422 big picture: Where are we With reading papers?

**Deterministic**
- Logics
  - First Order Logics
- Ontologies
  - Full Resolution
  - SAT

**Stochastic**
- Belief Nets
  - Approx. : Gibbs
- Markov Chains and HMMs
  - Forward, Viterbi...
  - Approx. : Particle Filtering
- Undirected Graphical Models
  - Markov Networks
  - Conditional Random Fields
- Markov Decision Processes and Partially Observable MDP
  - Value Iteration
  - Approx. Inference
- Reinforcement Learning

**Applications of AI**

**Representation**

**Reasoning Technique**
Extracting Knowledge from Evaluative Text

Giuseppe Carenini, Raymond T. Ng, Ed Zwart
Computer Science Dept.
University of British Columbia
Vancouver, CANADA
Motivation and Focus

• Large amounts of info expressed in **text form** is constantly produced
  – News, Reports, Reviews, Blogs, Emails,….

• Pressing need to extract and summarize key/strategic info

• Considerable work but limited **factual** info

**Our Focus:** evaluative text about single entity (**good** vs. **bad**, **right** vs. **wrong**)

• Customer reviews
• Travel logs
• Job candidate evaluations…… etc.
KCAP from evaluative text (single entity)

• Extract relevant knowledge

A. What features of the entity are evaluated in the reviews? [Hu, Liu AAAI ’04] [Popescu Etzioni HLT ’05]

B. For each feature:

1. what is the polarity of the evaluation? (good vs. bad) [Hu, Liu KDD ’04]
2. what is the strength of the evaluation? (rather good vs. extremely good) [Wilson et al. AAAI ’04]

• Summarize and present extracted knowledge to user ............
Outline

• Feature Extraction: limitations of previous work and our solution

• Evaluation of our approach

• Benefits in term of KCAP

• Conclusion and Demo of Future work 😊
...... the canon computer software used to download, sort, ... is very nice and very easy to use. The only two minor issues I have with the camera are the lens cap (it is not very snug and can come off too easily). ... .

The menus are easy to navigate and the buttons are easy to use. It is a fantastic camera and well worth the price.
Feature Extraction: sample form corpus

[Hu&Liu 2004]

...... the canon computer software used to download, sort, . . . is very nice and very easy to use. the only two minor issues i have with the camera are the lens cap (it is not very snug and can come off too easily). . . .

the menus are easy to navigate and the buttons are easy to use. it is a fantastic camera and well worth the price.
Limitations of previous approach

Key problems with extracted features (for KCAP):

• May be too many and redundant (often > 100)
• Flat, unstructured list (lack hierarchical organization)
• May be expressed in an unfamiliar terminology (for target user)

- spot metering
- metering option
- remote control
- battery
- night mode
- light automatic correction
- battery life
- remote
- battery charging system
- low light focus
- ...

Battery

Lighting
Example Ideal Mapping

UDFs

1. Canon G3 PS Digital Camera [canon, canion PS g3, digital camera, camera,...]
   1.1 User Interface [button, menus, lever]
   1.2 Convenience [
   ├── Battery [battery life, battery charging system, battery]
   │   ─ Self Timer [
   │       ├── Burst Mode [speed, mode]
   │       └── Rapid Fire Shot [speed]
   │       └── Delay Between Shots [unresponsiveness, delay, speed, lag time, lag]
   └── ....

2. Not Placed [manual, function, quality, strap, service, shoot, learning curve,...]
Our Solution

• Map extracted features (Crude Features (CF)) in a hierarchy of product features at different levels of abstraction. Two alternatives:
  – Learn the hierarchy
  – Have the user provide a hierarchy of User Defined Features (UDF)

• Such a mapping will:
  – Eliminate redundancy (CFs with same referent mapped in the same UDF)
  – Provide conceptual structure
  – Increase user familiarity with CFs
Mixed-initiative Process

Corpus of Evaluative Text

[Hu, Liu AAAI '04]

CFs

Mapping method

Merged Features CF → UDF

UDFs

User can revise UDF

User can revise mapping
Our Mapping Method

• Map each CF in the “most similar” UDFs
• CFs and UDFs are terms (i.e., sequences of 1 or more words)

So need a measure of term similarity

• Our approach to term similarity: combine similarity between constituent words

• So need a measure of word similarity and a function to combine similarities
Word Similarity Metrics \(wm\)

- **String Matching**: baseline

- **WordNet Synset Matching**: 1 if the two words appear in the same synset. e.g., \((\text{photograph, photo, exposure, picture})\)

- **Wordnet Distance Matching**: a set of measures that compute the semantic distance between the synsets of the two words

[Patwardan, Pedersen 2003] CPAN module
Term similarity: Combine word scores

\[ cf = \{v_1, \ldots, v_n\}; \quad udf = \{w_1, \ldots, w_m\} \]

- **MAX**: terms' score is the maximum score of comparing all possible word pairs

  \[ \max_{i,j} \{wm(v_i, w_j)\} \]

- **AVG**: terms' score is the average of the max of all \(i\) with \(j\), and vice versa (to avoid a high score of one word dominate the whole term's score)

  \[
  \frac{\sum_{i=1}^{n} \max_{j} \{wm(v_i, w_j)\}}{n} + \frac{\sum_{j=1}^{m} \max_{i} \{wm(v_i, w_j)\}}{m} / 2
  \]
Mapping Algorithm

Algorithm
• Each CF is mapped to the UDF with which it receives the greatest similarity score
• In case of tie scores CF is mapped more than once
• But mapping occurs only if score greater than a given threshold

Threshold
• For **string** and **synset** matching the threshold was set to 0.
• For **Wordnet distance** similarity measures was set by varying a parameter $\theta$
Outline

• Feature Extraction: limitations of previous work and our solution

• Evaluation of our approach

• Benefits in term of KCAP

• Conclusion and Demo of Future work 😊
Results DigCam for AVG

<table>
<thead>
<tr>
<th>Wordnet distance (lin)</th>
<th>p_dist</th>
<th>redun</th>
</tr>
</thead>
<tbody>
<tr>
<td>str_match</td>
<td>.38</td>
<td>.23</td>
</tr>
<tr>
<td>syn_score</td>
<td>.39</td>
<td>.27</td>
</tr>
<tr>
<td>θ</td>
<td>.35</td>
<td>.29</td>
</tr>
<tr>
<td>-0.2</td>
<td>.39</td>
<td>.33</td>
</tr>
<tr>
<td>-0.4</td>
<td>.43</td>
<td>.40</td>
</tr>
<tr>
<td>-0.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mapping quality measures

↓😊  ↑😊
## Results DVD for AVG

### Word Similarity metrics

<table>
<thead>
<tr>
<th>Wordnet distance (lin)</th>
<th>p_dist</th>
<th>redund</th>
</tr>
</thead>
<tbody>
<tr>
<td>str_match</td>
<td>.27</td>
<td>.21</td>
</tr>
<tr>
<td>syn_score</td>
<td>.28</td>
<td>.25</td>
</tr>
<tr>
<td>(\theta)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>.40</td>
<td>.38</td>
</tr>
<tr>
<td>-0.4</td>
<td>.47</td>
<td>.45</td>
</tr>
<tr>
<td>-0.6</td>
<td>.55</td>
<td>.52</td>
</tr>
</tbody>
</table>

**:smiley:** :wink: :smiley:**
Outline

• Feature Extraction: limitations of previous work and our solution

• Evaluation of our approach

• Benefits in term of KCAP

• Conclusion and Demo of Future work 😊
Benefits in term of KCAP

Key questions for manufactures and potential customers
- what product features are more frequently mentioned?
- how do customers evaluate those features?
- do they agree?

**CFs only**

<table>
<thead>
<tr>
<th>Crude Feature</th>
<th>Total</th>
<th>Pos</th>
<th>Neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>camera</td>
<td>57</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>picture</td>
<td>15</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>viewfinder</td>
<td>12</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lcd</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>image quality</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>image</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>display</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>shot</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Benefits in term of KCAP

... answer the same questions
- different levels of abstraction
- less redundancy
- more familiar terms

CF UDF mapping

<table>
<thead>
<tr>
<th>Table 5: Informative Mapping Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDF</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Editing/Viewing</td>
</tr>
<tr>
<td>LCD Display</td>
</tr>
<tr>
<td>Viewfinder</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4: Reduced Redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDF</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Image</td>
</tr>
<tr>
<td>Image Type</td>
</tr>
<tr>
<td>TIFF</td>
</tr>
<tr>
<td>...</td>
</tr>
<tr>
<td>Resolution</td>
</tr>
<tr>
<td>Effective Pixels</td>
</tr>
<tr>
<td>...</td>
</tr>
</tbody>
</table>

image
picture
shot
Conclusions

• Novel approach to feature extraction step in KCAP from evaluative text
• Mixed-initiative mapping of flat list of extracted CF into a UDF hierarchy
• Term similarity metrics

• Evaluation of these metrics on two corpora of customer reviews: reasonable accuracy, substantial reduction in redundancy

• Beneficial in term of captured knowledge
Future Work

• Improve mapping method
  – Try other term similarity measures (corpus based)
  – Inject more sophisticated NLP (e.g., weight scoring considering headword of a term)

• Develop interface to support user revision of the mapping and of the UDF hierarchy

• Summarize and present extracted knowledge to user ……….. Combine text and graphics…. Adapt techniques for generating evaluative text
Questions 2015-2

• Is WordNet the best online lexical database?!? 😊
• Who is the user?
• UDFs / CFs / Gold Standard
• Unplaced CFs
• CF extraction and polarity (how many methods?)
• Constructing large UDF
• Different Languages
• Threshold
• Future
  – Microsoft Research took this over in 2007-8
  – Interactive Multimedia Summarization (Visualization)
  – Lexical Similarity vs. corpus-based
  – Automatically create UDFs: Extract Hierarchy from the reviews/ from existing ontologies - Speech input... Sarcasm
Aspect-based sentiment summarization
US 8799773 B2

ABSTRACT

Phrases in the reviews that express sentiment about a particular aspect are identified. Reviewable aspects of the entity are also identified. The reviewable aspects include static aspects that are specific to particular types of entities and dynamic aspects that are extracted from the reviews of a specific entity instance. The sentiment phrases are associated with the reviewable aspects to which the phrases pertain. The sentiment expressed by the phrases associated with each aspect is summarized, thereby producing a summary of sentiment associated with each reviewable aspect of the entity. The summarized sentiment and associated phrases can be stored and displayed to a user as a summary description of the entity.

IMAGES (9)
Data and Gold Standard

Two products: Digital Camera and DVD

- **CFs** from Hu&Liu annotated corpora: 101 CFs for digital camera, 116 for DVD
- **UDFs** developed by domain experts: 86 UDFs for digital camera, 38 for DVD

Gold Standard Development:
- We manually developed initial mappings
- User study: we asked 7 subjects to fix our mappings with some random errors
- Based on their input a final version was created
Measures of mapping quality

- (Graphical “distance” from correct placement)

\[ \text{placement}_\text{distance}(c_{fi}) = \text{avg}(\text{edgeCount}(c_{fi})) \]

- (Fraction of redundant CF's)

\[ \text{redundant}_\text{reduc} = \frac{|\text{placedCF}| - |\text{nonEmptyUDF}|}{|\text{CF}|} \]

Can be maximized by placing all CFs in one UDF but…

\[ \text{redundant}_\text{reduc} \text{ in Gold Stand. DCam} = .45 ; \text{ DVD} = .43 \]
\[
\begin{align*}
\text{sim}_{\text{path}}(c_1, c_2) &= -\log \text{pathlen}(c_1, c_2) \\
\text{sim}_{\text{Resnik}}(c_1, c_2) &= -\log P(\text{LCS}(c_1, c_2)) \\
\text{sim}_{\text{Lin}}(c_1, c_2) &= \frac{2 \times \log P(\text{LCS}(c_1, c_2))}{\log P(c_1) + \log P(c_2)} \\
\text{sim}_{\text{je}}(c_1, c_2) &= \frac{1}{2 \times \log P(\text{LCS}(c_1, c_2)) - (\log P(c_1) + \log P(c_2))} \\
\text{sim}_{\text{eLesk}}(c_1, c_2) &= \sum_{r, q \in \text{RELS}} \text{overlap}(\text{gloss}(r(c_1)), \text{gloss}(q(c_2)))
\end{align*}
\]
Talk Summary

Corpus of Evaluative Documents

[Corpus of Evaluative Documents]

Extract evaluative info: feature hierarchy annotated with evaluations

[Extract evaluative info: feature hierarchy annotated with evaluations]

Generate NL Summary

[Generate NL Summary]

Allow access to original sources:
- Text footnotes
- Treemap zooming

[Allow access to original sources: Text footnotes Treemap zooming]

Present As Treemaps

[Present As Treemaps]

Multimedia Summary

[Media Summary]

Interactive

[Interactive]

Present NL Summary

[Present NL Summary]

SEA (NLG abstractor)
MEAD* (extractor)

[SEA (NLG abstractor) MEAD* (extractor)]

[KCAP ’05]

[KCAP ’05]

[EACL ’06]

[EACL ’06]

[INLG ’08]

[INLG ’08]

[IUI ’06] [IUI ’09]

[IUI ’06] [IUI ’09]


Corpus of Evaluative Documents

Extract evaluative info

Generate NL Summary

Present in Graphics

Present NL Summary

Multimedia Summary

Interactive

Allow access to Original sources
Extracted evaluative info after mapping

- Merged Features hierarchy annotated with all the evaluations each feature received in the corpus

Canon G3 PS Digital Camera \([-1,-1,+1,+2,+2,+3,+3,+3]\)

1. User Interface \([+2]\)
   - Button \([+1]\)
   - Menus \([+2,+2,+2,+3,+3]\)
   - Lever \([\ ]\)

2. Convenience \([\ ]\)
   - Battery \([\ ]\)
     - Battery life \([-1,-1,-2]\)
     - Battery charging system \([\ ]\)

3. ....
Conveying extracted info with graphics

Visualization should convey:
• Hierarchical organization of the features
• For each feature
  – # of evaluations
  – polarity/strength of the evaluations (average?)

Treemaps: space-filling technique for visualizing hierarchical information structures
• Each node in the hierarchy is represented as a rectangle
• Descendants of a node are represented as nested rectangles
• Rectangle size and colour can express information about the node
Treemap: stock market
One possible Treemap

• Each product feature is represented as a rectangle
• The hierarchy is represented by nesting
• Rectangle *size* expresses # of evaluations
• Rectangle *colour* expresses avg polarity/strength of evaluations (black for neutral, the more positive/negative the more green/red)
Apex DVD Player

Disk Format

- CD
- DVD Audio
- DVD-
- Disc Formats self

Extra Features

- Extra Features self
- Universal Remote Control

User Interface

- Surround Sound Sup
- Surround Sound

Video Output

- Progr
- S
- Video Output self
Another possible Treemap

- Each evaluation is represented as a rectangle
- The hierarchy is represented by nesting
- Rectangle *colour* expresses polarity/strength of the evaluation (black for neutral, the more *positive/negative* the more green/red)

- Note: More effective in conveying controversiality
### Apex AD2600 Progressive Scan DVD Player

#### Root

<table>
<thead>
<tr>
<th>Disc Formats</th>
<th>DVD Formats</th>
<th>Extra Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD</td>
<td>DVD</td>
<td></td>
</tr>
<tr>
<td>audio</td>
<td>disc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>disc_copy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dvd</td>
<td></td>
</tr>
<tr>
<td></td>
<td>format</td>
<td></td>
</tr>
</tbody>
</table>

#### JPEG

<table>
<thead>
<tr>
<th>JPEG player</th>
<th>player</th>
</tr>
</thead>
</table>

#### Surround Sound

<table>
<thead>
<tr>
<th>Surround Sound Setup</th>
<th>User Interface</th>
<th>Video Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>audio</td>
<td>display</td>
<td>color</td>
</tr>
<tr>
<td>sound</td>
<td>windows sound</td>
<td>freeze</td>
</tr>
</tbody>
</table>

#### Video Output

<table>
<thead>
<tr>
<th>Video Output</th>
<th>Video Output self</th>
</tr>
</thead>
<tbody>
<tr>
<td>picture</td>
<td>picture</td>
</tr>
<tr>
<td>video</td>
<td>video</td>
</tr>
</tbody>
</table>
Multimedia Interactive Approach

- Corpus of Evaluative Documents
  - Extract evaluative info
  - Generate NL Summary
  - Present in Graphics
  - Present NL Summary
  - Multimedia Summary

- Allow access to Original sources

Interactive
Summary of customer reviews for: Apex AD2600 Progressive-scan DVD player

Most customers disliked the Apex AD2600. Although many customers found the user interface to be good, many users thought the available video outputs was poor. However, many users liked the range of compatible disc formats, even though many customers found the compatibility with DVD audio discs to be very poor.

For the price, it’s a very nice dvd player. The front door is misaligned on my unit and you have to manually lift it up just so slightly for the door to close, a very annoying thing after a while. It does play a wide range of formats as advertised which is very nice. And so far have not had any problems with dvds not being able to play. Recommended to anyone looking to purchase a low priced dvd player and not expecting any bells or whistles from a brand name one like Sony.
Multimedia Interactive Summary: Formative Evaluation

• **Procedure** *(similar to study-1 and study-2)*

• Interested in testing effectiveness of text graphics combination (redundancy / support)

• Very positive feedback (Details in IUI-06 paper)

• **Recent Extension** to comparison of two entities (see paper in IUI-09)
Talk Summary

Corpus of Evaluative Documents

Extract evaluative info: feature hierarchy annotated with evaluations

Generate NL Summary

Present As Treemaps

Present NL Summary

Multimedia Summary

Allow access to original sources:
- Text footnotes
- Treemap zooming

Interactive

[IUI ’06] [IUI ’09]

[SEA (NLG abstractor) MEAD* (extractor)]

[KCAP ’05]

[EACL ’06] [INLG ’08]
Questions 2015

• UDFs / CFs / Gold Standard
• Unplaced CFs
• Clarification Placement distance
• CF extraction and polarity
• Constructing large UDF
• Different Languages
• Trade-off Placement and Redundancy
• Future
  – MSR
  – Interactive Multimedia Summarization
  – Extract Hierarchy from the reviews (automatically create UDFs)….. Speech input… Sarcasm

11/9/2015

KCAP 2005
Placement Distance

• The accuracy of a CF term in the research is assessed by considering the hierarchical path distance between where it is placed by the mapping algorithm and where it is placed by the gold standard (control mapping). Does the research assume that path lengths all encode the same semantic distance? (e.g. that pixels (parent) ⟷ resolution (child) has a semantic subset distance equal to image (parent) ⟷ image type (child))


Results for DVD

Table 2: Placement distance and redundancy reduction scores for DVD player with term metric $avg$

<table>
<thead>
<tr>
<th></th>
<th>1st Run</th>
<th></th>
<th>No Repetition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p_dist</td>
<td>redund</td>
<td>p_dist</td>
<td>redund</td>
</tr>
<tr>
<td>str_match</td>
<td>.31</td>
<td>.19</td>
<td>.27</td>
<td>.21</td>
</tr>
<tr>
<td>syn_score</td>
<td>.30</td>
<td>.23</td>
<td>.28</td>
<td>.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>sim_score (res)</th>
<th>sim_score (lin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.2</td>
<td>.39</td>
<td>.30</td>
</tr>
<tr>
<td>-0.4</td>
<td>.49</td>
<td>.44</td>
</tr>
<tr>
<td>-0.6</td>
<td>.59</td>
<td>.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>sim_score (res)</th>
<th>sim_score (lin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.2</td>
<td>.42</td>
<td>.36</td>
</tr>
<tr>
<td>-0.4</td>
<td>.49</td>
<td>.43</td>
</tr>
<tr>
<td>-0.6</td>
<td>.57</td>
<td>.50</td>
</tr>
</tbody>
</table>
## Results DigCam for AVG

**Table 1: Placement distance and redundancy reduction scores for DigCam with term metric $avg$**

<table>
<thead>
<tr>
<th></th>
<th>First Run</th>
<th>No Repeat</th>
<th>After Revision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p_dist</td>
<td>redun</td>
<td></td>
</tr>
<tr>
<td>str_match</td>
<td>.43</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>syn_score</td>
<td>.45</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>.42</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>.42</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>.46</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.2</td>
<td>.44</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>.43</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>.43</td>
<td>.36</td>
<td></td>
</tr>
</tbody>
</table>
Support Analysis of Evaluative Arguments (about single entity)

Corpus of relevant Evaluative Arguments

SE-SEA

SentExt Summary

UDF

AMVF

KEEA

Merge Features + strength + polarity (for each evaluation)

NLG-SEA

NLG Summary

MultiMedia Interactive summary

Value treemap

MISEA

Treemap Engine
Results: Summary

• In both products Wordnet distance with $\theta = -.2$
From Mapping To Evaluation

• Given unsupervised CF extraction and unsupervised UDF\textless{}\textrightarrow{}CF mapping, need to evaluate UDF features

• Assume we can calculate strength and polarity of customer evaluations for each CF using existing methods (Hu & Liu 2004; Wilson et al. 2004),
  – then we can generate an evaluation for each UDF based on its CF's
Back to High-level process

- Information Extraction
- Summary generation
Plan for Summary Generation

• Adapt GEA (Generator of Evaluative Arguments) (Carenini & Moore 2001) for
  – Content selection and organization
  – Microplanning (partially)
  – Realization

• Adapt existing MEAD (Radev et. al. 2001) software as baseline “domain/task independent” summarizer

• Evaluation: Compare system against baseline with human judges
Generator of Evaluative Arguments (GEA)

- Generates evaluations of entities based on:
  - properties of entity
  - user preferences about that entity
- Entity is represented as a set of attributes and values (e.g. (Zoom range . 12x))
- User Preferences are modelled using an AMVF (Additive Multiattribute Value Function)
  - This is a hierarchical set of preferences about entity, with attributes as leafs
GEA example: AMVF

House Value

  Location
    0.7
    0.3
  Amenities
    0.8
    0.2

Neighborhood
  0.4

Park-Distance
  0.6

Deck-Size
  0.8

Porch-Size
  0.4

House Value

0.3
GEA example: Attributes

- **Location**
  - **Porch**
    - **Size**
      - 0.9
      - 0.25
      - 0.6

- **Amenities**
  - **Neighborhood**
    - **Park-Distance**
      - n2
      - 0.5 km
      - 20 m
      - 2
      - 36 m
    - **Deck-Size**
      - 0.25
      - 0.6
  - **Porch-Size**

- **House Value**

<table>
<thead>
<tr>
<th>+</th>
<th>Likes it</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>Does not like it</td>
</tr>
</tbody>
</table>

- **Domain Values**
  - n2
  - 0.5 km
  - 20 m
  - 36 m

- **House-A**

- **Attribute Evaluation**
  - Component Value
  - Function

- **Domain Value**
GEA example: Opposing/Supporting Evidence

<table>
<thead>
<tr>
<th>Likes it</th>
<th>Does not like it</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$o$</th>
<th>$\text{Parent}(o)$</th>
<th>relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>supporting</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>supporting</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>opposing</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>opposing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supporting</th>
<th>Opposing</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

**Location**

- 0.7 +
- 0.3 -

**Amenities**

- 0.8 +
- 0.2 -

**Porch Size**

- 0.6 +
- 0.2 -

**Deck Size**

- 0.8 +
- 0.2 -

**Park Distance**

- 0.4 +
- 0.6 -

**Neighborhood**

- 0.9 +
- 0.6 -

**House Value**

- 0.7 +
- 0.3 -

- 0.64 +
- 0.3 -

**n2**

- 0.5 km
- 20 m²
- 36 m²
Measure of Importance [Klein 94]

For each attribute $a$:
Importance $a = w_a \max [v_a, [1 - v_a]]$
Argumentative Strategy

Based on guidelines from argumentation theory
[Miller 96, Mayberry 96]

**Selection:** include only “important” evidence
(i.e., above threshold on measure of importance)

**Organization:**

1. **Main Claim** (e.g., “This house is interesting”)
2. **Opposing** evidence
3. **Most important supporting** evidence
4. **Further supporting** evidence -- ordered by importance with *strongest last*

Strategy applied recursively on supporting evidence
Adapting GEA

**GEA**
- AMVF hierarchy ->
- AMVF weights ->
- Component ->
  Value Function

**Customer Reviews**
- UDF hierarchy
- Relative frequency of UDFs in corpus
- Aggregation of polarity/strength of UDF features
Adapting GEA (cont'd)

• Differences
  – Customers may evaluate non-leaf elements (e.g. “Location”) directly
  – in GEA domain, entities had only one evaluation for each attribute
    • For customer reviews, must give some indication of distribution of customer opinions on each attribute
Example: Some (fake) Reviews

“I really liked the Canon G3[+2]. The 12x zoom is really useful[+1]! The only thing I didn't like was its poor [-1] focussing in low light.”

“The Canon G3 is a great deal. The lens features were the best I've seen for a camera of its price [+2]. The menu system is very intuitive [+1], but I wish the camera could take RAW images [-1].”

“I really didn't like this camera [-2]. It focussed very poorly [-2] indoors (when I use it most) and I found myself wishing there were more modes on the dial [-1] rather than in the menu system. I returned mine already.”
Adapted GEA

Diagram:

- Canon G3
  - Lens
    - +0.57
    - 0.29
    - 0.5
  - Auto-Focus
    - +0.25
    - 0.17
    - 0.66
  - Interface
    - 0.43
    - 0.26
    - 0.39
  - Menu
    - 0.33
    - 0.28
    - 0.83
  - Dial
    - +0.66
    - 0.495
    - 0.25

Evaluation Aggregation Function:

- +2
- +1
- -1, -2
- +2
- -1, -2

Strength/Polarity of User Evaluations:

Attribute Evaluation
Output of GEA

• What GEA gives us:
  – High-level text plan (i.e. content selection and ordering)
  – Cue phrases for argumentation strategy ("In fact", "Although", etc.)

• What GEA does not give us:
  – Appropriate micro-planning (lexicalization).
    • Need to give indication of distribution of customer opinions
Hypothetical GEA Output

The Canon G3 is a good camera. However, the interface feature is poor. Although the menu system is good, the dial system is terrible.
Target Summary

Most users thought Canon G3 was a good camera. However, several users did not like interface. Although most users liked the menu system, many thought the dial was terrible.
Evaluation

• Current idea: task-based (extrinsic) evaluation
  – Give human test subject summary
  – Then, allow user some fixed time (e.g. 5 minutes) to scan a corpus of reviews (20-30?)
  – User should then answer e.g.
    • if summary provides “all” (?) important information
    • if summary left out information
    • if missing information was important
    • if summary is representative of corpus
  – Also evaluate fluency with known methods

Others?
Future Directions

• Current method of adapting GEA is just a first pass.
  – Could change e.g. Measure of Importance.

• We may leverage GEA's ability to create user-tailored evaluative arguments for generating user-tailored summaries (long term)
IE Key Sub-tasks

A. What features of the objects are evaluated in the reviews?

B. For each feature:

i. what is the **polarity** of the evaluation? (good vs. bad)

ii. what is the **strength** of the evaluation? (rather good vs. extremely good)
(User-Specific) Summarization of Multiple Customer Reviews

The Goal:
An automatic solution to the problem of summarizing a potentially large set of documents that contain evaluative language about a given entity (e.g., a product, a location, a job candidate, etc.)

User Specific: the summary is tailored to user’s conceptualization of the entity (now) model of the user’s preferences (long term)
Example: Some (fake) Reviews

“I really liked the Canon G3. The 12x zoom is really useful! The only thing I didn't like was its poor focussing in low light.”

“The Canon G3 is a great deal. The lens features were the best I've seen for a camera of its price. The menu system is very intuitive, but I wish the camera could take RAW images.”

“I really didn't like this camera. It focussed very poorly indoors (when I use it most) and I found myself wishing there were more modes on the dial rather than in the menu system. I returned mine already.”
Example: Target Summary

Most users liked the Canon G3. Many found the zoom feature to be good. Although many users did not like the auto focus, a few users liked the menu system. Only 1 user did not like the camera.
Example Target Summary

• Features
  – Selection of content (flash range not mentioned)
  – Discourse cues (cue phrases, order of evidence)
  – Contrasting and supporting evidence for summary of camera
  – Lexicalization of numerical tallies (2/3 => “most”)
High-level process

• Information Extraction

• Summary generation
Example of Learned Features for a Digital Camera

- noise
- function
- button
- camera
- four megapixel
- remote control
- software
- manual

- remote
- lever
- price
- Canon G3
- strap
- low light focus
- tiff format
- use
Ideal Extraction: sample form corpus  
[Hu&Liu 2004]

...... the canon computer software [+2] used to download, sort, . . . is very nice and very easy to use. The only two minor issues I have with the camera are the lens cap [-1] (it is not very snug and can come off too easily). . . .

The menus [+1] are easy to navigate and the buttons [+1] are easy to use. It is a fantastic camera [+3] and well worth the price.
Where we are now...

- Increase in accuracy scores from measure to measure, and from MAX to AVG, but it's small.

- We need to understand better how the similarity measures are working to better take advantage of them.