Application of quantum k-NN and Grover's algorithm for recommendation big-data system

Marek Sawerwain, Marek Wróblewski
e-mail: {M.Sawerwain, M.Wroblewski}@issi.uz.zgora.pl,

Institute of Control & Computation Engineering
University of Zielona Góra

QIPLSIGML
Machine Learning meets Quantum Computation
Kraków, Poland, 26.04.2018

Content

Motivations and area of applications

The recommendation system is a collection of prediction mechanisms and analytic algorithms. The main task of the recommendation system is to determine the elements with the greatest degree of matching on the basis of the data set and input information. Real profit and universality of use make them used in areas such as:

- e-commerce
- vod
- price comparison websites
- social networks
- Internet search engines
- expert systems

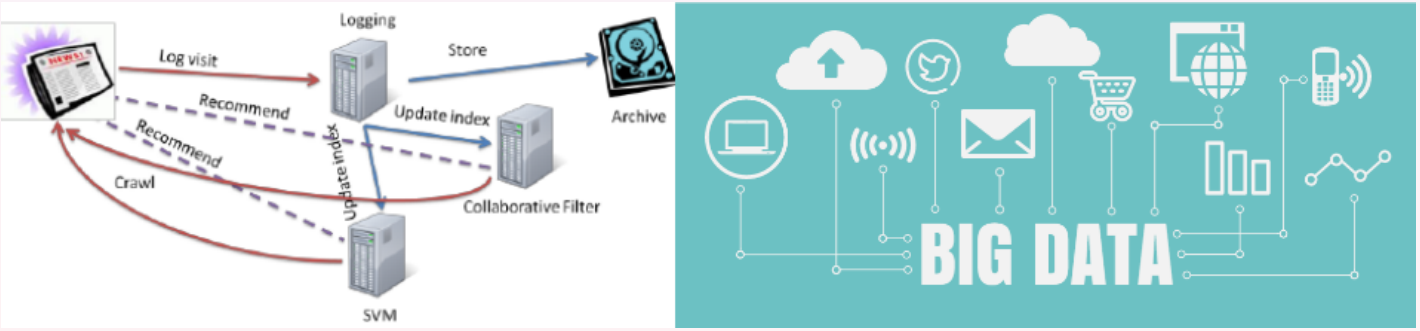


Figure: Source: Fortuna, B.,C., Dunja, C.,M.: Real-Time News Recommender System, https://www.360logica.com

Introduction to quantum computation 1/4

Qubit (qubit state) is the basic unit of quantum information:

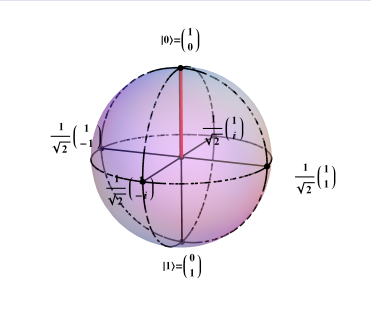
$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \alpha, \beta \in \mathbb{C}, \\ |\alpha|^2 + |\beta|^2 = 1 \tag{1}$$

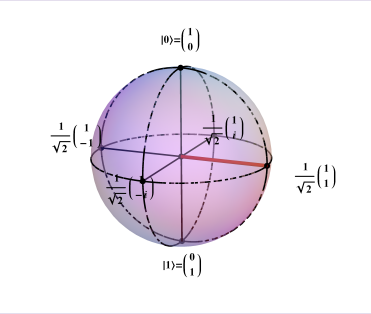
However, qudit is a natural generalization of the qubit concept for d-dimensions

$$|\phi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle + \ldots + \alpha_k|k\rangle, \quad \alpha_i \in \mathbb{C}, \\ \sum_{i=0}^{k} |\alpha_i|^2 = 1 \tag{2}$$

The quantum register is a system composed of a finite number of qubits or qudites:

$$|\Psi\rangle = |\psi_0\rangle \otimes |\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle \tag{3}$$





Remark

Not always quantum register can be represented as a tensor product subsystems, in which case the state of the register is a tangled state

Introduction to quantum computation 2/4

A density matrix of an unknown pure state of the qubit $|\psi\rangle$ is represented by The representation of a density matrix of an unknown pure state of the qubit $|\psi\rangle$ is as follows

$$\rho = |\psi\rangle\langle\psi| = \begin{bmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{bmatrix} \tag{4}$$

where the vector $\langle\psi|$ denotes the transposed vector $|\psi\rangle$.
The representation of a quantum register composed of density matrices is given as

$$\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|, \tag{5}$$

where λ_i represents the probability for state $|\psi_i\rangle$ and $\sum_i \lambda_i = 1$. In quantum mechanics such states are called mixed states.

Exponential increase in quantum data

The size of a quantum register described by the vector state is equal to 2^n or d^n where n is a number of qubits/qudits and d represents the level of a qudit. Size of a density matrix is equal to $d^n \times d^n$.
A vector state of 16 qubits requires 512 kb of memory, while a matrix density needs 256MB, if a single precision arithmetic is used. For 32 qubits a vector state requires 32GB of memory, whereas the density matrix — a 1ZB (zetta bytes $= 10^{21}$) !!!

Introduction to quantum computation 3/4

The realisation of application of a unitary operation U for a state expressed as a vector state is governed by a very simple equation:

$$U|\psi_0\rangle = |\psi_1\rangle. \tag{6}$$

In case of a density matrices, application of a unitary matrix U is described as follows:

$$U\rho_0 U^\dagger = \rho_1. \tag{7}$$

The technique of making a suitable unitary operator for a given quantum register is a formidable task from the computational point of view. If we want to modify e.g. the first and the third qudit then a suitable unitary operator is constructed as follows:

$$U = u_1 \otimes I \otimes u_2 \tag{8}$$

Note about controlled gates

The construction of controlled gates is also difficult, as those require the use of not only tensor products but also additional matrix projectors. For example, one of the possible realisations of a qubit *controlled not gate* is given by

$$U = |000\rangle\langle 000|I + |010\rangle\langle 010|X + |111\rangle\langle 111|I \tag{9}$$

Hadamard's gate, marked as H, is a valid goal due to the fact of entering the so-called superposition of quantum states. Let $|x\rangle$ be n-qubits:

$$|x\rangle = |x_0\rangle \otimes |x_1\rangle \otimes \cdots \otimes |x_{n-2}\rangle \otimes |x_{n-1}\rangle. \tag{10}$$

Hadamard gate impact on the state $|x\rangle$ is described as follows:

$$\begin{aligned} H|x\rangle &= H(|x_0\rangle \otimes |x_1\rangle \otimes \cdots \otimes |x_{n-2}\rangle \otimes |x_{n-1}\rangle) = \\ &= H|x_0\rangle \otimes H|x_1\rangle \otimes \cdots \otimes H|x_{n-2}\rangle \otimes H|x_{n-1}\rangle = \\ &= \bigotimes_{i=0}^{n-1} H|x_i\rangle = \frac{1}{\sqrt{2^n}} \left[\bigotimes_{i=0}^{n-1} \left(|0\rangle + (-1)^{x_i} |1\rangle \right) \right]. \end{aligned} \tag{11}$$

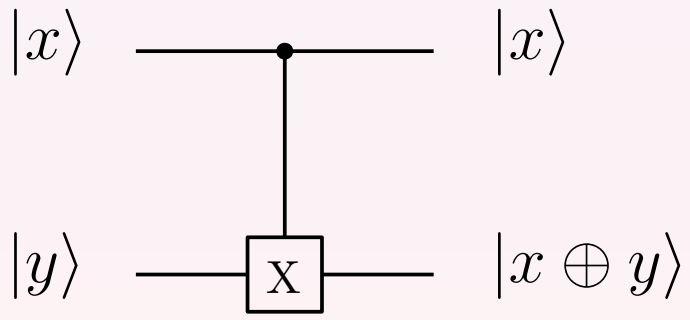
As you can see the application of the Hadamard gate, it gives an amplitude a value equal to the absolute value $\frac{1}{\sqrt{2^n}}$ in all states that a given quantum register can accept.

Gate CNOT (XOR)

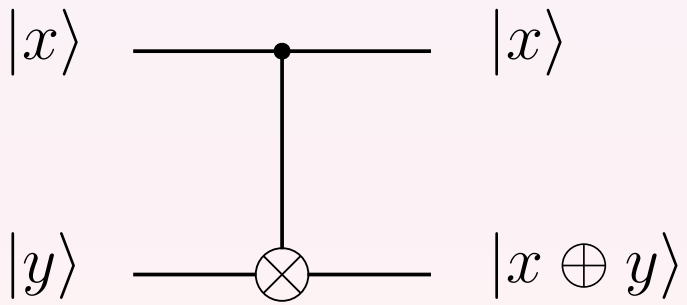
Action CNOT gate and its matrix representation:

$$\begin{aligned}
\mathrm{CNOT}|00\rangle &= |00\rangle, \\
\mathrm{CNOT}|01\rangle &= |01\rangle, \\
\mathrm{CNOT}|10\rangle &= |11\rangle, \\
\mathrm{CNOT}|11\rangle &= |10\rangle,
\end{aligned}
\qquad
\mathrm{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Graphical representation CNOT gate, as is used for quantum circuits:



(a)



(b)

Introduction to quantum computation 4/4

The realisation of the standard (von Neumann) quantum measurement begins with the preparation of an observable:

$$M = \sum_i \lambda_i P_i \tag{12}$$

where P_i is a projector associated with an eigenvalue λ_i of operator M. The results of measurement (λ_i) are known only to a probability $P(\lambda)$, given by: The results of the measurement are represented by the eigenvalues $lambda_i$. The measurement of the state of the register as opposed to the classical computer science is a probabilistic operation and the individual results occur with the probability determined in the following way:

$$P(\lambda_i) = \langle\psi|P_i|\psi\rangle \tag{13}$$

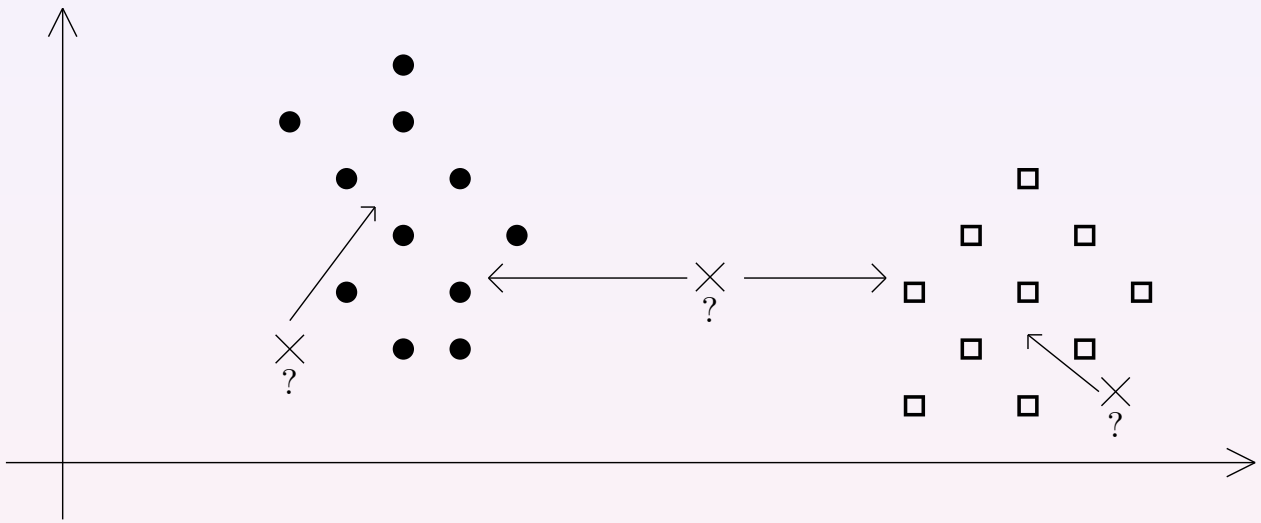
The obtained result λ_i means that the register $|\psi\rangle$ is collapsed to the state:

$$\frac{P_i|\psi\rangle}{\sqrt{\lambda_i}} \tag{14}$$

From the computational point of view two operations must be performed: calculation of the probability of distribution and the transformation of quantum register is performed.

The k-algorithm - the nearest neighbors

Classification of the tested pattern based on the distance to k nearest neighbors.



Notes on performance and the KNN:

- we have n of d patterns, and access to the training set,
- the rating of one sample is $O(d)$,
- distance rating for all samples $O(nd)$
- an additional $O(nk)$ to find k nearest neighbors,
- the total time is: $O(nk + nd)$.

Despite the polynomial class and currently available technology, alg. kNN is still "expensive" for large data sets.

The Hamming distance

The Hamming distance is a metric in the vector space of N words. It meets three conditions:

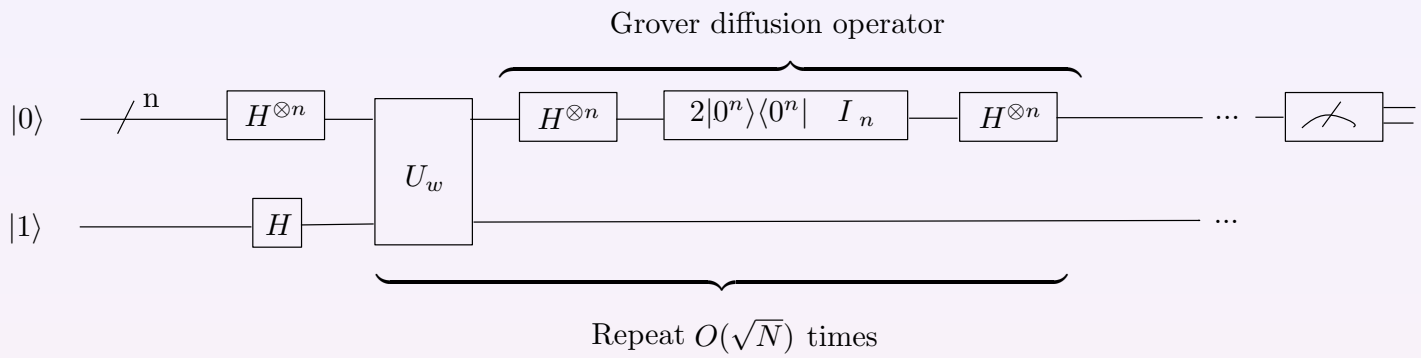
- is non-negative,
- symmetrical,
- satisfies the inequality of the triangle.

For binary strings a and b the Hamming distance is equal to the number of "ones" in the word $a \,\mathrm{XOR}\, b$, where XOR is the exclusive logic function (exclusive-or).
Examples of the Hamming distance (the operator $\bowtie$):

$$00101 \bowtie 00101 = 0, \quad 00101 \bowtie 00111 = 1, \quad 00101 \bowtie 10111 = 2 \qquad (15)$$

where $\bowtie$ means the comparison operation in the sense of Hamming distance.

Grover's algorithm

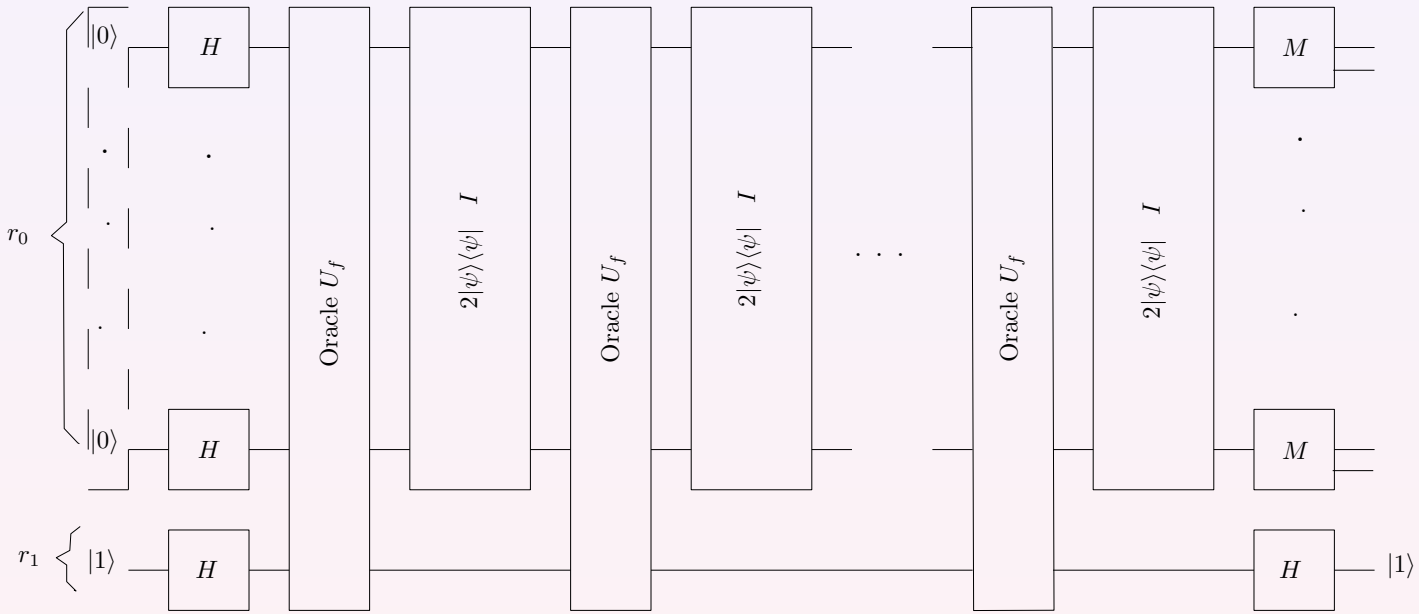


The most important steps in the Grover algorithm:

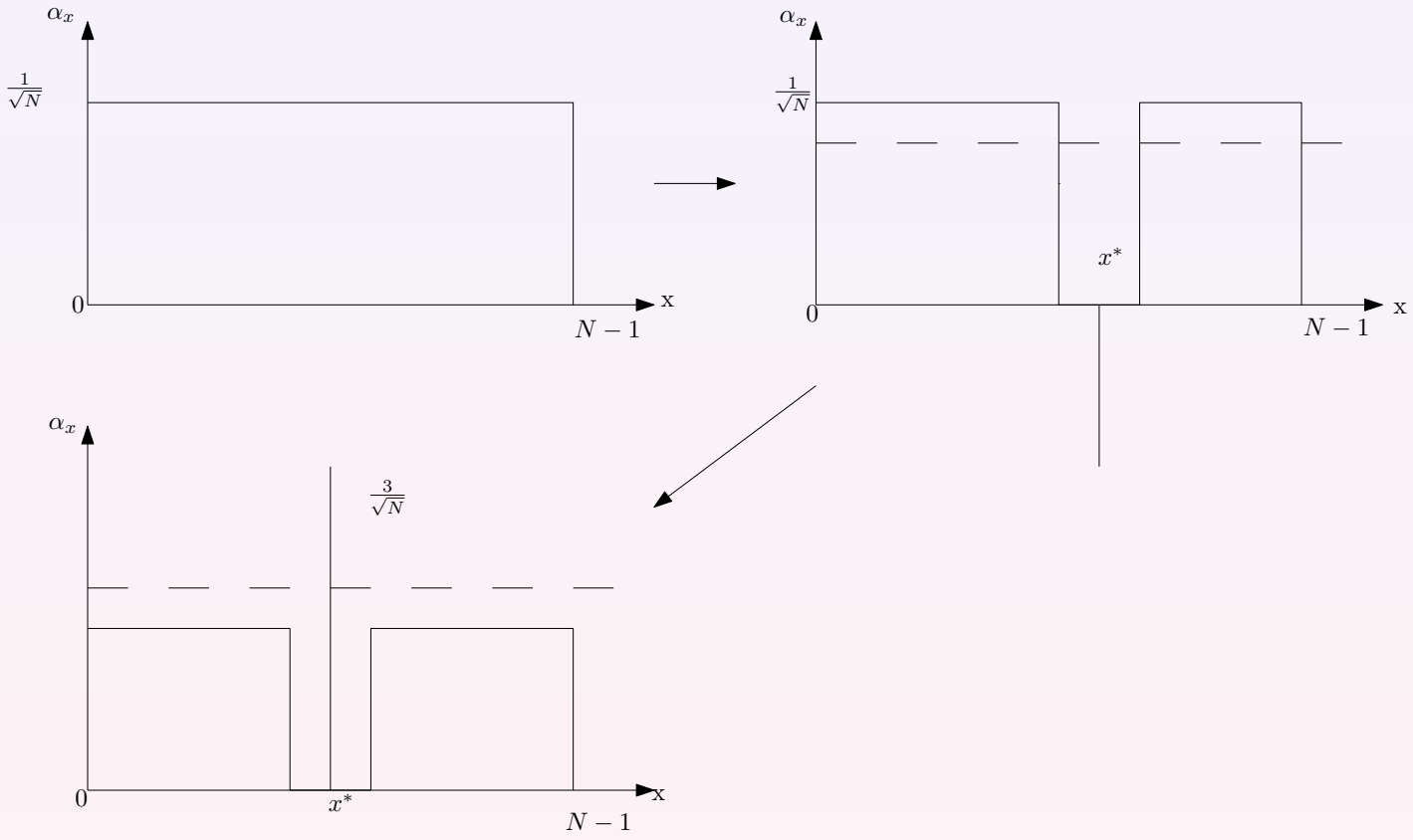
- initial state: $|\psi\rangle = |0\rangle^{\otimes n}|1\rangle$,
- initialization of the database registry: $|\psi\rangle \rightarrow H^{\otimes(n+1)}\frac{1}{\sqrt{2^n}}\sum_{t=0}^{2^n-1}|t\rangle|-\rangle = (\cos\frac{\theta}{2}|\alpha\rangle + \sin\frac{\theta}{2}|\beta\rangle)|-\rangle$,
- application of the amplify operator probability amplitudes: $G^R[\cos(\frac{2R+1}{2}\theta)|\alpha\rangle + \sin(\frac{2R+1}{2}\theta)|\beta\rangle]|-\rangle$,
- measurement of the first n computational qubits database type: $|x\rangle|-\rangle$,

where $|\alpha\rangle$ is a sub-register denoting weakened states, while $|\beta\rangle$ states for which we want to build up probability amplitudes.

Grover's algorithm



Grover's algorithm



Given $f : 0, ..., N-1 -> 0, 1$ such that $f(x) = 1$ for exactly one x, find x.

The classic version of the database

As part of the experiment, a classic database was designed and made based on real data downloaded from the OMDb website. They contain information about records describing individual films. The "movies" table contains the following columns:

- movie id
- year of publication
- title
- duration
- genre
- language
- imdb rating
- imdb votes
- type
- hamming

The classic version of the database

The database contains 12051 records. The following types of films exist in the database:

• News • Music • Game-Show • Talk-Show • Reality-TV • Sport • Animation • Family • Documentary	• Musical • Fantasy • Comedy • Biography • History • Romance • Adventure • Mystery • Western	• Action • Sci-Fi • Drama • Thriller • Crime • War • Horror • Adult

The classic version of the database

Id	Year	Title	Time	Genres	Language	IMDb rating	IMDb votes	Type	Hamming
266	1993	The Legend	100 min	Action, Comedy, History	Cantonese	7.3	5,475	movie	101100011100100000
334	1994	Princess Caraboo	97 min	Comedy, Drama, History	English	5.9	2,073	movie	111001100001000000
351	1998	Frank Lloyd Wright	146 min	Documentary, Biography, History	English	7.8	470	movie	100010011110100000
371	2001	The Inner Tour	97 min	Documentary, History	Arabic	7.5	105	movie	100010100000000000
426	1988	Young Einstein	91 min	Comedy, History	English	5.0	7,034	movie	111001000000000000
500	1987	Matewan	135 min	Drama, History	English, Italian	7.9	5,739	movie	110000100000000000
545	2004	Japan: Memoirs of a Secret Empire	123 min	Documentary, History	English	7.6	110	series	100010100000000000
551	1998	America's Journey Through Slavery	360 min	History	English	7.4	55	series	100000000000000000
850	1981	Quest for Fire	100 min	Adventure, Drama, History	French	7.4	15,079	movie	100110110000100000
857	1935	Mutiny on the Bounty	132 min	Adventure, Drama, History	English, Polynesian	7.8	14,972	movie	100110110000100000
898	1978	Gray Lady Down	111 min	Adventure, Drama, History	English, Norwegian	6.2	1,87	movie	100110110000100000
994	1961	King of Kings	168 min	Biography, Drama, History	English	7.1	4,3	movie	111101100001000000
1048	2004	Ben Franklin	210 min	Documentary, History	English	7.4	30	movie	100010100000000000
1061	1921	Orphans of the Storm	150 min	Drama, History	English	7.9	3,565	movie	110000100000000000
1070	2000	Uncle Saddam	63 min	Documentary, Comedy, History	English	6.4	221	movie	100010011100100000
1086	1997	Waco: The Rules of Engagement	165 min	Documentary, History	English	8.0	2,514	movie	100010100000000000
1147	1996	The Crucible	124 min	Drama, History	English	6.8	26,826	movie	110000100000000000

Figure: Structure of the "movies" table

Diagram of the division of quantum registers

$|Q\rangle =$ Database (2752) (12051) (100%)

$|q_0\rangle =$ Musical (22) (64) (0,531%)

$|q_1\rangle =$ Fantasy (3) (64) (0,531%)

$|q_2\rangle =$ Comedy (574) (2876) (23,865%)

$|q_3\rangle =$ Biography (3) (386) (3,203%)

$|q_4\rangle =$ History (7) (12) (0,100%)

$|q_5\rangle =$ Documentary (418) (1023) (8,489%)

$|q_6\rangle =$ Romance (16) (78) (0,647%)

$\vdots$

$|q_{25}\rangle =$ Adventure (17) (546) (4,531%)

News	(0) (0) (0,000%)
Music	(130) (207) (1,718%)
Game-Show	(0) (3) (0,025%)
Talk-Show	(2) (2) (0,017%)
Reality-TV	(7) (10) (0,083%)
Sport	(13) (13) (0,108%)
Animation	(32) (584) (4,846%)
Family	(36) (124) (1,029%)
Mystery	(8) (69) (0,573%)
Western	(72) (78) (0,647%)
Action	(69) (1743) (14,464%)
Sci-Fi	(26) (75) (0,622%)
Drama	(948) (2642) (21,923%)
Thriller	(102) (136) (1,129%)
Crime	(9) (701) (5,817%)
War	(2) (16) (0,133%)
Horror	(216) (575) (4,771%)
Adult	(20) (24) (0,199%)

The structure of the quantum register

The structure of the quantum register in the recommendation system:

$$|\psi\rangle = \underbrace{\overset{|\psi_q\rangle}{\boxed{\text{IDs of entity}}}}_{\mathrm{q}}\underbrace{\overset{|\psi_{f_f}\rangle}{\boxed{\text{feature vectors}}}}_{\mathrm{f}}\underbrace{\overset{|\psi_{f_u}\rangle}{\boxed{\text{user vector}}}}_{\mathrm{f}}\underbrace{\overset{|\psi_k\rangle}{\boxed{\text{additional suplemental qubits}}}}_{\mathrm{k}}$$

The total number of qubits:

$$\mathsf{N} = \mathsf{q} + \mathsf{f} + \mathsf{f} + \mathsf{k}\ .$$

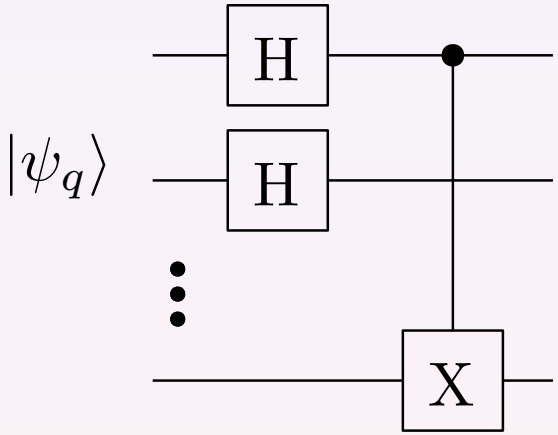
The vector features (feature vector) and user preferences vector are the same length. Examples of states:

$$\begin{array}{rcl} \dots & \dots & \dots \\ |\psi_{i-1}\rangle & = & |010|111001|101001|\psi_k\rangle, \\ |\psi_i\rangle & = & |011|111011|101001|\psi_k\rangle, \\ |\psi_{i+1}\rangle & = & |100|111100|101001|\psi_k\rangle, \\ \dots & \dots & \dots \end{array}$$

Coding identifiers and features

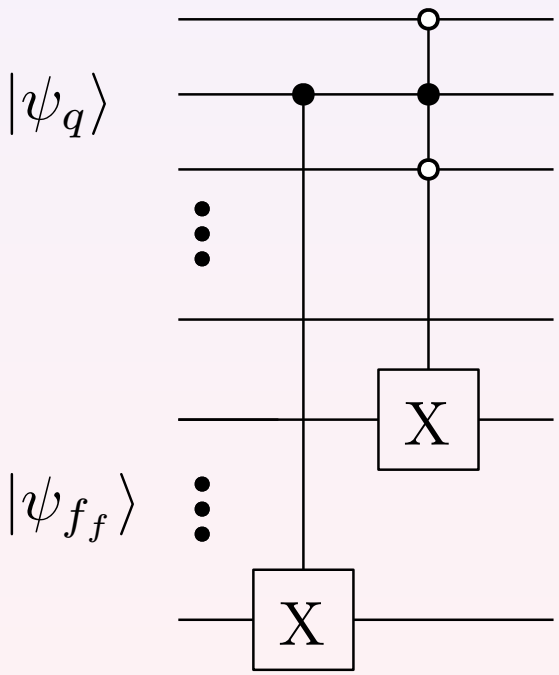
Examples of circuits for creating identifiers and features using multi qubit gates of controlled negation:

Initialization of IDs



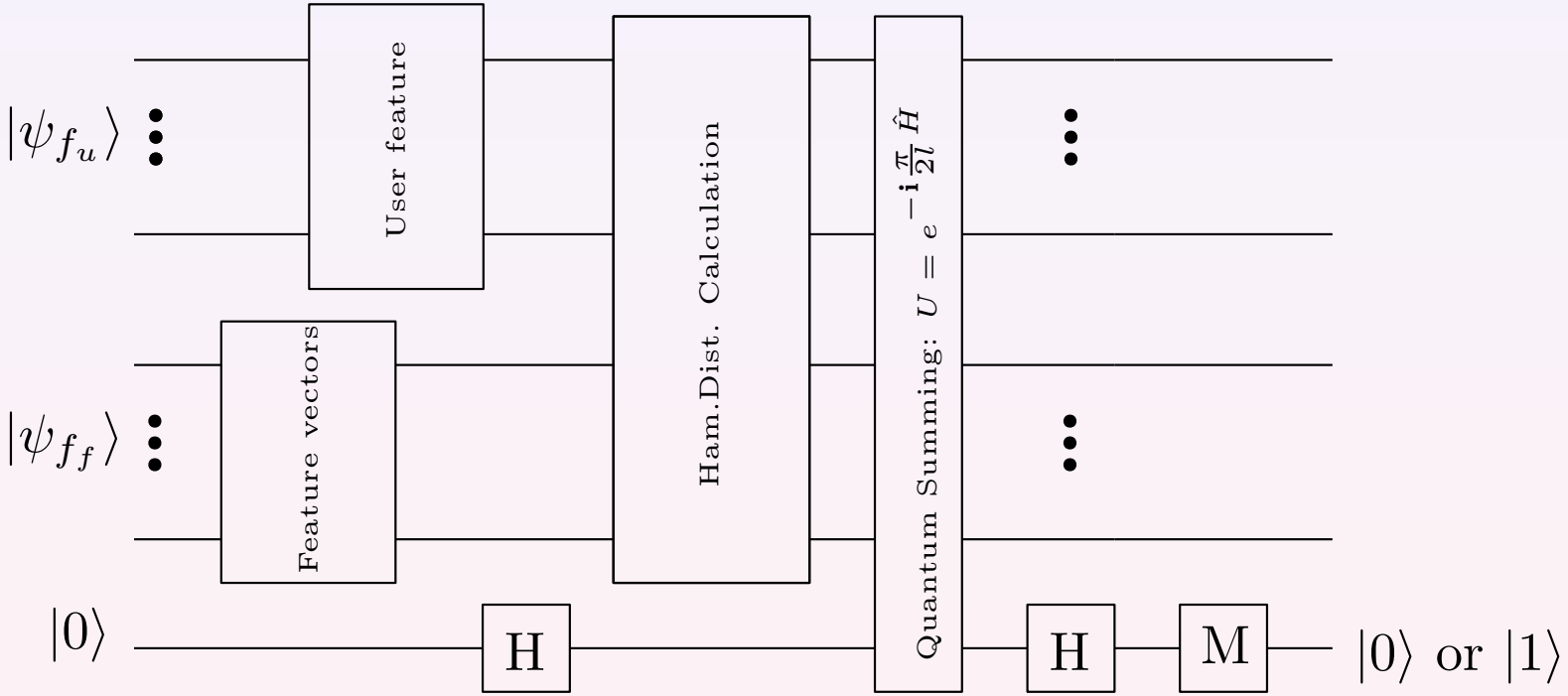
(a)

Features encoding

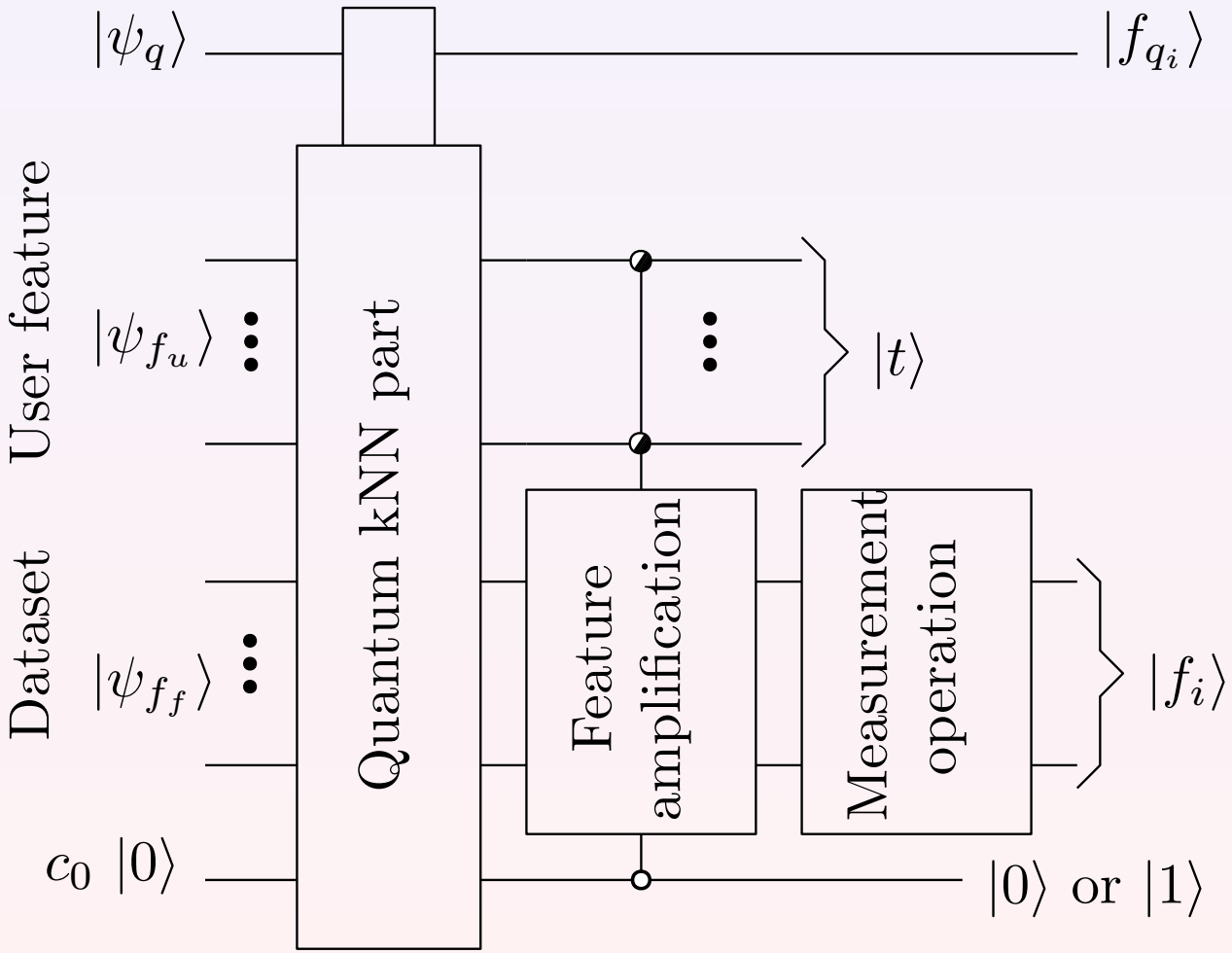


(b)

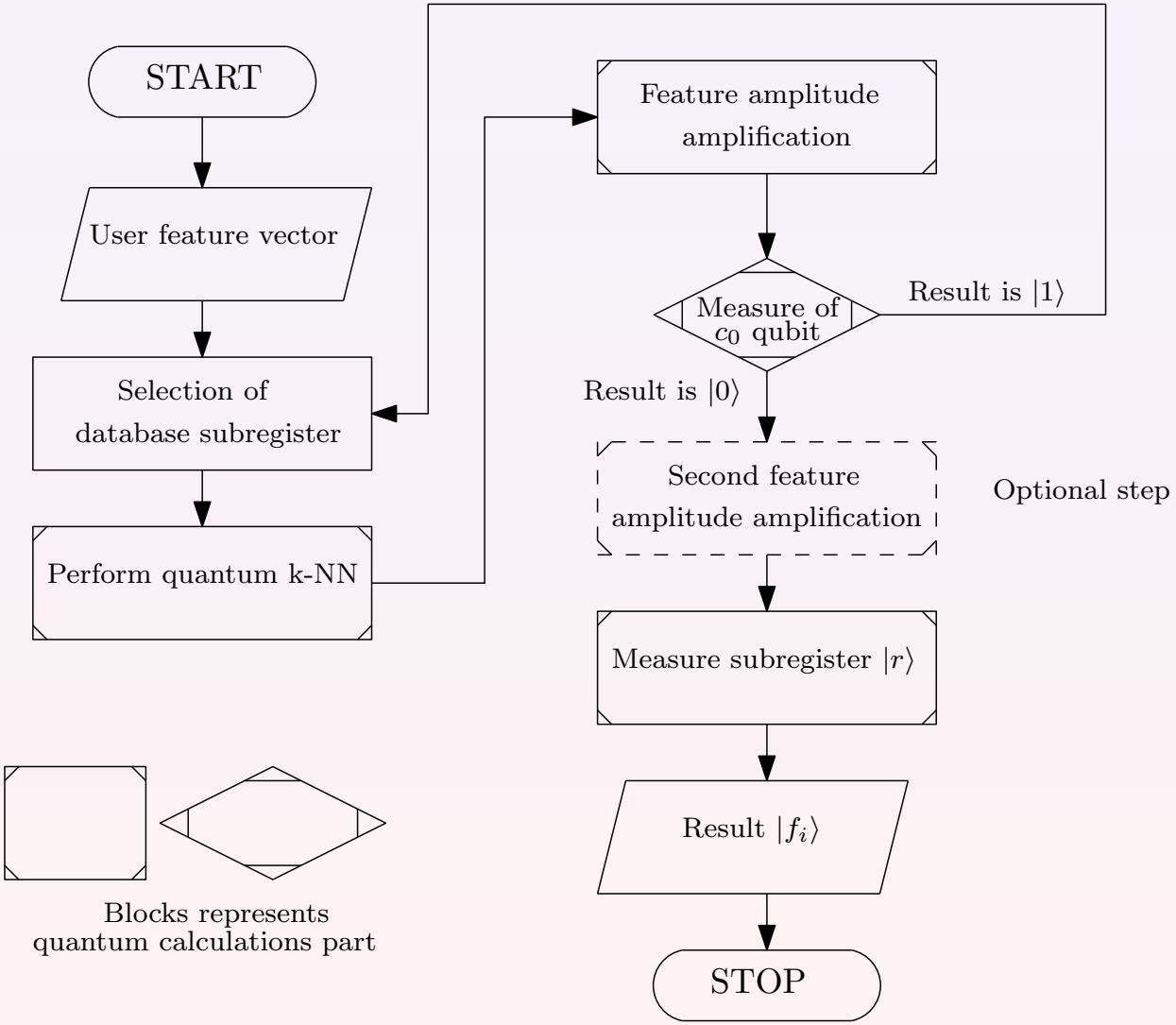
Diagram of the quantum circuit for quantum kNN



General diagram of the quantum circuit recommendation system



The control flow proposed for quantum recomendation algorithm



Realization of the qkNN part

The initial state of the quantum register during the implementation of qkNN (omitted for simplicity will be the identifier describing the film qubits):

$$|\psi_0\rangle = |0\rangle^{\otimes 2l+1}. \tag{16}$$

The first l qubits represent the database file, the second l qubitów is a user feature vector. Initialization of the sub-register with the base of features:

$$|\Psi_1\rangle = \frac{1}{\sqrt{L}} \sum_{p=1}^{L} |r_1^p, \ldots, r_l^p\rangle, \tag{17}$$

where $L = 2^l$. Specifying the form of the "database" and the user's vector results in the following state:

$$|\Psi_2\rangle = \frac{1}{\sqrt{L}} \left(\sum_{p=1}^{L} |r_1^p, \ldots, r_l^p\rangle \right) \otimes |t_1, \ldots, t_l\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{18}$$

Calculation of the Hamming distance between the set of features and the user's vector is carried out using the CNOT gates:

$$CNOT(r_i^p, t_i,) = (d_i^p, t_i), \quad i = 1, \ldots, I. \tag{19}$$

After performing this operation, the registry status is as follows:

$$|\Psi_3\rangle = \frac{1}{\sqrt{L}} \left(\sum_{p=1}^{L} |d_1^p, \ldots, d_I^p\rangle \right) \otimes |t_1, \ldots, t_I\rangle \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \tag{20}$$

The next step is to perform the operation to determine the so-called the sum of Hamming distances. This operation is performed by the block `Hamm.Summ`. Described by the unit operation U:

$$U = e^{-\mathbf{i}\frac{\pi}{2I}\hat{H}}, \quad \hat{H} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^{\otimes I} \otimes I_{I\times I} \otimes \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{21}$$

After applying the U operation, we get the state:

$$|\psi_4\rangle = \frac{1}{\sqrt{L}} \sum_{p=1}^{L} \left(e^{\mathbf{i}\frac{\pi}{2l} d(t,r^p)} |d_1^p, \ldots, d_l^p\rangle \otimes |t_1, \ldots, t_l\rangle \otimes |0\rangle + \right. \\ \left. + e^{-\mathbf{i}\frac{\pi}{2l} d(t,r^p)} |d_1^p, \ldots, d_l 1^p\rangle \otimes |t_1, \ldots, t_l\rangle \otimes |1\rangle \right). \tag{22}$$

After the Hadamard operation on the additional qubits c_0, we get a state describing the state after the quantum algae has been performed. k-NN:

$$|\psi_5\rangle = \frac{1}{\sqrt{L}} \sum_{p=1}^{L} \left(\cos\left(\frac{\pi}{2l} d(t, r^p)\right) |d_1^p, \ldots, d_l^p\rangle \otimes |t_1, \ldots, t_l\rangle \otimes |0\rangle + \right. \\ \left. + \sin\left(\frac{\pi}{2l} d(t, r^p)\right) |d_1^p, \ldots, d_l^p\rangle \otimes |t_1, \ldots, t_l\rangle \otimes |1\rangle \right). \tag{23}$$

The probability of measuring zero on state c_0, i.e. success, that we will go through the desired probability distribution with the correct indication of the recommended element:

$$P(c_0) = \frac{1}{L} \sum_p \cos^2\left(\frac{\pi}{2l} d(t, r^p)\right). \tag{24}$$

Strengthening the success of recommendations

Assuming that zero was obtained by measuring the state c_0, we get another state described in the following way:

$$|\psi_6\rangle = \frac{1}{\sqrt{L}} \sum_{p=1}^{L} m_p |\psi_{rcmd}\rangle, \tag{25}$$

where $|\psi_{rcmd}\rangle = |d_1^p, \ldots, d_l^p\rangle \otimes |t_1, \ldots, t_l\rangle \otimes |0\rangle$, while m_p represent the amplitudes of probability obtained after the measurement. In general, the register can be distinguished by g amplitudes for recommended movies with the highest compatibility of the feature m_p^r and $L-g$ with lower compatibility m_p^{nr}:

$$|\psi_7\rangle = \frac{1}{\sqrt{L}} \left(\sum_{p=1}^{g} m_p^r |\psi_{rcmd}\rangle + \sum_{p=g+1}^{L} m_p^{nr} |\psi_{rcmd}\rangle \right). \tag{26}$$

Based on the paper [Biham1996, Phys. Rev. A, 60, 2742] we introduce the mean and the variance for amplitudes:

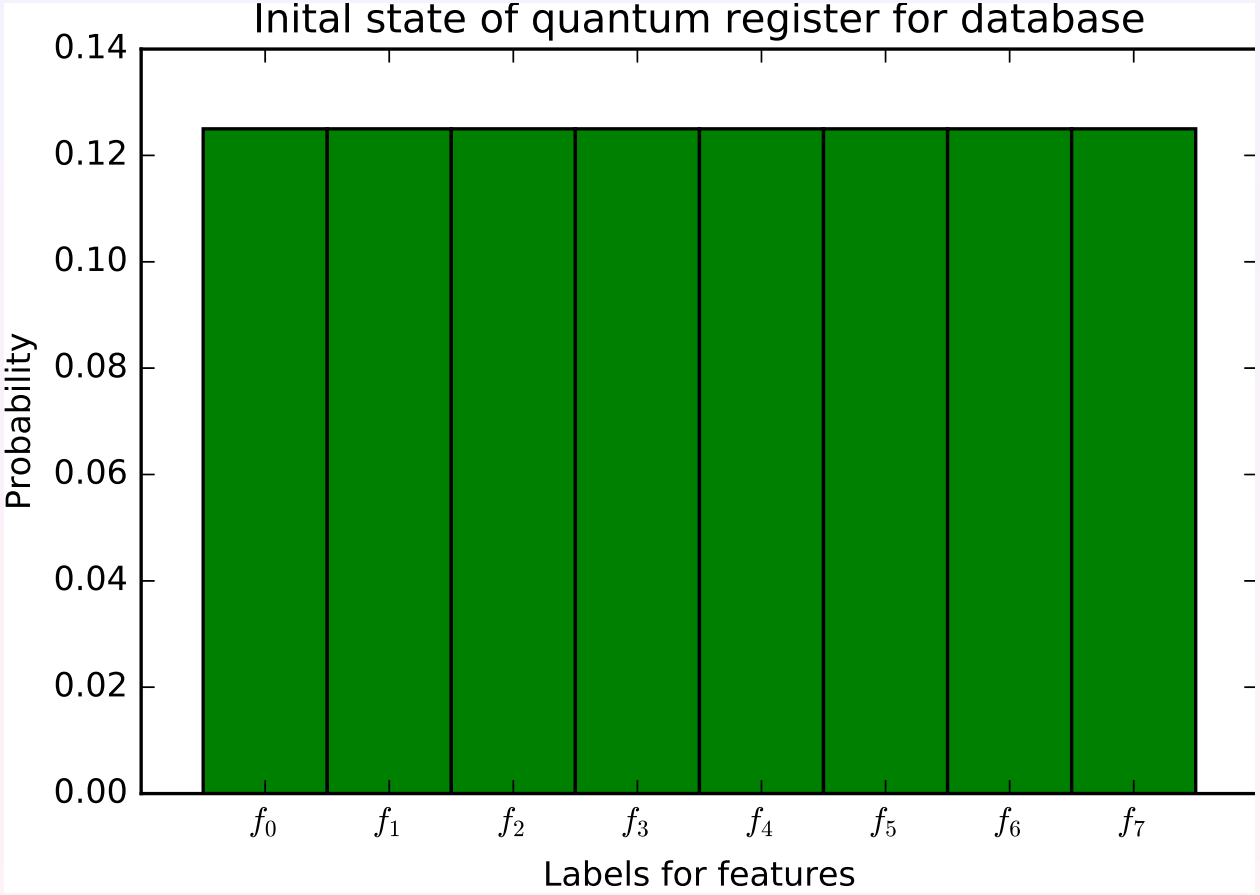
$$\overline{m^r}(t) = \frac{1}{q}\sum_{p=1}^{q} m_p^r(t), \quad \overline{m^{nr}}(t) = \frac{1}{q}\sum_{p=q+1}^{L} m_p^{nr}(t),$$
$$\sigma_{nr}^2(t) = \frac{1}{L-q}\sum_{p=q+1}^{L} |m_p^{nr}(t) - \overline{m^{nr}}(t)|^2 \tag{27}$$

where t is the iteration number (time) of the algae operation. Grover, for $t = 0$ we have an initial distribution of probability amplitudes. The maximum probability of measuring recommended elements from the database is defined as:

$$P_{max} = 1 - (L-q)\sigma_{nr}^2 - \frac{1}{2}\left((L-q)|\overline{m^{nr}}(0)|^2 + q|\overline{m^r}(0)|^2\right) + \\ + \left(\frac{1}{2}|(L-q)\overline{m^{nr}}(0)^2 + q\overline{m^r}(0)^2|\right), \tag{28}$$

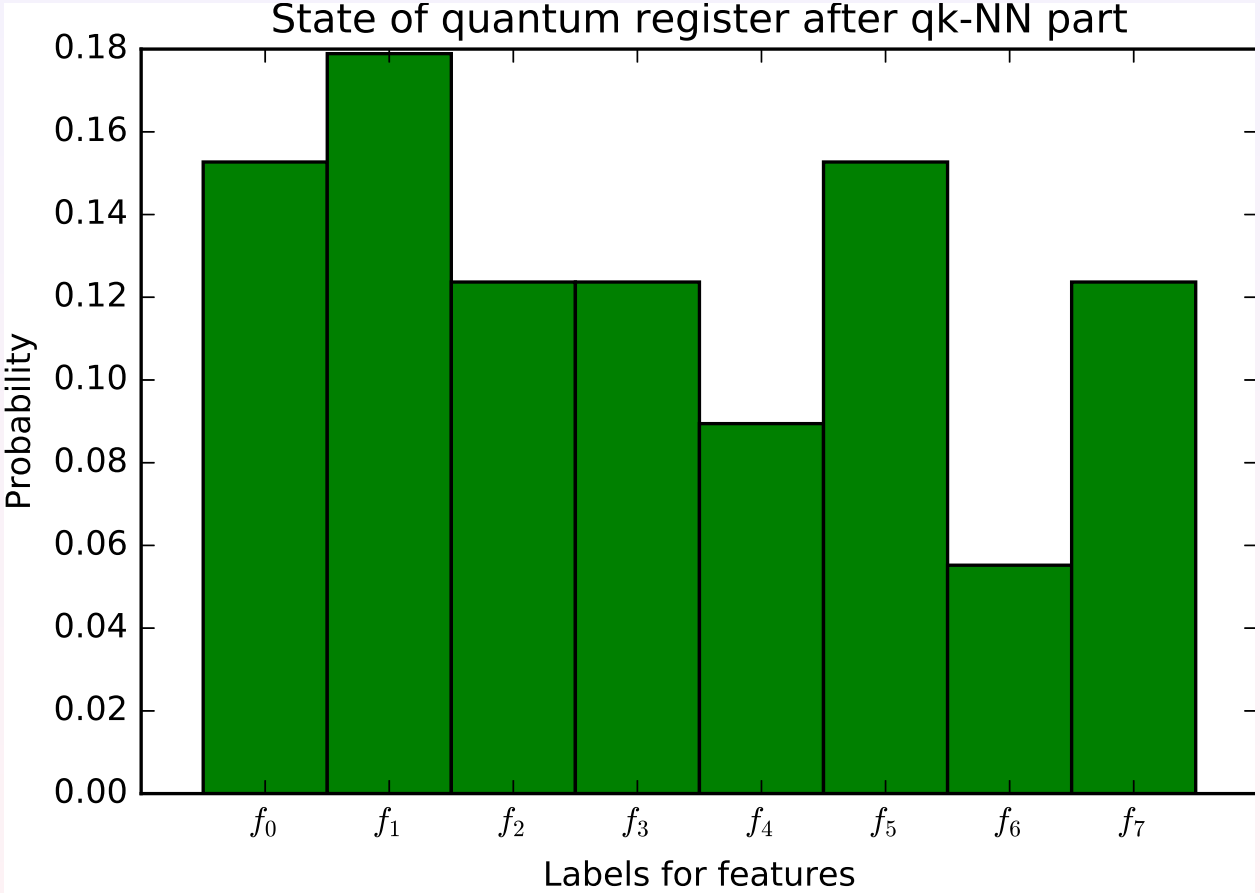
when executing $O(\sqrt{\frac{L}{q}})$ iteration.

The initial state of the registry with the database



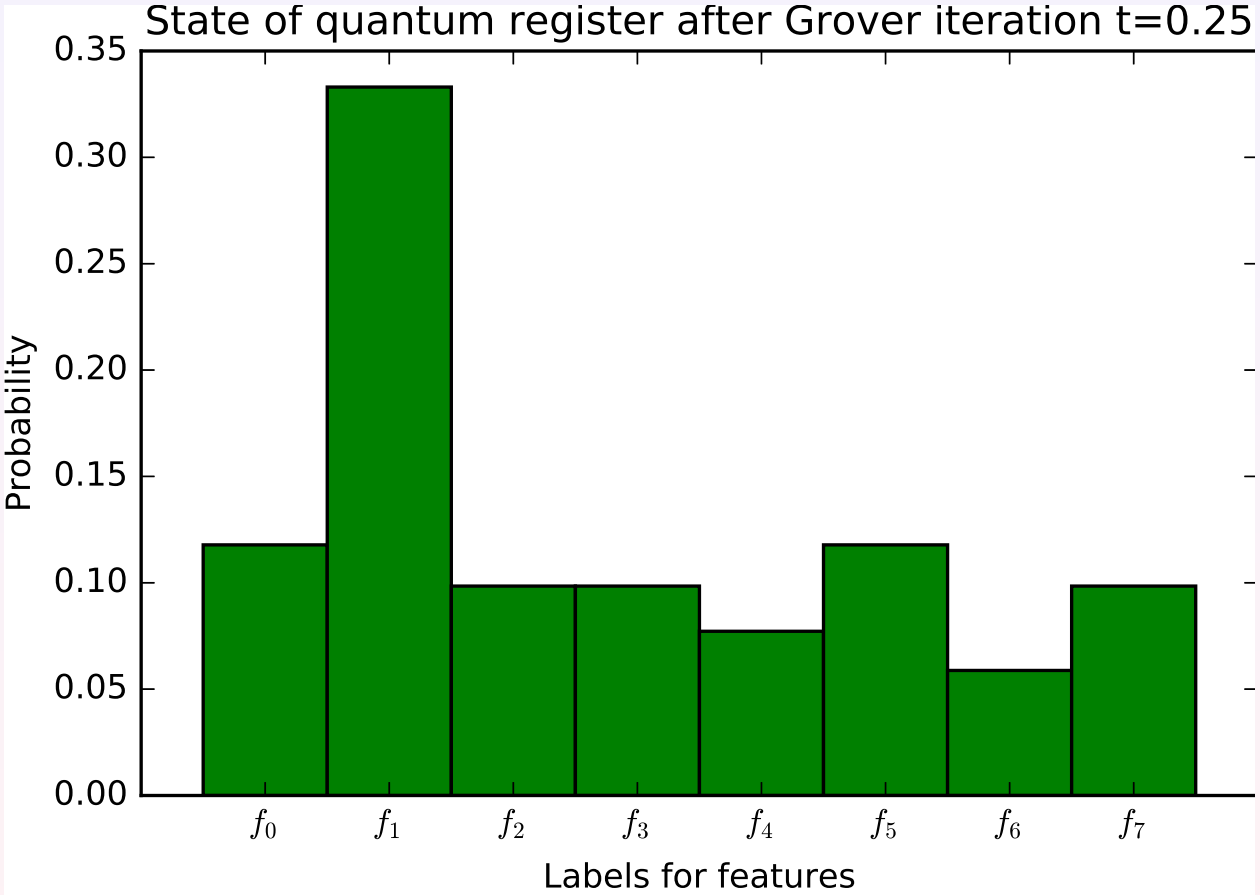
The initial state of the selected database with elements for the recommendation. We assume that we are interested in the feature f_1.

Initial state of the register after qk-NN



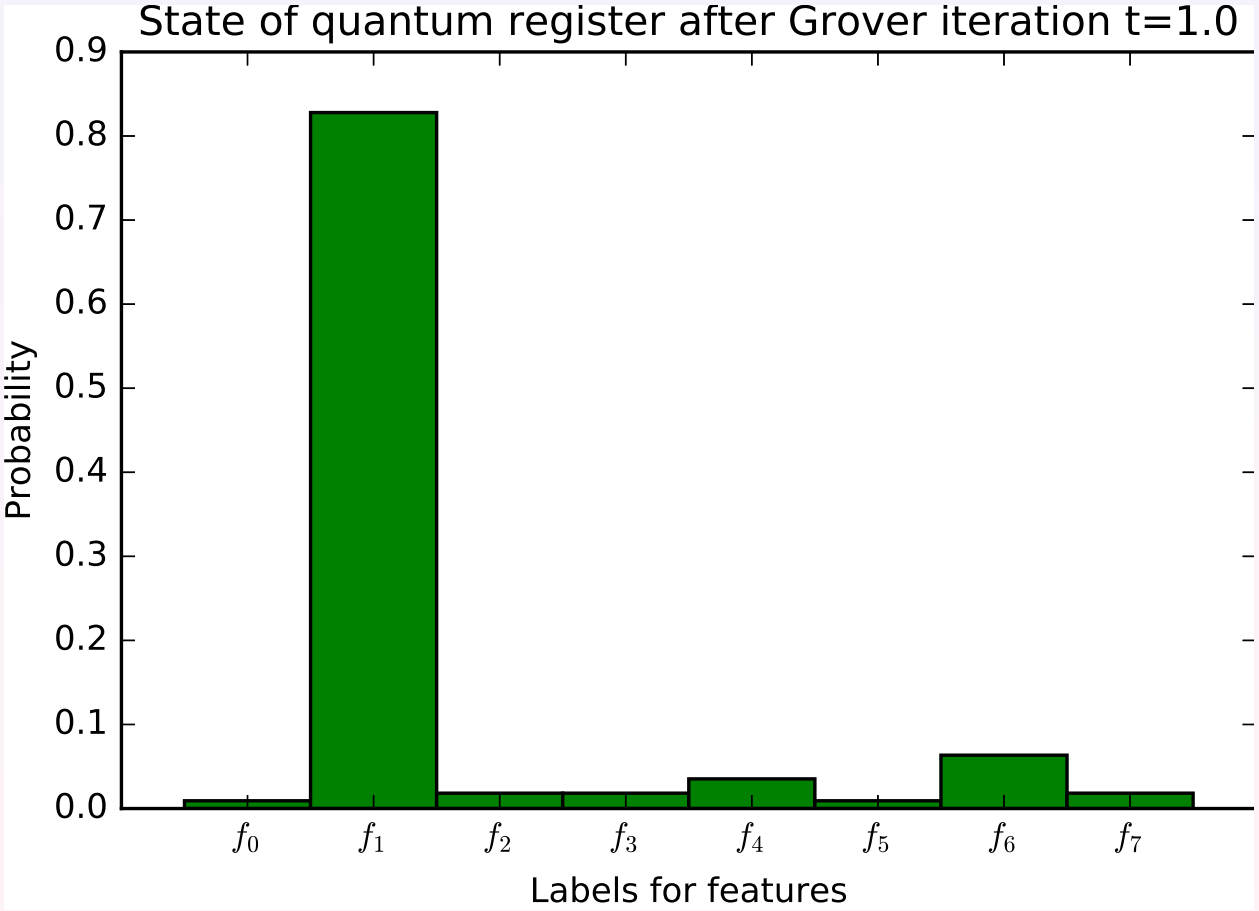
The status of the selected database after the implementation of the part qk-NN, the feature f_1 has the highest probability of measurement.

Amplified Grover for t = 0.25



The status of the selected database after the implementation of the Grover gain part for $t = 0.25$, the f_1 feature again has the highest probability of measurement.

Amplified Grover for t = 1.00



The status of the selected database after the implementation of the Grover gain part for $t = 1.00$, the feature f_1 again has the highest probability of measurement.

Summary

Conclusions and open problems:

- presented a one of the ways of practical use of quantum algorithms,
- made a Algebraic analysis of the problem,
- made a numerical experiments,
- in the future practical - verification of the solution on an IBM quantum computer.

Thank you for the attention!

Bibliography

- Schuld, S., Sinayskiy, I., Petruccione, F.: Quantum Computing for Pattern Classification. PRICAI 2014: Trends in Artificial Intelligence,Springer International Publishing, p. 208-220 (2014)
- Mateus, P., Omar, Y.: Quantum Pattern Matching. arXiv preprint: quant-ph/0508237v1 (2005)
- Pinkse P.W.H., Goorden S.A., Horstmann M., Skoric B., Mosk A.P.:Quantum pattern recognition. Lasers and Electro-Optics Europe (CLEO EUROPE/IQEC), 2013 Conference on and International Quantum Electronics Conference, Munich, pp. 1-1. (2013)
- Schaller G., Schützhold R.: Quantum algorithm for optical-template recognition with noise filtering. Phys. Rev. A 74, pp. 012303 (2006)
- Schuld, M., Sinayskiy, I., Petruccione, F.: Quantum computing for pattern classification. In: 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI) and Also Appear in the Springer Lecture Notes in Computer Science 8862 (2014)
- Trugenberger, C.A.: Quantum Pattern Recognition. Quantum Information Processing 1(6), 471–493, Springer (2002)
- Yue Ruan, Hanwu Chen, Jianing Tan, Xi Li: Quantum computation for large-scale image classification. Quantum Information Processing 15(10), 4049–4069, Springer (2016)

Bibliography

- Wiebe N., Kapoor A., Svore K.M.: Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning, Quantum Information & Computation, Vol. 15 Issue 3-4, pp. 316 – 356 (2015)
- Erdal, A.: An Information-Theoretic Analysis of Grover's Algorithm. Quantum Communication and Information Technologies, Springer Netherlands, 339–347 (2003)
- Alpaydin, E.: Introduction to machine learning. MIT press (2004)
- Armbrust, M., Fox, A., Griffith, R., Joseph, D.A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A view of cloud computing. Commun. ACM. 4, 50-58 (2010)
- Brassard, G., H.P.: An exact quantum polynomial-time algorithm for simon's problem. IEEE Computer Society Press pp. 12-23 (1997)
- Busemeyer, J., B.P.: Quantum models of cognition and decision. Cambridge University Press (2012)
- Hechenbichler, K., S.K.: Weighted k-nearest-neighbor techniques and ordinal classification. Sonderforschungsbereich p. 399 (2004)
- Fortuna, B.,C., Dunja, C.,M.: Real-Time News Recommender System. Machine learning and knowledge discovery in databases. European conference, ECML PKDD 2010, Barcelona, Spain, September 20–24, 2010. Proceedings, Part III (pp.583-586)
- Biham, E.,O., Biron, D., Grassl, M., Lidar, D.: Grover's quantum search algorithm for an arbitrary initial amplitude distribution. Phys. Rev. A 60, 2742 (1999)