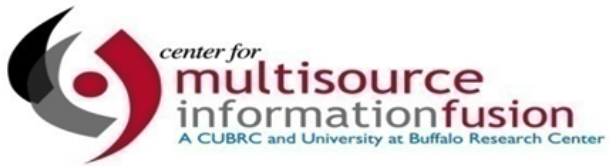
Evaluating Spreading Activation for Soft Information Fusion

Michael Kandefer and Stuart C. Shapiro

Department of Computer Science and Engineering,
and Center for Multisource Information Fusion
University at Buffalo, Buffalo, NY 14260
{mwk3,shapiro}@buffalo.edu



Outline

Topic

Report study of two algorithms

that use spreading activation

to retrieve background and contextual information

relevant to an English message concerning insurgency.

Find best settings of two parameters for each algorithm.

Compare the two algorithms.

Motivation

- Analyze natural language message.
  - Counterinsurgency domain
- Enhance with **relevant** background and contextual information
  - to enhance information value,
  - to facilitate fusion with other information.

Tractor

1. Syntactic Processing.
2. Propositionalizer.
   - Produce propositional graph.
3. Contextual Enhancement.
   - Add relevant ontological and other background information found via Context-Based Information Retrieval (CBIR).
4. Fuse contextually enhanced propositional graphs.

Example of Enhancement

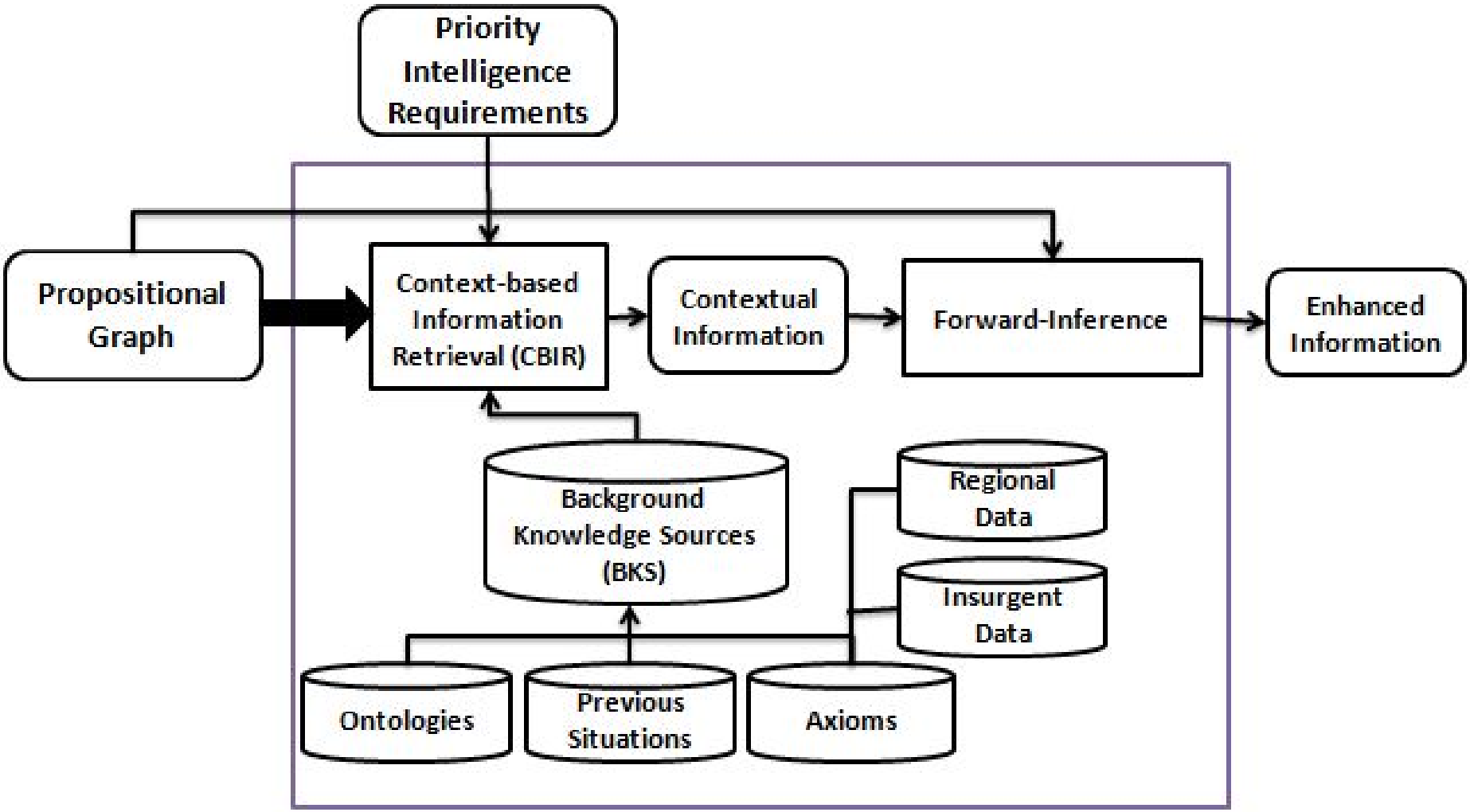
Message 2:

> Source said a Sunni youth he knows to be about 20 years old, Khalid Sattar, has become increasingly vocal in denouncing the U.S. at several mosques in Adhamiya.

Retrieved relevant background information:

- Khalid Sattar is a Person.
- Denouncing the U.S. is an indicator of terrorist activity.
- If someone is engaged in a type of terrorist activity, they may be an insurgent.

CBIR Process



CBIR by Spreading Activation

1. Represent the message as a propositional graph.
2. Represent the Background Knowledge (BK) as a propositional graph.
3. Embed the message graph in the BK graph.
4. Apply "pulse" to nodes from message (cue nodes).
5. Spread the pulse to reachable nodes.
6. Collect activated nodes.

Propositional Graphs

- A Knowledge Representation (KR)
  - A kind of associative network.
  - Labeled directed acyclic graph
  - with formal syntax
  - and formal semantics

Propositional Graphs

- A Knowledge Representation (KR)
  - A kind of associative network.
  - Labeled directed acyclic graph
  - with formal syntax
  - and formal semantics
- Atomic Node
  - Corresponds to individual constant.
  - No outgoing arcs.
  - Denotes entity in domain.
  - Node ID = any symbol.

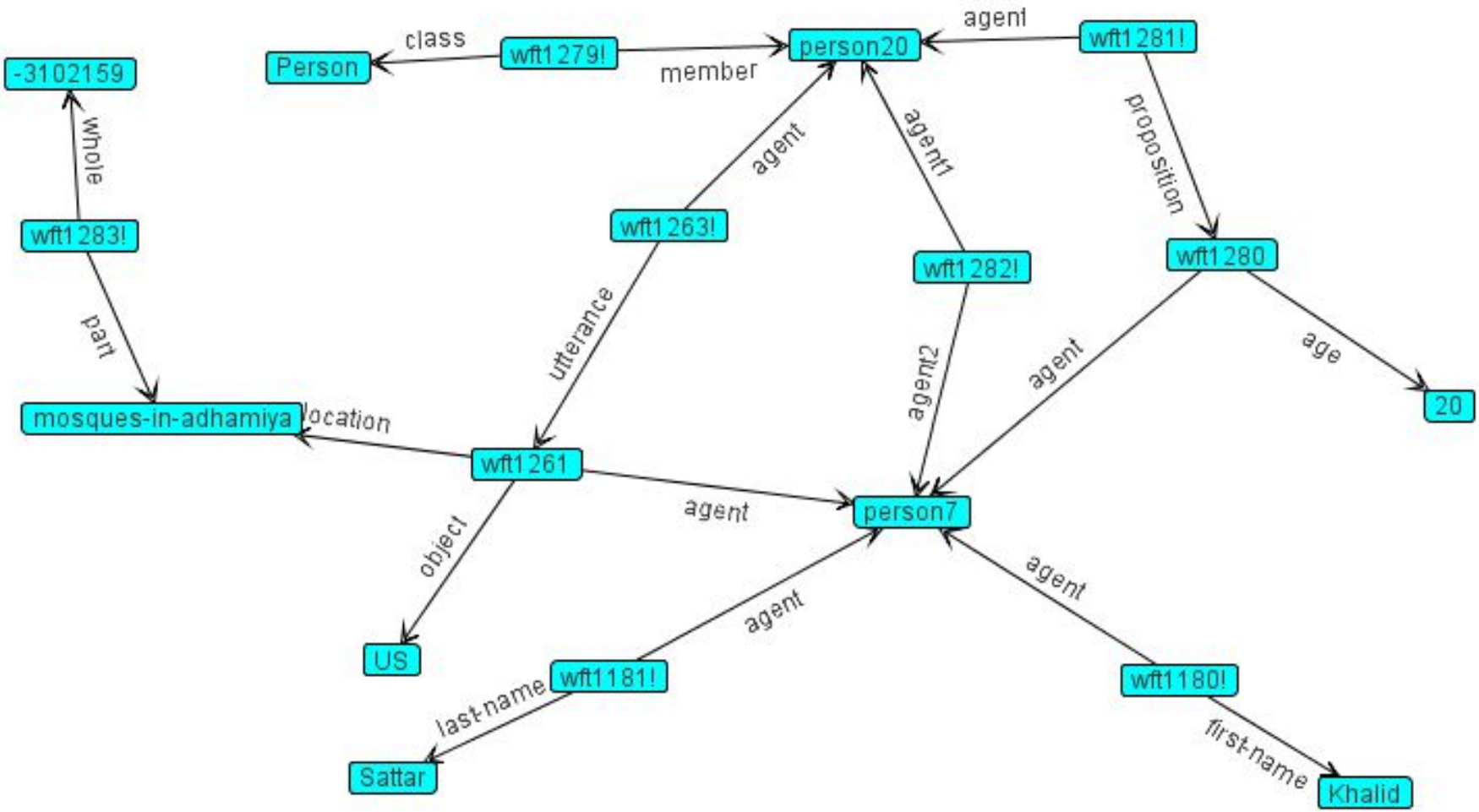
Propositional Graphs

- A Knowledge Representation (KR)
  - A kind of associative network.
  - Labeled directed acyclic graph
  - with formal syntax
  - and formal semantics
- Atomic Node
  - Corresponds to individual constant.
  - No outgoing arcs.
  - Denotes entity in domain.
  - Node ID = any symbol.
- Molecular Node
  - Corresponds to functional term.
  - Outgoing arcs labeled with argument position (role).
  - Compositional semantics.
  - Denotes
    - Entity in domain
    - Possibly a proposition.
  - Node ID = `wft`i[!]

Basic Principles

- Comprehensiveness
  - Every entity, person, category, property, value, etc.
    Every proposition, belief, fact, etc.
    is represented by a node.
- Uniqueness Principle
  - No two nodes with same ID.
  - No two molecular nodes with same labeled arcs to same nodes.
  - No two nodes representing the (obviously) same domain entity.
  - Base case for fusing propositional graphs
    - and for embedding message graph into BK graph.

A Propositional Graph for Message 2



General Spreading Activation Algorithm

for all $n \in CueNodes$ **do**
  $A_n \leftarrow InitialActivationLevel \geq threshold$
  Add n to $NewlyActivated$
end for
while $NewlyActivated \neq \phi$ **do**
  $Frontier \leftarrow NewlyActivated$; $NewlyActivated \leftarrow \phi$
  for all $n \in Frontier$ **do**
    for all m *adjacent to* n **do**
      Calculate A_m
      if $A_m \geq threshold$ **then**
        Put m in $NewlyActivated$
      end if
    end for
  end for
end while
return $\{n \mid A_n \geq threshold\}$

Parameters and Procedures

- $0.0 \leq A_m \leq 1.0$
- $\mathit{InitialActivationLevel} = 1.0$
- *threshold* determined by study
- Procedure to calculate A_m: two studied

Study Two Spreading Activation Algorithms

- Texai Algorithm
  - Standard spreading activation algorithm.
  - Designed for information retrieval.
- ACT-R Declarative Memory Activation Algorithm
  - Developed for ACT-R cognitive model.
  - Designed to rank memory chunks for relevance to new information, and choose the best.

Texai Algorithm

To spread the activation pulse:

$$A'_j = A_j + \sum_{i \in N} A_i * \frac{1}{|N|} * D$$

- A'_j: New activation level of node j.
- A_j: Previous activation level of node j.
  All non-cue nodes start at 0.
- N: Set of nodes adjacent to node j.
- $\frac{1}{|N|}$: to provide equal weighting of arcs connected to j.
- D: Decay factor.
  Pulse weakens as it spreads.
  Determined by study.

ACT-R Declarative Memory Activation Algorithm

To spread the activation pulse:

$$A_i = S - \frac{1}{|Frontier|} \sum_{j \in Frontier} ln(deg(j))$$

A_i: Activation level of node i.

S: Maximum associative strength.
Determined by study.

(This is an adaptation of the original ACT-R Declarative Memory Activation Algorithm. See the paper.)

Methodology of Information Retrieval Evaluations

- Retrieve a subset of available documents, BK
- that are relevant to a cue, or set of keywords
- Rel = set of ground truth relevant documents
- Ret = set of retrieved documents
- Recall: fraction of relevant documents that are retrieved
  $r = \frac{|Rel \cap Ret|}{|Rel|}$
- Precision: fraction of retrieved documents that are relevant
  $p = \frac{|Rel \cap Ret|}{|Ret|}$
- F-measure: Combination of recall and precision
  $F = \frac{2rp}{r+p}$

  $0.0 \leq F \leq 1.0$, the higher, the better.

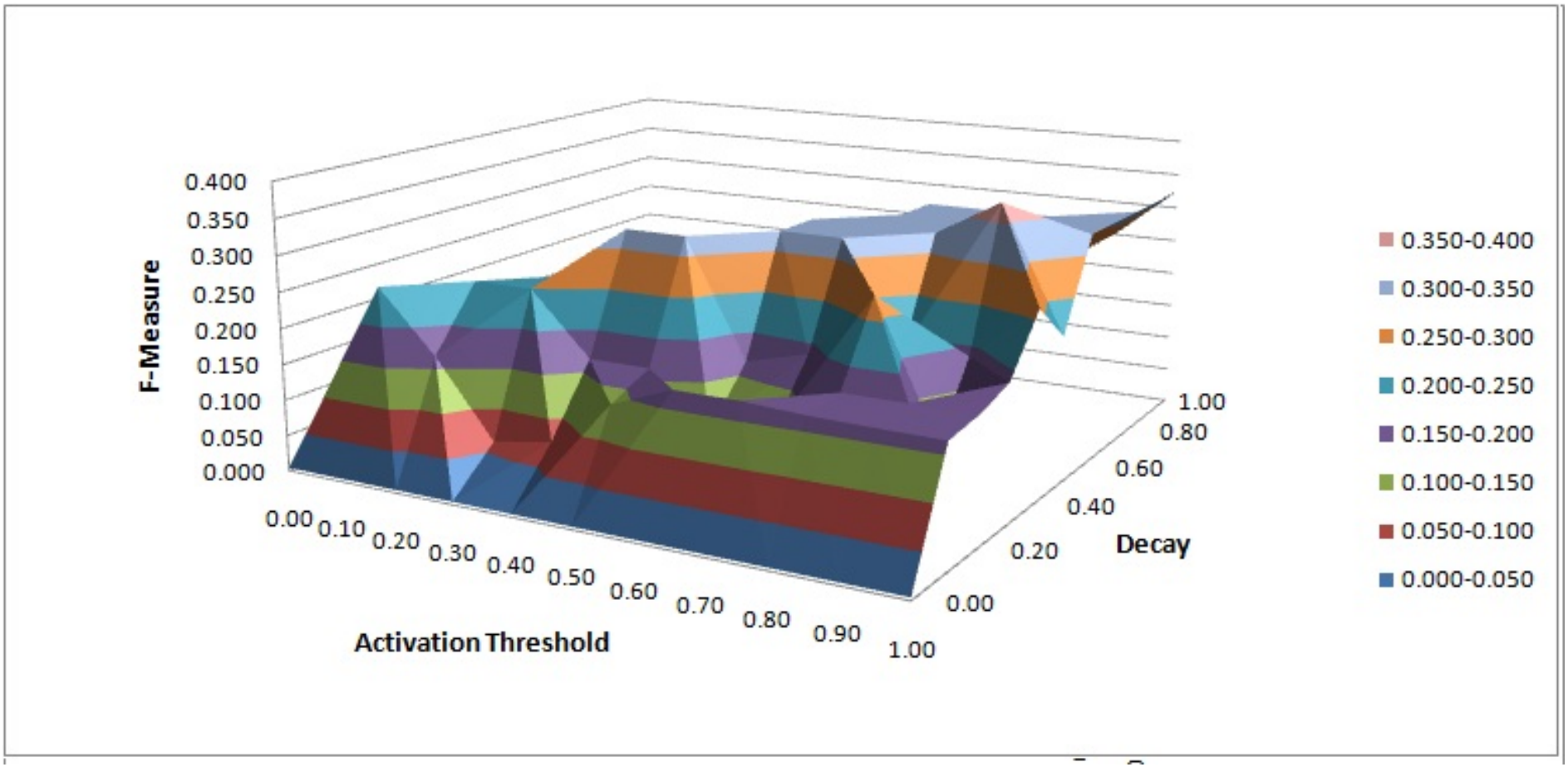
"Documents" Used in CBIR Evaluations

- BK = Propositional graph representations of
  - Iraqi subset of National Geospatial-Inelligence Agency GEOnet Names Server data.
  - Hand-crafted information about people in the domain
  - Seven hand-crafted rules for reasoning about the domain
- Cue = Propositional graph representation of a message.
- *Rel* = BK propositions actually used by full forward inference from Cue and BK.
- *Ret* = set of retrieved BK propositions.

Texai Parameter Evaluation

for all messages in 4 chosen messages **do**
  for threshold from 0.0 to 1.0 by 0.1 **do**
    for decay from 0.0 to 1.0 by 0.1 **do**
      Compute F-measure
    end for
  end for
end for

Average F-measures for Texai

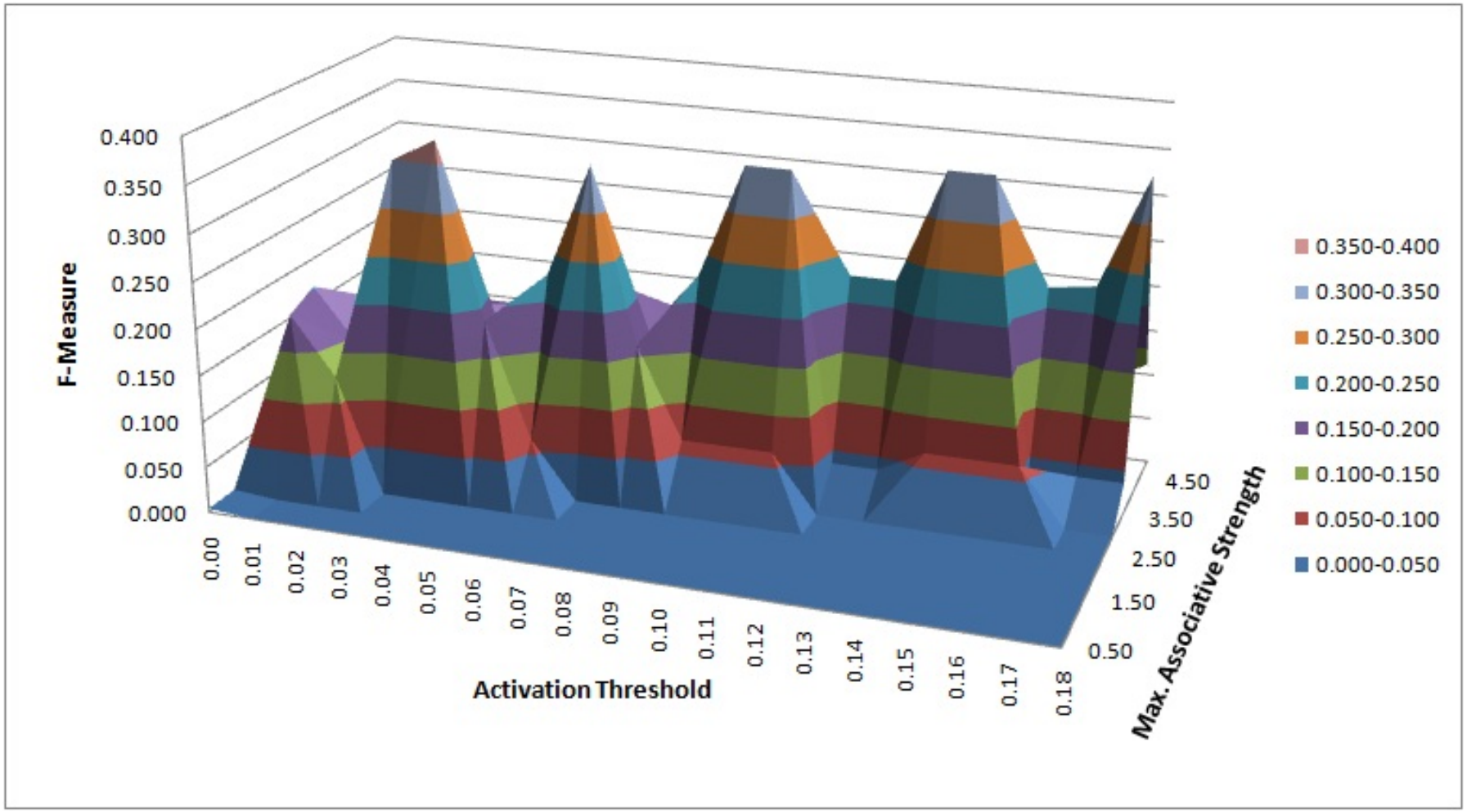


Max average F-measure of 0.375 at threshold = 0.5, decay = 0.9.

ACT-R Parameter Evaluation

for all messages in 4 chosen messages **do**
  for threshold from 0.0 to 0.19 by 0.01 **do**
    for max associative strength (S) from 0.5 to 5.0 by 0.5 **do**
      Compute F-measure
    end for
  end for
end for

Average F-measures for ACT-R



Max average F-measure of 0.375 at threshold = 0.04, S = 2.0.

Comparison of Texai vs. ACT-R

F-measure for each algorithm at best settings of parameters:

Message	Texai	ACT-R
Message 1	0.5	0.0
Message 2	1.0	0.75
Message 3	0.0	0.5
Message 4	0.0	0.25
Mean	0.375	0.375
Standard Deviation	0.41	0.28

ACT-R declarative memory activation judged better due to smaller standard deviation.

Conclusions

- Methodology successful
  - for finding best parameters
  - for comparing algorithms
- Both Texai and ACT-R declarative memory activation algorithms are good for CBIR.

Future Work

- Use larger datasets
- Evaluate ACT-R's base-level-learning method
  that rewards nodes for relevance to recent previous messages.

Acknowledgements

- Multidisciplinary University Research Initiative (MURI)
  - Grant (Number W911NF-09-1-0392) for "Unified Research on Network-based Hard/Soft Information Fusion"
  - Issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery.
- Center for Multisource Information Fusion
  - Dr. James Llinas, Executive Director

- Contact information

  - Michael Kandefer (email: mwk3@buffalo.edu)
  - Dr. Stuart C. Shapiro (email: shapiro@buffalo.edu)