Bounds on Lengths of Real Valued Vectors Similar with Regard to the Tanimoto Similarity

Marzena Kryszkiewicz
Institute of Computer Science
Warsaw University of Technology

Deriving Bounds on Lengths of Tanimoto Similar Real Valued Vectors…

Property. Let u and v be vectors. Then:

$$u \cdot v \leq |u||v|$$

Proof.

The inequality holds trivially if u or v is a zero vector.

Otherwise, $cos(\angle(u,v)) = \frac{u \cdot v}{|u||v|}$ and $cos(\angle(u,v)) \leq 1$.

Hence, $u \cdot v = cos(\angle(u,v))|u||v| \leq |u||v|$. □

Deriving Bounds on Lengths of Tanimoto Similar Real Valued Vectors…

Theorem. Let u and v be non-zero vectors such that $T(u, v) \geq \varepsilon$ and $\varepsilon \in (0,1)$. Let

$$\alpha_1 = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) - \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right), \quad \alpha_2 = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) + \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right).$$

Then: $|v| \in [\alpha_1 |u|, \alpha_2 |u|]$.

Proof. $\varepsilon \leq T(u,v) = \frac{u \cdot v}{|u|^2 + |v|^2 - u \cdot v} \leq \frac{|u||v|}{|u|^2 + |v|^2 - |u||v|}$.

So, $\varepsilon |v|^2 - (\varepsilon + 1)|u||v| + \varepsilon |u|^2 \leq 0$ … □

Deriving Bounds on Lengths of Tanimoto Similar Real Valued Vectors

Property. Let:

$$\alpha_1 = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) - \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right), \quad \alpha_2 = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) + \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right).$$

Then: $\alpha_1 = \frac{1}{\alpha_2}$.

Theorem. Let u and v be non-zero vectors such that $T(u, v) \geq \varepsilon$ and $\varepsilon \in (0,1)$. Let $\alpha = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) + \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right)$.

Then: $|v| \in \left[\frac{1}{\alpha}|u|, \alpha |u|\right]$.

Example: Using bounds on lengths for searching Tanimoto similar vectors for $\varepsilon = 0.85$…

Id	(non-zero dimension, value) pairs	length
v5	{(4, 4.0), (5,-3.0), (7, 3.0)}	5.83
v9	**{(4,-2.0), (5, 3.0), (7, 3.0), (9, 5.0)}**	**6.86**
v3	**{(3, 6.0), (4, 4.0)}**	**7.21**
v7	**{(3, 6.0), (4, 4.0)}**	**7.21**
v8	**{(2, 4.0), (4, 4.0), (9, 5.0)}**	**7.55**
v2	**{(1, 3.0), (2,-2.0), (6, 5.0), (8, 6.0)}**	**8.60**
v4	**{(2,-2.0), (4, 4.0), (6, 5.0), (8,-5.0)}**	**8.37**
v10	**{(2,-2.0), (3,-9.0)}**	**9.22**
v1	**{(1,-3.0), (2, 4.0), (5, 3.0), (6, 5.0), (7, 3.0), (8, 6.0)}**	**10.20**
v6	**{(3,-9.0), (4, 4.0), (9, 5.0)}**	**11.05**

Vector length ∈ [6.72, 15.49]

Let us consider vector $u = v1$ and let $\varepsilon = 0.85$. Then, only vectors the lengths of which belong to the interval $\left[\frac{|u|}{\alpha}, \alpha|u|\right] \subseteq [6.72, 15.49]$, where $\alpha = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) + \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right)$, have a chance to belong to ε-neighbourhood of vector u. □

Example: Using bounds on lengths for searching Tanimoto similar vectors for $\varepsilon = 0.98$

Id	(non-zero dimension, value) pairs	length
v5	{(4, 4.0), (5,-3.0), (7, 3.0)}	5.83
v9	{(4,-2.0), (5, 3.0), (7, 3.0), (9, 5.0)}	6.86
v3	{(3, 6.0), (4, 4.0)}	7.21
v7	{(3, 6.0), (4, 4.0)}	7.21
v8	{(2, 4.0), (4, 4.0), (9, 5.0)}	7.55
v2	{(1, 3.0), (2,-2.0), (6, 5.0), (8, 6.0)}	8.60
v4	{(2,-2.0), (4, 4.0), (6, 5.0), (8,-5.0)}	8.37
v10	**{(2,-2.0), (3,-9.0)}**	**9.22**
v1	**{(1,-3.0), (2, 4.0), (5, 3.0), (6, 5.0), (7, 3.0), (8, 6.0)}**	**10.20**
v6	**{(3,-9.0), (4, 4.0), (9, 5.0)}**	**11.05**

Vector length ∈ [8.84, 11.77]

Let us consider vector $u = v1$ and let $\varepsilon = 0.98$. Then, only vectors the lengths of which belong to the interval $\left[\frac{|u|}{\alpha}, \alpha|u|\right] \subseteq [8.84, 11.77]$, where $\alpha = \frac{1}{2}\left(\left(1+\frac{1}{\varepsilon}\right) + \sqrt{\left(1+\frac{1}{\varepsilon}\right)^2 - 4}\right)$, have a chance to belong to ε-neighbourhood of vector u; that is, vectors: $v1$, v6 and $v10$. □

References

- Marzena Kryszkiewicz: Bounds on Lengths of Real Valued Vectors Similar with Regard to the Tanimoto Similarity. ACIIDS (1) 2013: 445-454.
- Marzena Kryszkiewicz: Using Non-Zero Dimensions and Lengths of Vectors for the Tanimoto Similarity Search among Real Valued Vectors. ACIIDS (1) 2014: 173-182