EE 6850 Lecture #10  (Nov. 13, 2002)

- Part II
  - Personalized Highlight Generation
  - Learning User Preference
  - Manual Annotations
- References

---

Personalized Adaptation

- Limited resources (time, bandwidth)
  - Summarize
- Relevant information is user specific
  - Personalize
- Interest in semantic information
  - Use high-level metadata
MPEG-7 increase in semantic (textual) metadata
MPEG-7 Authoring System (IBM TRL)

- Event-based indexing
  - Domain-specific pre-defined event list (keywords)
  - Operator manually annotates video (close to real time)

IBM TRL’s MPEG-7 Authoring System
Soccer Video

- Metadata & User Preferences

**Time: Event: Team**

- 00:15:51:63:Pass:Paraguay
- 00:15:54:38:Pass cut:Japan
- 00:15:54:38:Pass:Paraguay
- 00:15:57:46:Pass:Japan
- 00:15:59:18:Dribble:Japan
- 00:16:03:62:Long pass:Japan
- 00:16:04:74:Goal:Japan
- 00:16:35:08:Replay start:Japan

- How do we find the highlights for specific users?

Event Statistics

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Freq</th>
<th>O/D</th>
<th>Event</th>
<th>Freq</th>
<th>O/D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red card</td>
<td>1</td>
<td>-</td>
<td>Goal kick</td>
<td>64</td>
<td>D</td>
</tr>
<tr>
<td>Replay corner kick</td>
<td>3</td>
<td>-</td>
<td>Replay foul</td>
<td>75</td>
<td>-</td>
</tr>
<tr>
<td>Replay free kick</td>
<td>3</td>
<td>-</td>
<td>Replay shoot</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Bar/Post</td>
<td>5</td>
<td>O</td>
<td>Keeper cut</td>
<td>585</td>
<td>D</td>
</tr>
<tr>
<td>Missing pass</td>
<td>7</td>
<td>-</td>
<td>Shoot</td>
<td>165</td>
<td>O</td>
</tr>
<tr>
<td>Replay injury</td>
<td>7</td>
<td>-</td>
<td>Free kick</td>
<td>762</td>
<td>O</td>
</tr>
<tr>
<td>Replay keeper cut</td>
<td>7</td>
<td>-</td>
<td>Throw in</td>
<td>195</td>
<td>-</td>
</tr>
<tr>
<td>Goal</td>
<td>11</td>
<td>O</td>
<td>Foul</td>
<td>205</td>
<td>-</td>
</tr>
<tr>
<td>Replay centering</td>
<td>19</td>
<td>-</td>
<td>Ball cut</td>
<td>280</td>
<td>D</td>
</tr>
<tr>
<td>Yellow card</td>
<td>22</td>
<td>-</td>
<td>Through pass</td>
<td>364</td>
<td>O</td>
</tr>
<tr>
<td>Offside</td>
<td>24</td>
<td>O</td>
<td>Centering</td>
<td>323</td>
<td>O</td>
</tr>
<tr>
<td>Replay goal</td>
<td>32</td>
<td>-</td>
<td>Long pass</td>
<td>337</td>
<td>O</td>
</tr>
<tr>
<td>Player change</td>
<td>35</td>
<td>-</td>
<td>Pass cut</td>
<td>792</td>
<td>D</td>
</tr>
<tr>
<td>Shoot block</td>
<td>35</td>
<td>D</td>
<td>Dribble</td>
<td>1250</td>
<td>-</td>
</tr>
<tr>
<td>Clear</td>
<td>57</td>
<td>D</td>
<td>Pass</td>
<td>1687</td>
<td>-</td>
</tr>
<tr>
<td>Corner kick</td>
<td>57</td>
<td>O</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Proposed Framework

MPEG-7 annotation

Summaries of relevant information (not entire video)

User preferences

Highlight Generation

Approach 1: Fixed keyword-weights

- **Keyword weights (user profile) + time budget** determine highlights

| Goal 10 | Corner-kick 8 | Keeper-cut 7 | Pass 2 ...
|---------|--------------|--------------|-----------|

User profile

Indicates Event importance

EE 6850, F02, Chang, Columbia U.
Trigger Functions

- Events represented by “trigger” functions

**More Trigger Functions**

- **Digital**
  
  \[ f(t) = w \]

- **Oval**
  
  \[ f(t) = \frac{\sqrt{1 - a(t - t_1)^2}}{b} \]

- **Linear**
  
  \[ f(t) = at + b \]

- **Gaussian**
  
  \[ f(t) = w \cdot \exp\left\{ -\frac{(t - t_1)^2}{2\sigma^2} \right\} \]

- **Pulse**
  
  \[ f(t) = -\frac{\sqrt{1 - a(t - t_1 + s)^2}}{b} + w \]
Limitations & A Learning Approach

- Constructing profiles is difficult
  - How to assign event weights / determine trigger functions/parameters?

- Exploit domain-specific information
- Use high-level metadata features over time
- Learn preferences

Domain-Specific Information

- Some keywords unambiguously represent important events
  - Goal, Corner-kick, Red Card …

- BUT different instances of the same event cannot be distinguished, e.g., same event at different times
  - Centering (2:30)
  - Centering (8:38)

Use domain-specific information AND time windows
Activity Windows

- Treat metadata as discrete time signal
- Extract metadata features
  - Overlapping windows $[t_1, t_2], [t_1+\delta, t_2+\delta], \ldots$
- Features:
  - Number of events, No. of location events, No. of interruptions, No. of offenses/defenses, play time

![Diagram showing overlapping windows and events]

High-level Metadata Features

- Detailed domain-specific analysis yields domain-specific event-categories

- **Throw-in**: Start action
- **Foul**: End action
- **Pass**: Ball in play
- **Red card**: Ball not in play
- **Corner-kick**: Determine location
- **Red-card**: Affect outcome
- **Goal**: Change score
- **Block**: Defensive
- **Shoot**: Offensive
Learning Highlights

Operator annotates video

User watches video
Selects highlight segments
ONLY clicks on interesting segments

User(s)-specific Segment Classifier
• Highlight
• No highlight

Machine Learning Algorithm

Viewing Highlights

Operator annotates video

Activity window
Feature Extraction

User selects time budget & views highlights

User(s)-specific Segment Classifier
• Highlight
• No highlight

EE 6850, F02, Chang, Columbia U.
Experimental Setup

- 6 soccer games (~ 90 mins. each) with detailed event metadata
- 1 subject selected personal highlights → ground truth
- Features
  - No. of events
  - No. location events
  - No. Interruptions
  - No. defensive events
  - No. offensive events
  - Play time

Experimental Results

10-fold Cross-validation

<table>
<thead>
<tr>
<th>Set</th>
<th>Method</th>
<th>Prec.</th>
<th>Recall</th>
<th>Prec.</th>
<th>Recall</th>
<th>Prec.</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1-Nearest N.</td>
<td>78%</td>
<td>69%</td>
<td>72%</td>
<td>80%</td>
<td>54%</td>
<td>36%</td>
</tr>
<tr>
<td>A</td>
<td>3-Nearest N.</td>
<td>80%</td>
<td>77%</td>
<td>80%</td>
<td>80%</td>
<td>60%</td>
<td>35%</td>
</tr>
<tr>
<td>B</td>
<td>Neural N.</td>
<td>79%</td>
<td>83%</td>
<td>80%</td>
<td>85%</td>
<td>63%</td>
<td>40%</td>
</tr>
<tr>
<td>C</td>
<td>Human</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>89%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Training Sets

- A: All highlights, same No. of neg. examples (3,156 instances)
- B: 67% of the highlights, same No. of neg. examples (1,986 instances)
- C: All examples (14% positive, 86% negative) (10,917 instances)
Recap [Jaimes et al 02]

- **Accurate semantic information annotation**
  - Annotation can be done in *almost* real time

- **Application of using metadata to generate personalized video highlights**
  - Cell phones, PDAs, etc.

- **Determining highlights event from detailed metadata**
  - Not trivial

- **A framework that learns user preferences to generate video highlights**
  - Exploits *domain-specific* information
  - Uses high-level metadata features over in time windows
  - Learns preferences

- **Other scenarios**
  - *Fast paced segment*:
    - Play time, No. interruptions
  - *Offensive segment*:
    - No. offensive/defensive events, etc.
  - *Exciting segment*:
    - Average score, no. of events, etc.