and-control *(open facebook)*, dialogue *(how are you?)*, joke *(will you marry me?) etc.*
- Voice queries are more difficult than typed queries due to the errors introduced by the automatic speech recognizer
- In this work we bring the complexities of voice search and intent classification together in proposing a multi-stage classifier for intent classification

Intent Classification

Example Queries	Find nearest Subway	Map
	Call mom	Command-and-Control
	Best buy com	Navigational
	What is pi	Knowledge
	How to spell	Endpoint

Multi-Class Naive Bayes (Vocab+Yahoo+POS) → F_1, F_2 ... F_n → Binary Regression Classifier 1, Binary Regression Classifier 2 ... Binary Regression Classifier K → Final Prediction (argmax with thresholding)

- Stage 1 – Multi-class *Naïve Bayes* classifier uses *bag-of-words, part-of-speech tag* information and *domain words* of the query as features and predicts top K probable classes for a query.
  - Domain words (DW) of a query are extracted from url's of top ranked retrieved pages from search engine. For example, the search engine retrieves the DW's *imdb marvel en.wikipedia youtube etc.* for the query *the avengers.*
  - DW introduce external knowledge in classifier (*Movies, Music, Sports etc.*) and keeps online data requirement low
  - POS tags detect patterns for *Command-and-Control, Map, Knowledge, Dialogue,* broken queries *(Endpoint) etc.*
- Stage 2 – Top K ranked classes are taken with some additional *features* in a logistic regression classifier.
  - Features like DW and url ranking, DW and query overlap (*Navigational*), substrings having maximum information gain for a class help differentiate between close query categories from Stage 1
- Stage 3 – Prediction of the independent binary classifiers are combined into a single prediction

- Navigational – Query DW overlap
- Map – Presence of substrings *find, close, direction, near, where* in query
- Movie – Presence of *IMDB* as the topmost DW
- Command-and-Control – Query starting with a *Verb*
- Websearch – Presence of *Wikipedia* in the topmost two DW's

Evaluation

- 52,282 unique queries, with total 1,04,950 impressions, are collected from an android voice search application and manually tagged. Avg. number of words per query is 2.3. Avg. word error rate of ASR engine is 20%. The figure below shows the F_1 score comparison of different models over 13 most frequent query classes.

Legend: ☐ Reg+Vocab+POS+Yahoo ■ Vocab+POS+Yahoo ☐ Vocab+Yahoo ■ Vocab+POS ■ Vocab

Categories: weather, map, nav, command, knowledge, movies, dialogue, restaurant, web, music, sports, acceptor, endpoint