SkyLens: Visual Analysis of Skyline on Multi-dimensional Data

Xun Zhao, Yanhong Wu, Weiwei Cui, Xinnan Du, Yuan Chen, Yong Wang, Dik-Lun Lee, and Huamin Qu
Background

• Multi-criteria decision making

Employee recruitment  University selection  Car comparison
Background

• Suppose you are a college basketball coach, how do you recruit the best players?

<table>
<thead>
<tr>
<th>PLAYER</th>
<th>TEAM</th>
<th>AGE</th>
<th>GP</th>
<th>W</th>
<th>L</th>
<th>MIN</th>
<th>OFFRTG</th>
<th>DFRTG</th>
<th>NETRTG</th>
<th>AST%</th>
<th>AST/TO</th>
<th>AST Ratio</th>
<th>DREB%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJ Hammons</td>
<td>DAL</td>
<td>24</td>
<td>22</td>
<td>18</td>
<td>4</td>
<td>7.4</td>
<td>102.2</td>
<td>102.8</td>
<td>-0.6</td>
<td>3.8</td>
<td>0.40</td>
<td>6.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Aaron Brooks</td>
<td>IND</td>
<td>32</td>
<td>65</td>
<td>36</td>
<td>29</td>
<td>13.7</td>
<td>101.5</td>
<td>104.6</td>
<td>-3.0</td>
<td>21.6</td>
<td>1.89</td>
<td>24.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Aaron Gordon</td>
<td>ORL</td>
<td>21</td>
<td>80</td>
<td>29</td>
<td>51</td>
<td>28.7</td>
<td>105.4</td>
<td>108.2</td>
<td>-2.8</td>
<td>9.7</td>
<td>1.69</td>
<td>12.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Aaron Harrison</td>
<td>CHA</td>
<td>22</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>3.3</td>
<td>83.3</td>
<td>101.9</td>
<td>-18.6</td>
<td>37.5</td>
<td>0.00</td>
<td>38.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Adrian Payne</td>
<td>MIN</td>
<td>26</td>
<td>18</td>
<td>13</td>
<td>7</td>
<td>7.5</td>
<td>102.6</td>
<td>101.8</td>
<td>0.8</td>
<td>8.9</td>
<td>0.88</td>
<td>9.0</td>
<td>6.9</td>
</tr>
<tr>
<td>Al Horford</td>
<td>BOS</td>
<td>31</td>
<td>68</td>
<td>46</td>
<td>22</td>
<td>32.3</td>
<td>110.7</td>
<td>105.8</td>
<td>5.0</td>
<td>23.9</td>
<td>2.93</td>
<td>25.7</td>
<td>4.9</td>
</tr>
<tr>
<td>Al Jefferson</td>
<td>IND</td>
<td>32</td>
<td>66</td>
<td>33</td>
<td>33</td>
<td>14.1</td>
<td>102.3</td>
<td>108.1</td>
<td>-5.8</td>
<td>11.4</td>
<td>1.73</td>
<td>9.5</td>
<td>8.6</td>
</tr>
<tr>
<td>Al-Farouq Aminu</td>
<td>POR</td>
<td>26</td>
<td>61</td>
<td>33</td>
<td>28</td>
<td>29.1</td>
<td>107.7</td>
<td>105.9</td>
<td>1.8</td>
<td>8.2</td>
<td>1.05</td>
<td>13.8</td>
<td>4.9</td>
</tr>
<tr>
<td>Alan Anderson</td>
<td>LAC</td>
<td>34</td>
<td>30</td>
<td>20</td>
<td>10</td>
<td>10.3</td>
<td>103.1</td>
<td>114.0</td>
<td>-10.8</td>
<td>5.2</td>
<td>1.57</td>
<td>10.5</td>
<td>1.1</td>
</tr>
<tr>
<td>Alan Williams</td>
<td>PHX</td>
<td>24</td>
<td>47</td>
<td>11</td>
<td>36</td>
<td>15.1</td>
<td>105.6</td>
<td>105.8</td>
<td>-0.3</td>
<td>4.9</td>
<td>0.62</td>
<td>6.1</td>
<td>13.8</td>
</tr>
<tr>
<td>Alec Burks</td>
<td>UTA</td>
<td>25</td>
<td>42</td>
<td>26</td>
<td>16</td>
<td>15.5</td>
<td>105.0</td>
<td>104.9</td>
<td>0.1</td>
<td>7.4</td>
<td>0.86</td>
<td>8.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Alex Abrines</td>
<td>OKC</td>
<td>23</td>
<td>68</td>
<td>37</td>
<td>31</td>
<td>15.5</td>
<td>106.0</td>
<td>108.3</td>
<td>-2.3</td>
<td>5.5</td>
<td>1.21</td>
<td>9.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Alex Len</td>
<td>PHX</td>
<td>24</td>
<td>77</td>
<td>21</td>
<td>56</td>
<td>20.3</td>
<td>99.4</td>
<td>110.5</td>
<td>-11.1</td>
<td>4.3</td>
<td>0.43</td>
<td>6.3</td>
<td>10.4</td>
</tr>
</tbody>
</table>
Introduction – Skyline

• **Skyline algorithm**: automatically select the *skyline* of the dataset

• In database, skyline algorithm is an important and extensively studied problem
Introduction – Skyline

- **Skyline algorithm**: automatically select the *skyline* of the dataset
Introduction – Skyline Definition

- **Skyline**: a set of superior points that are not dominated by other points in the dataset

- **Dominance**:
  - If \( p \) dominates \( q \), then:
    - \( p \) is not worse than \( q \) in all attributes
    - \( p \) is at least better than \( q \) in one attribute
**Introduction – Skyline Example**

- **Skyline**: a set of superior points that are not dominated by other points in the dataset

<table>
<thead>
<tr>
<th>Players</th>
<th>Block</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Bob</td>
<td>20</td>
<td>25</td>
</tr>
</tbody>
</table>

Bob dominates Alan (*block & rebound*)
Introduction – Skyline Example

- **Skyline**: a set of superior points that are not dominated by other points in the dataset

<table>
<thead>
<tr>
<th>Players</th>
<th>Block</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Bob</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Calvin</td>
<td>25</td>
<td>20</td>
</tr>
</tbody>
</table>

Calvin dominates Alan (**block** & **rebound**)
Introduction – Skyline Example

- **Skyline**: a set of **superior** points that are not **dominated** by other points in the dataset

<table>
<thead>
<tr>
<th>Players</th>
<th>Block</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Bob</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Calvin</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Daniel</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Daniel dominates Alan *(block)*
Introduction – Skyline Example

- **Skyline**: a set of *superior* points that are not *dominated* by other points in the dataset

<table>
<thead>
<tr>
<th>Players</th>
<th>Block</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Bob</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Calvin</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Daniel</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

Points: Daniel > Calvin > Bob
Rebound: Bob > Calvin > Daniel
Introduction – Skyline Example

- **Skyline**: a set of **superior** points that are not **dominated** by other points in the dataset

<table>
<thead>
<tr>
<th>Players</th>
<th>Block</th>
<th>Rebound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Bob</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Calvin</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Daniel</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

**Skyline**: Bob, Calvin, Daniel
Introduction – Challenges

Scalability

The size of skyline increases with the number of attributes

Interpretation

The reasons that make a point in skyline is unclear

Comparison

The strength and weakness of each skyline point is implicit
SkyLens – Visual Components

Projection View

Tabular View

Comparison View

Skyline point A
Skyline point B
Skyline point C
SkyLens – Demo!
SkyLens – Video
Projection View: provide an overview of skyline (clusters and outliers)
Projection View

- Methods: t-SNE projection and skyline glyphs
Projection View – Skyline Glyph

- **Normal mode**: show the attribute value distribution of skyline

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attr. I</td>
<td>5</td>
</tr>
<tr>
<td>Attr. II</td>
<td>3</td>
</tr>
<tr>
<td>Attr. III</td>
<td>7</td>
</tr>
<tr>
<td>Attr. IV</td>
<td>1</td>
</tr>
<tr>
<td>Attr. V</td>
<td>3</td>
</tr>
<tr>
<td>Attr. VI</td>
<td>1</td>
</tr>
</tbody>
</table>

…
Projection View – Skyline Glyph

- **Normal mode**: show the attribute value distribution of skyline
- **Dominating score (superiority metric)**:
  - # of points dominated by this point

![Projection View](image)

- *Rebound*
- *Points*

![Skyline Glyph](image)

- *Higher*
- *Lower*

- Dominating score
- Calvin
Projection View – Skyline Glyph

- **Focus mode**: highlight how other points differ from a focused one

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Point A</th>
<th>Point B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attr. I</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Attr. II</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Attr. III</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Attr. IV</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Attr. V</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Attr. VI</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

Point A (focused point)

Point B
**Projection View – Skyline Glyph**

- **Focus mode**: highlight how other points differ from a focused one using color map

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Point A</th>
<th>Point B (diff.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attr. I</td>
<td>5</td>
<td>3 (-2)</td>
</tr>
<tr>
<td>Attr. II</td>
<td>3</td>
<td>4 (+1)</td>
</tr>
<tr>
<td>Attr. III</td>
<td>7</td>
<td>6 (-1)</td>
</tr>
<tr>
<td>Attr. IV</td>
<td>1</td>
<td>5 (+4)</td>
</tr>
<tr>
<td>Attr. V</td>
<td>3</td>
<td>5 (+2)</td>
</tr>
<tr>
<td>Attr. VI</td>
<td>1</td>
<td>3 (+2)</td>
</tr>
</tbody>
</table>

...
Projection View – Case Studies

Games played

Field goal percentage

**Two outliers** with high values on games played and FG%
Projection View – Case Studies

Lamar Odom has the largest dominating score (central circle color)
Projection View – Case Studies

Switching to focus mode: three clusters can be found.
Projection View – Case Studies

Switching to *focus mode*: three clusters can be found

Point made-related Attr.

Lamar Odom

Dwight Howard
(defense player)

LeBron James
(pointer player)
Projection View – Case Studies

Switching to focus mode: three clusters can be found

LeBron James (point player)
Lamar Odom (defense player)
Chris Paul (assist player)
Dwight Howard

Assist-related Attr.
SkyLens – Tabular View

Tabular View: infer the underlying reasons that make a point in skyline
Tabular View

- Methods: matrix representation & in-cell bar chart visualization
Tabular View

- Methods: matrix representation & in-cell bar chart visualization
- Each row represents a skyline point
Tabular View

• Methods: matrix representation & in-cell bar chart visualization
  • Each row represents a skyline point
  • Each column represents an attribute
Tabular View

- Table cell – divergent bar chart visualization
- Goal: summarize the overall differences among skyline points
Tabular View

• Table cell – divergent bar chart visualization
  • Goal: summarize the overall differences between skyline points

Attr. I

Point A

Each **vertical bar** represents a **skyline point**:  
- current point (Point A)  
- other points
Tabular View

- Table cell – divergent bar chart visualization
- Goal: summarize the overall differences between skyline points

**Attr. I**

**Point A**

**Bar order**: sorted by the current attribute value (Attr. I)
Tabular View

- Table cell – divergent bar chart visualization
- Goal: summarize the overall differences between skyline points

<table>
<thead>
<tr>
<th>Point A</th>
<th>Positive diff.</th>
<th>Negative diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attr. 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bar length: other skyline points’ average value differences compared with point A
Tabular View – Case Studies

- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences between skyline points

Attr. I: games played
Attr. II: points
Attr. III: assist
Attr. IV: block

LeBron James:
- High ranking on points & assist
Tabular View – Case Studies

• Table cell – divergent bar chart visualization
  • Goal: summarize the overall differences between skyline points

LeBron James:
• High ranking in points & assist
• Better overall performance than most skyline points

Attr. I: games played

Attr. II: points

Attr. III: assist

Attr. IV: block
Tabular View – Case Studies

- Table cell – divergent bar chart visualization
  - Goal: summarize the overall differences between skyline points

Attr. I: game played

Attr. II: points

Attr. III: assist

Attr. IV: block

LeBron James:
- High ranking in points & assist
- Better overall performance than most skyline points
  - Dwight has an overall comparable performance with LeBron
Tabular View

• Table cell interaction: expanding a row for detailed information
Tabular View

• Table cell interaction: expanding a row for detailed information
Tabular View

- Table cell interaction: expanding a row for detailed information

Each column is an extension of the corresponding vertical blue bar and represents the same skyline point.
Tabular View

- Table cell interaction: expanding a row for detailed information

Matrix cell color: the value difference of the corresponding attribute (Attr. III)
Dwight Howard:
• No players has better performance than him in defense-related attributes
Tabular View – Case Studies

Dwight Howard:
• No players has better performance than him in defense-related attributes
• Many players outperform him in AST
SkyLens – Comparison View

Projection View

Tabular View

Comparison View

Comparison View: support a thorough comparison between skyline points
Comparison View

- Methods: radar charts & domination glyphs
  - Comparing attribute values
  - Examining dominating scores
  - Investigating dominated points

Goal: a thorough comparison on 2 ~ 5 skyline points
Comparison View

• Radial layout for the radar charts & domination glyphs

2-point comparison  3-point comparison  4-point comparison
The radar chart encodes: attribute values and dominating score.
Comparison View – Domination Glyph

Domination glyph: summarize the differences among skyline points from a domination perspective.
Comparision View – Domination Glyph

Inner sector angle: the ratio of points that are dominated by skyline point B.
Comparison View – Domination Glyph

Outer sector angle: the ratio of points that are exclusively dominated by skyline point B
Comparison View – Domination Glyph Interaction

Hovering interaction: pop-up window showing the overlaid radar chart

Gray polylines represent the points that are exclusively dominated by skyline point B
Comparison View – Case Studies

Comparing Dwight, LeBron, and Chris in the perspective of domination relation
Comparison View – Case Studies

Comparing Dwight, LeBron, and Chris in the perspective of domination relation

Dwight and LeBron have similar dominating scores. Chris has a smaller dominating score compared with them.
Comparison View – Case Studies

Hovering over the points that are exclusively dominated by Chris against LeBron
Comparison View – Case Studies

Hovering over the points that are exclusively dominated by Chris against LeBron
Evaluation – Case Studies

• Two case studies using the NBA and Numbeo quality-of-life data

• NBA 2010 - 2011 regular season statistics
  • 452 players and 12 numerical attributes

• Numbeo quality-of-life data
  • 176 cities and 8 numerical attributes
Evaluation – User Study

• Qualitative user study
  • 12 participants recruited from the local university
  • 10 tasks covering all important aspects in skyline analysis
  • 19 questions related with SkyLens usage in a post-session interview
Future Work

• Include nominal attribute analysis

• Support data with uncertain values

• Track temporal changes of skyline
Q&A

SkyLens: Visual Analysis of Skyline on Multi-dimensional Data

Xun Zhao
Contact: xzhaoag@ust.hk
Project page: http://zhaoxun.me/skylens
Agenda

• Background
• Introduction
• SkyLens
  • Projection View
  • Tabular View
  • Comparison View
• Evaluation
• Future Work
Comparison View – Radar charts

Attribute value: radar chart
Comparison View

Ranking on each attribute: radius of circle on axis
Comparison View

Dominating score: radius of dashed circles