The Pareto Frontier of Utility Models as a Framework for Evaluating Push Notification Systems
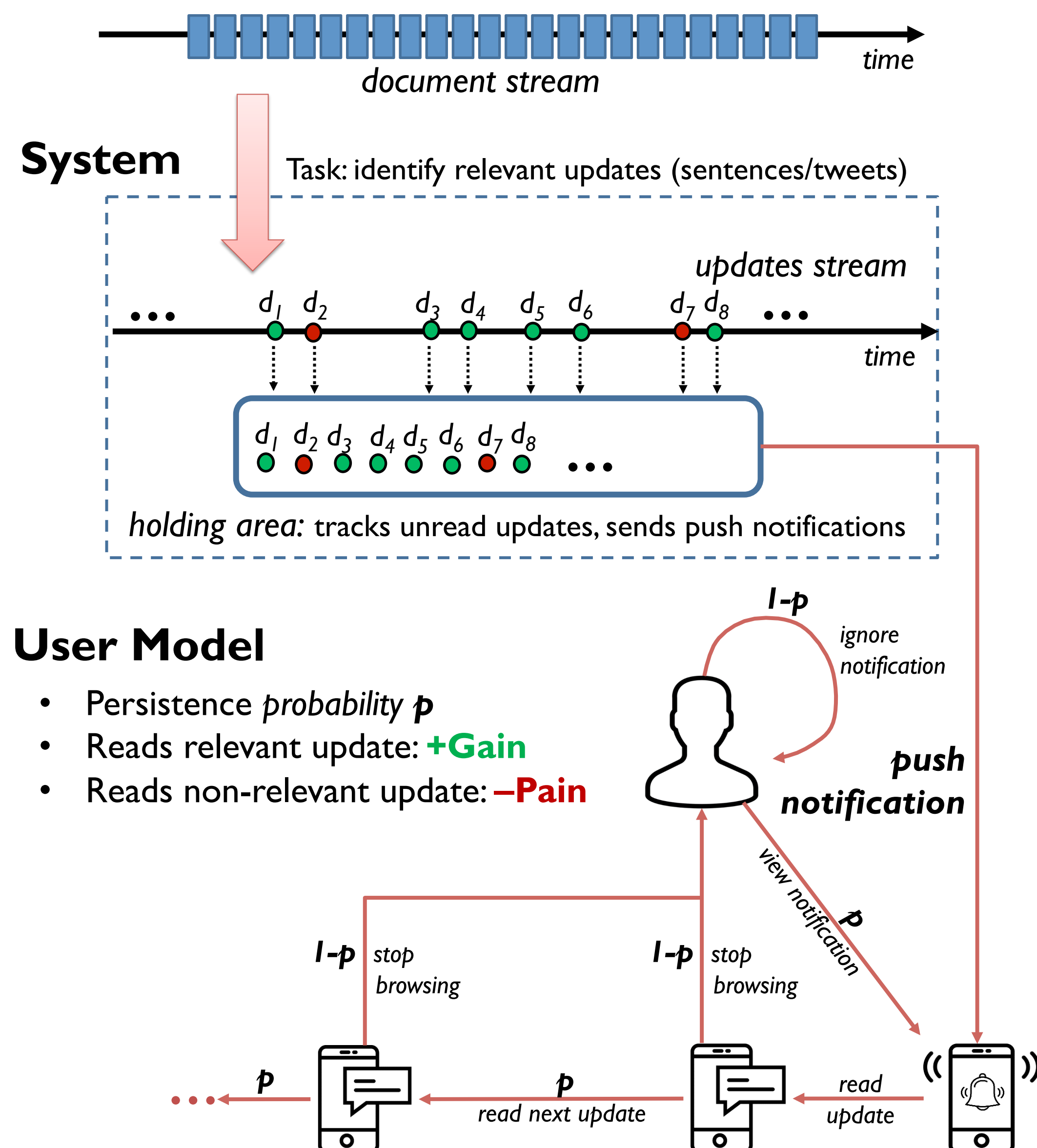
Gaurav Baruah and Jimmy Lin
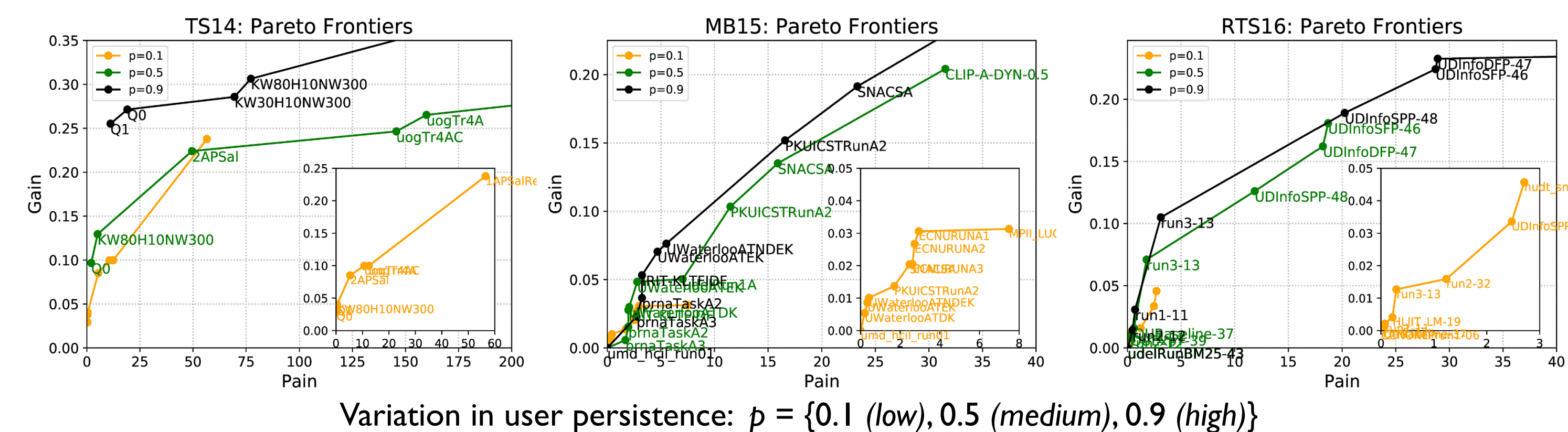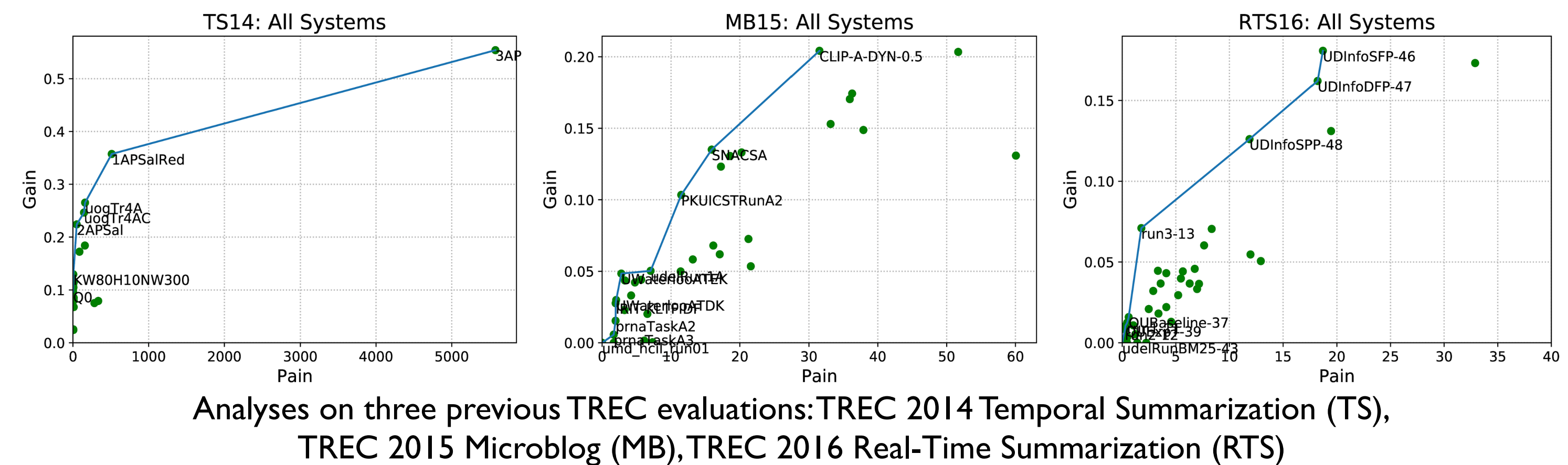
University of Waterloo

tl; dr – We present a unified, utility-based framework for evaluating push notification systems. Our key insight is to examine the Pareto frontier of a gain vs. pain tradeoff space to understand system effectiveness.

Prospective Information Needs

document stream *time*

System Task: identify relevant updates (sentences/tweets)

updates stream

d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 *time*

d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 ...

holding area: tracks unread updates, sends push notifications

User Model

- Persistence *probability* p
- Reads relevant update: **+Gain**
- Reads non-relevant update: **–Pain**

$1-p$ ignore notification

p *push notification*

view notification

$1-p$ *stop browsing* $1-p$ *stop browsing*

p read notification

p read update

p read next update

Key Idea: Evaluation Using Pareto Frontier of Gain vs. Pain

Systems on the frontier dominate effectiveness (minimum level of pain for given level of gain)

Analyses on three previous TREC evaluations: TREC 2014 Temporal Summarization (TS), TREC 2015 Microblog (MB), TREC 2016 Real-Time Summarization (RTS)

Variation in user persistence: $p = \{0.1$ *(low)*, 0.5 *(medium)*, 0.9 *(high)*$\}$

Unified Evaluation Framework
- Applicable to multiple tasks (TS, MB, RTS)
- Generalizable to newswire, tweets, etc.

Subsumes Previous Metrics
- Pain and gain are directly visible
- Precision reflected in vector angle

More Insightful Analyses
- Easy to separate volume from precision
- Explicit visualization of different tradeoff points

Extensibility
- Can be calibrated using real user data
- Supports more complex user and utility models