On (not) indexing quadratic form distance by metric access methods

Tomáš Skopal
Tomáš Bartoš
Jakub Lokoč

SIRET Research Group
Faculty of Mathematics and Physics
Charles University in Prague
Overview

- Quadratic Form Distance (QFD)
- **QMap model**
  - QFD to $L_2$ Space Transformation
- **QMap and MAMs**
  - Experimental evaluation
- Conclusion
Technical Background

- Content-based similarity searching
  - Pair-wise similarity
  - High-dimensional (feature) vectors

- Fast query processing
  - Efficiency
  - Effectiveness
Quadratic Form Distance (QFD)

- Similarity measuring:

\[
QFD_A (u,v) = \sqrt{(u - v)^T A(u - v)}
\]

- \( u, v \): feature vectors \((1 \times n)\)
- \( A \): similarity matrix/QFD matrix \((n \times n)\)
  - positive definite \((zA^z > 0)\)
  - static / dynamic correlations
  - data independent
Quadratic Form Distance (QFD)

- Applications
  - QBIC project (Querying Images by content)
  - 2D & 3D shapes
  - Protein structures
  - MindReader

- Advanced
  - SQFD (Signature QFD)
Indexing QFD

- Transformation approaches
  - QBIC system
- Lower-bounding (e.g. Faloutsos et. al 1994)
  - Contractive reduction techniques
  - SVD / KLT decompositions
- Combination (e.g. Hafner et. al 1995)
  - Transformation to k-dimensional $L_p$ space
Motivation

- Metric Access Methods (MAMs)
  - Effective/efficient similarity searching
  - Reduce distance computations
  - Complexity depends on distance function

- QFD is considered as expensive – $O(n^2)$
  - Indexing needed

- We show the transformation of QFD
  - Obtain cheaper distance function – $O(n)$
QFD vs. $L_2$ comparison (correlated dimensions)

<table>
<thead>
<tr>
<th>QFD</th>
<th>Euclidean ($L_2$) distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlated dimensions</td>
<td>✓</td>
</tr>
<tr>
<td>Expensive – $O(n^2)$</td>
<td>✗</td>
</tr>
<tr>
<td>Independent dimensions</td>
<td>✗</td>
</tr>
<tr>
<td>Cheap – $O(n)$</td>
<td>✓</td>
</tr>
</tbody>
</table>
QMap model

- Transform QFD space
  - $L_2$ instead of QFD
  - Preserving distances (homeomorphism)
QMap model

- Transform QFD space
  - $L_2$ instead of QFD
  - Preserving distances (homeomorphism)
QMap model

- **Rotate** (weighted L$_2$ space)
- **Scale** (L$_2$ space)

- Transformation matrix $B$
  - obtained by Cholesky decomposition: $A = BB^T$
QFD to $L_2$
Space Transformation

1. $QFD_A (u,v) = \sqrt{(u - v)A(u - v)^T}$
2. Cholesky decomposition: $BB^T = A$
3. $QFD (u,v) = \sqrt{(u - v)BB^T(u - v)^T}$
4. $QFD (u,v) = \sqrt{[(u - v)B][(u - v)B]^T}$
5. $QFD (u,v) = \sqrt{(uB - vB)(uB - vB)^T}$
6. $L_2 (u',v') = \sqrt{(u' - v')(u' - v')^T}$
Experiments

- Application of QMap in MAM
  - Sequential (SEQ) file
  - Pivot Table
  - M-tree

- 1,000,000 images (512 dimensional RGB histogram)

- Actions
  - Indexing
  - Querying
## Indexing - time complexity results

<table>
<thead>
<tr>
<th>Method (model)</th>
<th>“Winner”</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ file (QFD)</td>
<td>QFD</td>
</tr>
<tr>
<td>SEQ file (QMap)</td>
<td></td>
</tr>
<tr>
<td>Pivot Table (QFD)</td>
<td>QMap</td>
</tr>
<tr>
<td>Pivot Table (QMap)</td>
<td></td>
</tr>
<tr>
<td>M-tree (QFD)</td>
<td>QMap</td>
</tr>
<tr>
<td>M-tree (QMap)</td>
<td></td>
</tr>
</tbody>
</table>
Indexing - Experiments

Sequential file

- QMap model
- QFD model

M-tree

- QFD model
- QMap model

EDBT 2011, Uppsala, Sweden

23rd March 2011
<table>
<thead>
<tr>
<th>Method (model)</th>
<th>“Winner”</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ file (QFD) SEQ file (QMap)</td>
<td>QMap</td>
</tr>
<tr>
<td>Pivot Table (QFD) Pivot Table (QMap)</td>
<td>QMap</td>
</tr>
<tr>
<td>M-tree (QFD) M-tree (QMap)</td>
<td>QMap</td>
</tr>
</tbody>
</table>
Querying Experiments (kNN)

**Pivot table**

- Database size 1M

- QFD model
- QMap model

**M-tree**

- Database size 1M

- QFD model
- QMap model
Contributions

- QMap model
  - Space transformation: \( \text{QFD} \rightarrow L_2 \)
  - Distance-preserving (homeomorphic)
  - Data-independent
  - Output is explicitly formulated

- QMap model is separated from the usage of any access methods
  - Superior performance
Thank you for your attention.

SIRET Research Group
http://siret.ms.mff.cuni.cz