

SELECTING EXPERTS USING DATA QUALITY CONCEPTS

Carlos Eduardo Barbosa¹, Marcio Antelio^{1,2}, Jano Moreira de Souza¹, Jonice Oliveira³

¹COPPE, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

²CEFET, Rio de Janeiro, Brazil

³DCC, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

ABSTRACT

Personal networks are not always diverse or large enough to reach those with the right information. This problem increases when assembling a group of experts from around the world, something which is a challenge in Future-oriented Technology Analysis (FTA). In this work, we address the formation of a panel of experts, specifically how to select a group of experts from a huge group of people. We propose an approach which uses data quality dimensions to improve expert selection quality and provide quality metrics to the forecaster. We performed a case study and successfully showed that it is possible to use data quality methods to support the expert search process.

KEYWORDS

Learning Curves, Knowledge Depreciation, Data Quality, Expert Selection, Decision-making, Future-oriented Technology Analysis.

1. INTRODUCTION

Nowadays, budget constraints make it necessary to choose which scientific research works receive priority. Therefore, it has become necessary to develop ways to identify long-term technologies that will have the greatest impact on the future of science, the economy, and society. Several approaches for predicting the future of technology have been united under the term future-oriented technology analysis (FTA).

There are several FTA methods which depend on experts' opinions. Porter et al. [1] listed Delphi, Focus Groups, Interviews, and Participatory Techniques. These methods use a large group of experts, relying on the 'wisdom of crowds' [2].

In the other hand, Data quality techniques may be used to measure the expert search quality of the results. Uotila and Melkas [3] discussed the similarities and the differences between data quality, information, and knowledge. In the scope of this work, the terms 'data' and 'information' are interchangeable.

The purpose of this work is to provide a framework based on learning curves and knowledge depreciation to form a panel of experts, and to measure its intrinsic quality using data quality dimensions. To illustrate the framework, a small Delphi modification was performed.

In Section 2, a literature review is done, which covers FTA, learning curves, knowledge depreciation, and data quality concepts; Section 3 presents the expert selection framework; Section 4 presents a case study which shows the framework in practice; and Section 5 discusses the results and conclusions.

2. LITERATURE REVIEW

2.1. Future-oriented Technology Analysis

Future-oriented technology analysis (FTA) is an umbrella concept which encompasses any systematic methodology for supporting decisions related to emerging technology, including its development and future impacts [1]. For a short time, FTA was known as technology futures analysis (TFA), but its name was later altered to give the idea of future-oriented, not technology-oriented [4]. The FTA concept is an evolution of two terms which describe the FTA facets: technology foresight and technology forecasting.

The military is heavily involved with FTA, since war efforts have resulted in advances in several areas of knowledge; for example, the nuclear, computing, internet, and space domains [5]. In the 1950s and 1960s, FTA was used by the military in the Cold War as a tool for planning and prioritizing the research and development of military technology, something which involves long development times. Today, commercial technological competitors use FTA as a tool for business intelligence. Competitive technological intelligence is concerned with forecasting the technology developments of others in order to constrain corporate research and development. National technology foresight, as well as industry or product technology roadmaps, show that industry and government are interested in planning, prioritizing, and evaluating their research and development programs [6].

2.2. Learning curves

Wright [7] noted that labour costs for production are progressively reduced in a non-linear way as the production becomes cumulative. This may be expressed by a curve, defined by Argote and Epple [8]:

$$y = ax^{-b} \quad (1)$$

In which:

- y is the labour time for the x^{th} unit;
- a is the labour time for the first unit;
- x is the number of units; and
- b is the learning rate (i.e., rate for reduction in labour time).

According to Dorroh et al. [9], cost researchers have noted that simple log-linear curve models are inadequate, and so they proposed several alternative learning curve specifications to address this inadequacy. For this reason, in this work, the learning curve model is inspired by a Gompertz function [10].

Empirical studies have shown that unit costs do not continue to decline indefinitely with cumulative knowledge, a phenomenon that is referred to as the plateau effect [11].

2.3. Knowledge Depreciation

Knowledge depreciation (KD) refers to the erosion in the knowledge inside organizations or individuals. In organizations, employee turnover may lead to loss of knowledge and learning [11]. When leaving an organization, the worker carries his unique knowledge, thus depreciating the organization's knowledge. The KD rate varies widely across industrial settings. Knowledge management initiatives may be used to resist this depreciation [12]. At the individual level, there

is also evidence of KD [11]. Learning new skills may provoke unlearning in others. Temporal memory decay is the major factor that results in forgetfulness [13].

2.4. Quality of data, information, and knowledge

According to Uotila and Melkas [3], data refers to the most basic descriptive element which represents a perception or measurement, lacking in content, meaning, or intent. Information is more than a set of data; it is the output of a process that interprets and manipulates data. Knowledge consists of an assortment of inputs: information, experiences, beliefs, relationships, and techniques, that an individual mentally synthesizes together to determine what a specific situation means and how to handle it. The terms and their meanings are shown in Table 1.

Table 1. Definition of terms, from [3].

Term	Meaning
Data	A collection of attribute values about some entity or event organized as a record and stored in some type of file structure (e.g., relational table).
Information	The output that results when a specified process manipulates raw source data and semi-processed component data items into a predefined format.
Knowledge (of an area)	A process by which, justified true beliefs about relationships between ideas, information, and experiences relevant to a particular area are used to generate situational awareness.
Data quality, information quality, and knowledge quality	Data, information products, and knowledge are of high quality if they are fit for their intended use in conducting business operations, decision making, and planning.

A user's perception of what is accurate, current, important, or useful is also determined by why they seek it [14]. Similarly to search engine algorithms, a single query may convey different meanings or goals, and better results could be achieved with a better understanding of the intention of a query [15].

The widely accepted information quality frameworks vary in approach and application, but they all share several characteristics regarding their classifications of quality dimensions [15]. Data quality is multidimensional and may be described according to the quality of the values collected [3]. When discussing data quality dimensions, it is reasonable to replace data with information or knowledge without losing the meaning [3]. Common methods to identify the quality dimensions associated with data are shown in Table 2.

Table 2. Methods to identifying data quality dimensions, from [3].

Approach	Definition	Example
Intuitive	The selection of quality attributes in a specific study is based on the individual's experience or intuitive understanding about what attributes are important.	[16]
Systematic	Focuses on how the output of a process may become deficient.	[17]
Empirical	Collects input from consumers to determine the characteristics they use to assess whether the end product is fit for use in their tasks.	[18]

Wang and Strong [18] proposed a hierarchical conceptual framework, in which groups of data quality dimensions are labelled. These representative labels capture the essence of the dimensions

from the data quality groups proposed: intrinsic, contextual, representational, and accessible. A summary of the common information quality dimensions and their short definitions are shown in Table 3.

Intrinsic data quality includes accuracy, objectivity, believability, and reputation. This suggests that accurate and objective data also needs believability and reputation in order to be considered to be high quality data. This is somewhat analogous to product quality [18].

Table 3.Common dimensions of data quality, from [15].

Dimension	Definition
Accuracy	Extent to which data are correct, reliable, and certified as free of error.
Consistency	Extent to which information is presented in the same format and is compatible with previous data.
Security	Extent to which access to information is appropriately restricted to maintain its security.
Timeliness	Extent to which the information is sufficiently up-to-date for the task at hand.
Completeness	Extent to which information is not missing and is of sufficient breadth and depth for the task at hand.
Conciseness	Extent to which information is concisely represented without being overwhelming (i.e., brief in presentation, yet complete and to the point).
Reliability	Extent to which information is correct and reliable.
Accessibility	Extent to which information is available, or quick and easy to retrieve.
Availability	Extent to which information is physically accessible.
Objectivity	Extent to which information is unbiased, unprejudiced, and impartial.
Relevancy	Extent to which information is applicable and helpful for the task at hand.
Usability	Extent to which information is clear and easily able to be used.
Understandability	Extent to which data are clear and unambiguous and can be easily comprehended.
Amount of data	Extent to which the quantity or volume of available data is appropriate.
Believability	Extent to which information is regarded as true and credible.
Navigation	Extent to which data are easily found and linked to.
Reputation	Extent to which information is highly regarded in terms of source or content.
Usefulness	Extent to which information is applicable and helpful for the task at hand.
Efficiency	Extent to which data are able to quickly meet the information needs for the task at hand.
Value-addedness	Extent to which information is beneficial and provides advantages from its use.

Accessible data quality includes accessibility and security. Data consumers view this as an important aspect of data quality. Treating accessible data quality as a data quality category or incorporating its dimensions into other categories makes almost no difference. In any case, accessibility dimensions are indispensable [18].

Contextual data quality includes relevancy, value-addedness, timeliness, completeness, and amount of information. The contextual data quality shows that data quality must be analysed in the context of the current task [18].

Representational data quality includes concise representation, consistent representation, interpretability, and ease of understanding. These aspects are related to the data format and meaning. For data consumers, well-represented data must not only be concise and consistently represented, but also interpretable and easy to understand [18].

Naumann and Rolker [19] grouped the entities and factors involved in information quality into subject, object, and process criteria. The identified criteria and their assessment methods are shown in Table 4.

Table 4. Classification of information quality criteria, from [19].

Assessment class	Criterion	Assessment method
Subject criteria	Believability	User experience
	Concise representation	User sampling
	Interpretability	User sampling
	Relevancy	Continuous user assessment
	Reputation	User experience
	Understandability	User sampling
	Value-addedness	Continuous user assessment
Object criteria	Completeness	Parsing, sampling
	Customer support	Parsing, contract
	Documentation	Parsing
	Objectivity	Expert input
	Price	Contract
	Reliability	Continuous assessment
	Security	Parsing
	Timeliness	Parsing
Process criteria	Verifiability	Expert input
	Accuracy	Sampling, cleansing techniques
	Amount of data	Continuous assessment
	Availability	Continuous assessment
	Consistent representation	Parsing
	Latency	Continuous assessment
	Response time	Continuous assessment

3. THE EXPERT SELECTION FRAMEWORK

The building of an expert selection framework was divided into three steps: the first step is defining basic knowledge concepts for which the data quality dimensions could rely on; the second step is defining data quality dimensions in the expert selection, using the knowledge concepts; and the third step is defining measurements of expert-oriented FTA methodology quality. In order to facilitate the understanding of the concepts, all concepts defined in this work have a truth value ranging between 0 and 1. This decision enables the later use of fuzzy variables [20] during the implementation phase.

3.1. Knowledge measurement

In order to measure knowledge, we use four concepts: a learning curve model, knowledge depreciation multiplier, knowledge gap, and effective knowledge.

3.1.1. Learning Curve

Although the original Wright learning curve was an exponential function, critics suggested using other curves, which are more suitable. In this work, we chose to model a learning curve that was inspired by a Gompertz function [10] and is defined in equation (2):

$$K(x) = e^{-de^{-cx}} \quad (2)$$

In which:

- $K(x)$ is the person's knowledge in the area;
- x is the amount of experience (time) in the area;
- d is related to the initial knowledge; and
- c is the learning rate.

The parameter c should be adjusted to scale the learning time to the user's scale (which may be in years, months, weeks, etc.). The parameter d should be adjusted to inform the default initial knowledge; that is, the common knowledge of the area. At the initial knowledge, x is set to zero and thus, $K(0) = e^{-d}$. The parameters c and d must be set as positive values and, consequently, this Gompertz function forms learning curves as shown in Figure 1 and Figure 2.

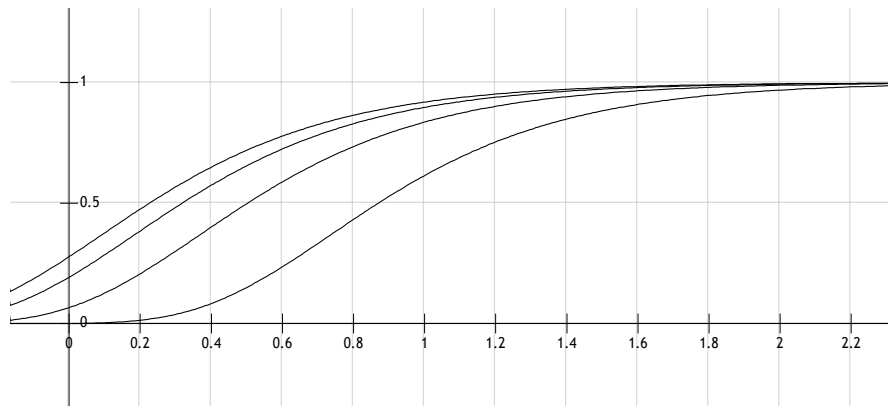


Figure 1. Learning curves in which only parameter d is varied.

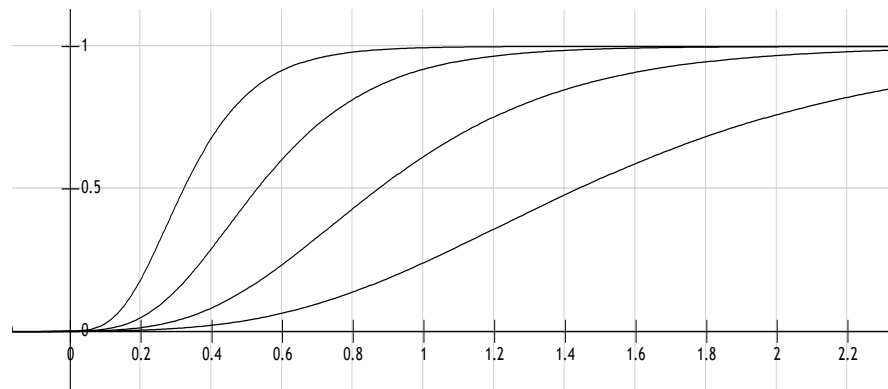


Figure 2. Learning curves in which only parameter c is varied.

3.1.2. Knowledge Depreciation (KD) multiplier

Another fundamental concept which is used in this work is Knowledge Depreciation. We defined a knowledge depreciation (KD) multiplier, which is given as equation (3):

$$KD(\Delta t) = 1 - \frac{\ln(r\Delta t + 1)}{\ln(r\Delta t + 1) + 1} \quad (3)$$

In which:

- KD is the knowledge depreciation multiplier;
- r is the knowledge depreciation (unlearning) rate; and
- Δt is the time without learning or practicing the knowledge.

The KD multiplier is the exponential decay in a person's knowledge of a specific area. This factor has a parameter r , which indicates the depreciation (unlearning) rate. The higher the value of r , the faster the knowledge is depreciated. Moreover, it is important to emphasize that the KD function is asymptotic, which means that a person needs an infinite amount of time to forget everything about some specific knowledge. Another interpretation of the KD multiplier is: What percentage of knowledge is lost once the possible KD values lie in the interval $[0, 1]$?

Examples of KD multiplier curves using progressive knowledge depreciation rates (r) are shown in Figure 3.

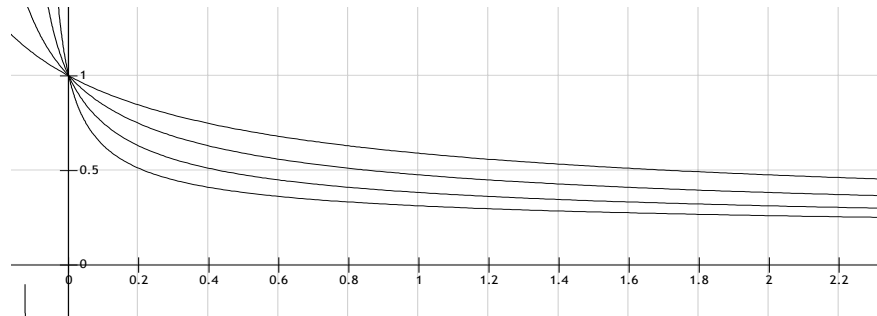


Figure 3. Knowledge Depreciation curves in which the parameter r is varied.

3.1.3. Knowledge Gap

The next concept is the Knowledge Gap (KG). The KG measures the amount of knowledge that could be learned if the person didn't stop learning. The KG is defined by equation (4):

$$KG(x, s) = K(x) - K(s) \quad (4)$$

In which:

- KG is the person's knowledge gap in the area;
- x is the amount of experience (time) in the area, if learning was not stopped; and
- s is the amount of experience (time) in the area at the moment in which the learning stopped.

Although the KG concept does not play a fundamental role in this work, it is important to define it and to show its significance in the learning curve graph, as shown in Figure 4.

3.1.4. Effective Knowledge

Finally, effective knowledge (EK) is the most important concept in the framework. This concept will be used to transport data quality ideas to the realm of expert selection. EK is defined as the depreciated knowledge, as equation (5) shows:

$$EK = K(x) KD(x - s) \quad (5)$$

In which:

- *EK* is the person’s effective knowledge in the area;
- *x* is the amount of experience (time) in the area, if learning is not stopped; and
- *s* is the amount of experience (time) in the area at the moment in which the learning stopped.

Due to KD being defined as exponential decay, as the *x - s* difference gets bigger, the KD multiplier decreases and, thus, the EK becomes lower. Figure 4 shows three curves: the original learning curve; the curve without knowledge depreciation, which stays constant; and the EK, considering the knowledge depreciation as defined in our framework.

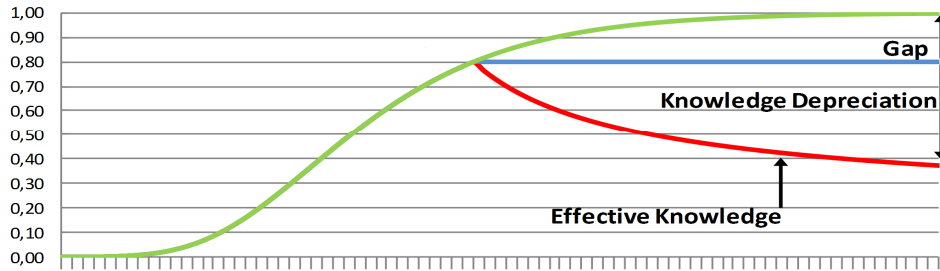


Figure 4. Knowledge Gap, Knowledge Depreciation, and Effective Knowledge.

3.1. Quality measurement for Expert Selection

In order to measure an expert selection, we use three data quality dimensions: accuracy, relevancy, and consistency. The data quality concepts were redefined to suit the expert selection framework and they depend only on the EK of the experts. It is important to remember that the EK lies in the [0,1] interval, which means that these dimensions lie in the same interval.

3.2.1. Accuracy

The first dimension to be defined in terms of EK is accuracy. Starting with the assumption that accuracy is the “extent to which data are correct”, in accordance with Table 3, accuracy is the mean effective knowledge available, as defined in equation (6):

$$Accuracy = \frac{\sum EK}{n} \quad (6)$$

In which:

- EK is the effective knowledge in the area; and
- n is the number of selected experts.

3.2.2. Relevancy

The second dimension is relevancy, which degrades faster than accuracy, especially if EK values are more accurate (i.e., closer to the mean value). At the maximum variation in EK values, relevancy overlaps accuracy, but never surpasses it. Relevancy is defined in equation (7):

$$Relevancy = \frac{\sum EK^2}{n} \quad (7)$$

In which:

- EK is the person's effective knowledge in the area; and
- n is the number of selected experts

3.2.3. Consistency

The third dimension is consistency, which measures the variation of the EK values. This result was achieved using the EK's standard deviation. We define the consistency dimension in equation (8):

$$Consistency = 1 - 2 \sqrt{\frac{\sum EK^2 - \frac{(\sum EK)^2}{n}}{n}} \quad (8)$$

In which:

- EK is the person's effective knowledge in the area; and
- n is the number of selected experts.

However, it is possible to replace the terms of equation (8) with the terms of equations (6) and (7), and reinterpret the consistency dimension. This shows that, in this framework, consistency is related to relevancy and accuracy. The result is equation (9):

$$Consistency = 1 - 2 \sqrt{Relevancy - Accuracy^2} \quad (9)$$

4. CASE STUDY

In order to measure the framework's capability to estimate the effective knowledge of an expert panel, the parameters c and d from equation (2), and the parameter r from equation (3), need to be estimated.

Thus, the experiment is divided into two phases: definition of parameters, and examples of the expert selection technique.

4.1. Definition of parameters

The first experiment is designed to produce the initial parameters c , d , and r , which are then used further. We selected 10 experts from our institution and asked 7 students to analyse their curricula available in Lattes Platform [21] – the Brazilian government's platform which integrates researcher's curricula, Research Groups, and Academic Institutions in a single information system. They had to analyse each expert's curriculum, in order to find publications or other proof of activity in the research subject and thus determine:

- Activity time in one specific area, in years (AcT);
- Inactivity time in the same area, in years (InT);
- Knowledge when leaving the area, in [0,1] interval (MaxK);
- Actual knowledge or effective knowledge (EK) in the area.

Computer supported collaborative work (CSCW) was the chosen area for this part of the experiment. The data collected was grouped in Table 5.

Table 5. Data collected from students.

Expert number	AcT		InT		MaxK		EK	
	Avg	σ	Avg	σ	Avg	σ	Avg	σ
1	18.43	1.72	2.86	0.38	0.84	0.13	0.68	0.19
2	10.43	6.53	7.00	6.19	0.63	0.22	0.47	0.30
3	25.14	2.41	1.14	3.02	1.00	0.00	0.97	0.08
4	9.00	2.83	1.57	2.44	0.69	0.10	0.63	0.13
5	11.14	1.21	0.29	0.49	0.81	0.13	0.79	0.17
6	27.86	7.60	1.29	3.40	0.97	0.08	0.94	0.11
7	6.57	2.64	5.57	3.60	0.59	0.20	0.36	0.32
8	9.43	3.51	1.57	1.13	0.73	0.27	0.64	0.27
9	14.75	4.45	2.66	1.70	0.78	0.15	0.68	0.19
10	14.75	4.24	2.66	1.86	0.78	0.09	0.68	0.10

In the next step, the data was plotted in a graph and the learning curve (Figure 5) and the knowledge depreciation curve (Figure 6) were interpolated.

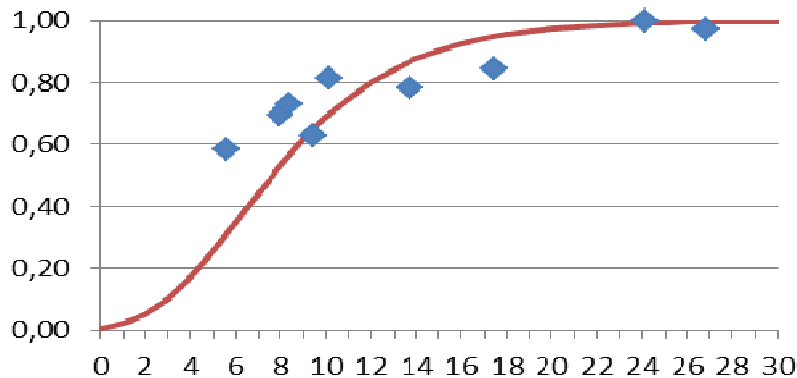


Figure 5. Learning curve from the first experiment.

Analysing the curve patterns, the parameters $d = 5$, $c = 0.26$, and $r = 0.07$ were chosen as the best fit. It is important to note that these parameters may vary; that is, other curves may have as good a fit as this one. The lack of non-experts in this experiment (people with less than 7 years of experience) made the plotting harder at the beginning of the curve. To solve this problem, we fixed the value for the parameter d to 5, in order to force the initial knowledge to be zero. Then, the c value, which minimizes the total distance of the data points to the curve, was calculated using iterative numerical methods. The parameter d was calculated using the same method as for parameter c .

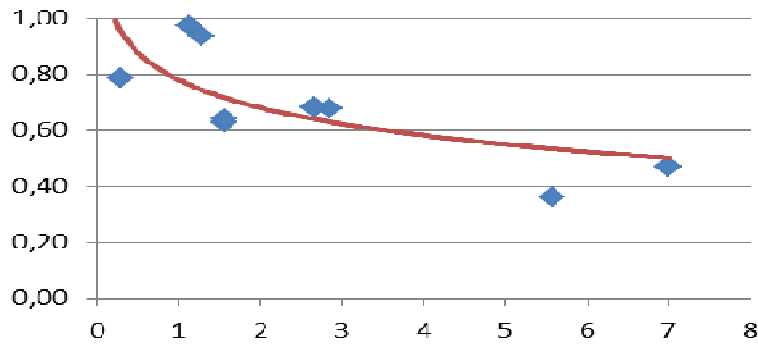


Figure 6. Knowledge depreciation for the first experiment.

Using the resulting data form the experiment, the depreciation curve of $r = 0.07$ was plotted to highlight the significance of the depreciation curve. The depreciation curve, limited to a 10-year window, is shown in Figure 7.

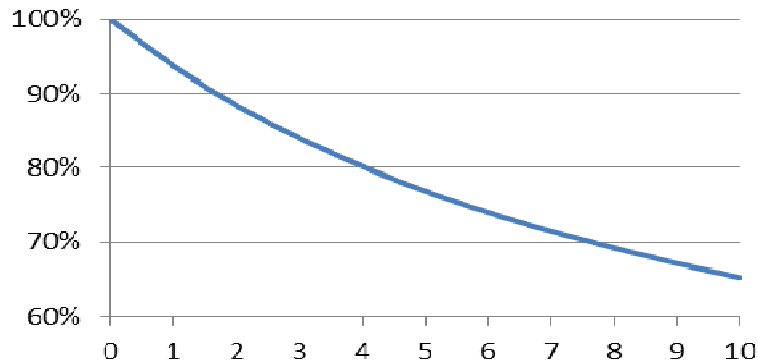


Figure 7. Knowledge depreciation in a 10-year window.

Analysing the depreciation curve, the proposed model indicates that an expert keeps only 65% of the original knowledge after 10 years of inactivity. This conclusion is made because the value of the function $KD(10) = 0.65$.

4.2. Examples of expert selection

The second experiment is a practical application of the expert selection framework. The objective in this experiment is to select two groups of five experts from an initial group of thirty candidates. The size of the initial expert group and the subset to be selected by the algorithm are completely arbitrary. In fact, this algorithm may be used to only rank the experts. The expert areas chosen to be used in this experiment were knowledge management and administration.

In the first step of the algorithm, the candidates were chosen ad hoc. In the second step, their EK in the two areas was measured. Seven students studied their curricula in Lattes Platform [21] to identify their activity time (AcT) and inactivity time (InT) in each area. We computed the average AcT and InT, as well as the standard deviation. With these average AcT and InT values, equations (2), (3) and (5) were used to estimate the EK of each expert in each area, with parameters from the first experiment ($d = 5$, $c = 0.26$, and $r=0.07$). The results are shown in Table 6.

In the third step, a final list of selected experts was made for each area, by selecting the experts with higher EK. In this theoretical example, there is no risk of experts refusing to participate. The lists for each area, ranked by expertise (the estimated EK), are shown in Table 7 and Table 8. In the next step, the expert selection quality was measured. To achieve the objective, the accuracy, relevancy, and consistency of both expert selections were calculated. The results are shown in Table 9.

According to our framework, the expert selection based on administration was slightly better than the expert selection based on knowledge management, since it has higher accuracy and relevancy. However, its consistency was a little bit lower, which means that the experts have higher EK variation.

Table 6. Activity, inactivity, and estimated effective knowledge.

Expert	Knowledge management					Administration				
	AcT		InT		Est. EK	AcT		InT		Est. EK
	AVG	σ	AVG	σ		AVG	σ	AVG	σ	
1	12.4	2.3	1.3	1.5	0.76	15.7	8.0	3.7	4.6	0.75
2	17.4	6.1	2.1	1.9	0.83	28.1	8.3	5.7	6.1	0.75
3	14.4	5.8	1.9	2.7	0.79	14.1	6.3	6.4	6.2	0.64
4	11.4	1.4	0.3	0.5	0.76	10.6	5.0	1.4	3.8	0.66
5	7.4	2.2	5.9	1.7	0.36	8.4	2.6	6.9	1.8	0.41
6	8.0	5.6	7.1	3.8	0.38	15.9	8.3	4.6	3.8	0.72
7	17.6	11.4	1.1	2.3	0.88	22.6	9.6	2.3	4.5	0.86
8	14.1	2.0	0.1	0.4	0.87	13.7	6.5	1.0	1.7	0.81
9	12.1	3.0	2.1	2.0	0.71	10.6	5.7	2.3	3.2	0.63
10	12.7	1.8	2.0	1.5	0.74	7.7	7.8	8.0	9.1	0.35
11	6.7	1.6	1.3	0.5	0.38	10.7	5.1	13.0	15.1	0.45
12	4.6	9.5	6.7	4.6	0.16	14.4	8.5	4.0	4.0	0.71
13	11.9	1.5	1.1	1.2	0.74	22.6	16.4	1.0	1.7	0.92
14	3.1	1.1	9.9	0.4	0.07	7.3	6.9	2.9	4.6	0.40
15	13.3	3.9	1.9	2.3	0.76	16.6	2.4	1.9	2.7	0.83
16	20.7	7.1	1.3	2.2	0.90	17.9	12.4	1.4	1.8	0.87
17	2.7	1.6	8.7	2.8	0.06	11.4	6.0	10.1	11.7	0.50
18	22.3	8.8	4.9	5.8	0.76	18.6	7.9	5.0	7.9	0.74
19	13.3	4.2	0.6	0.8	0.82	21.9	5.6	0.3	0.5	0.96
20	9.6	8.2	3.4	3.4	0.54	19.6	7.9	1.9	3.3	0.86
21	10.4	5.3	1.9	2.2	0.64	21.9	5.8	1.0	1.7	0.92
22	14.8	2.1	0.4	0.5	0.87	14.4	2.6	1.1	1.6	0.83
23	15.1	4.4	0.0	0.0	0.91	20.4	4.3	0.7	1.3	0.93
24	14.4	4.9	1.0	2.2	0.83	35.9	5.8	1.3	2.2	0.92
25	5.4	6.6	2.0	1.7	0.26	18.7	5.7	3.3	4.6	0.80
26	8.3	2.1	0.3	0.8	0.55	8.4	3.9	2.4	3.3	0.49
27	7.7	3.7	0.6	1.1	0.49	10.4	3.6	2.0	2.1	0.63
28	8.6	6.2	5.7	3.9	0.44	22.1	3.2	0.0	0.0	0.98
29	16.7	8.7	1.7	2.4	0.84	26.1	4.8	3.6	3.8	0.81
30	14.9	2.7	0.3	0.8	0.88	19.3	12.4	1.1	2.0	0.90

Table 7.Selected experts in knowledge management.

Rank	Expert	AcT	InT	EK
1	23	15.1	0.0	0.91
2	16	20.7	1.3	0.90
3	30	14.9	0.3	0.88
4	7	17.6	1.1	0.88
5	22	14.8	0.4	0.87

Table 8.Selected experts in administration.

Rank	Expert	AcT	InT	EK
1	28	22.1	0.0	0.98
2	19	21.9	0.3	0.96
3	23	20.4	0.7	0.93
4	13	22.6	1.0	0.92
5	21	21.9	1.0	0.92

Table 9.Expert selection quality.

Expert selection	Accuracy	Relevancy	Consistency
Knowledge management	0.89	0.79	0.97
Administration	0.94	0.89	0.95

5. CONCLUSIONS

This work presented a data quality framework for forming a panel of experts. The proposed framework defined a method for measuring effective knowledge, based on the learning curve and knowledge depreciation. The effective knowledge may be used to measure the accuracy, relevancy, and consistency of a group of experts; that is, the inherent quality of them.

The experiments achieved their objectives: we successfully estimated the effective knowledge of the experts, and we were able to select experts. The intrinsic quality of the selection was also successfully measured.

ACKNOWLEDGEMENTS

We would like to thank CAPES and CNPq for their financial support of this work.

REFERENCES

- [1] A. L. Porter, W. B. Ashton, G. Clar, J. F. Coates, K. Cuhls, S. W. Cunningham, K. Ducatel, P. van der Duin, L. Georgehiou, T. Gordon, H. Linstone, V. Marchau, G. Massari, I. Miles, M. Mogege, A. Salo, F. Scapolo, R. Smits, and W. Thissen, "Technology futures analysis: Toward integration of the field and new methods," *Technol. Forecast. Soc. Change*, vol. 71, no. 3, pp. 287–303, 2004.
- [2] J. Surowiecki, *The wisdom of crowds*. Anchor, 2005.
- [3] T. Uotila and H. Melkas, "Quality of data, information and knowledge in regional foresight processes," *Futures*, vol. 39, no. 9, pp. 1117–1130, 2007.
- [4] R. Johnston, "Historical Review of the Development of Future-Oriented Technology Analysis," in *Future-Oriented Technology Analysis*, C. Cagnin, M. Keenan, R. Johnston, F. Scapolo, and R. Barré, Eds. Springer Berlin Heidelberg, 2008, pp. 17–23.
- [5] V. W. Ruttan, *Is war necessary for economic growth? Military procurement and technology development*. Oxford University Press, USA, 2006.
- [6] A. L. Porter, "Tech forecasting an empirical perspective," *Technol. Forecast. Soc. Change*, vol. 62, no. 1–2, pp. 19–28, 1999.
- [7] T. P. Wright, "Factors affecting the cost of airplanes," *J. Aeronaut. Sci. Inst. Aeronaut.Sci.*, vol. 3, 1936.
- [8] L. Argote and D. Epple, "Learning Curves in Manufacturing," *Science*, vol. 247, no. 4945, pp. 920–924, Feb. 1990.
- [9] J. R. Dorroh, T. R. Gullledge Jr., and N. K. Womer, "A generalization of the learning curve," *Eur. J. Oper. Res.*, vol. 26, no. 2, pp. 205–216, 1986.
- [10] E. Tjørve, "Shapes and functions of species–area curves: a review of possible models," *J. Biogeogr.*, vol. 30, no. 6, pp. 827–835, 2003.
- [11] G. Li and S. Rajagopalan, "A learning curve model with knowledge depreciation," *Eur. J. Oper. Res.*, vol. 105, no. 1, pp. 143–154, 1998.
- [12] D. E. Rose and D. Levinson, "Understanding user goals in web search," in *Proceedings of the 13th international conference on World Wide Web*, New York, NY, USA, 2004, pp. 13–19.
- [13] A. Dedeke, "A Conceptual Framework for Developing Quality Measures for Information Systems," in *Proceedings of the 2000 Conference on Information Quality (IQ-2000)*, Cambridge, MA, USA, 2000, pp. 126–128.
- [14] T. C. Redman, *Data quality: the field guide*. Access Online via Elsevier, 2001.
- [15] Y. Wand and R. Y. Wang, "Anchoring data quality dimensions in ontological foundations," *Commun ACM*, vol. 39, no. 11, pp. 86–95, 1996.
- [16] R. Y. Wang and D. M. Strong, "Beyond accuracy: what data quality means to data consumers," *J Manage InfSyst*, vol. 12, no. 4, pp. 5–33, 1996.
- [17] F. Naumann and C. Rolker, "Assessment Methods for Information Quality Criteria," in *Proceedings of the International Conference on Information Quality (IQ)*, 2000.
- [18] M. L. Puri and D. A. Ralescu, "Fuzzy random variables," *J. Math.Anal.Appl.*, vol. 114, no. 2, pp. 409–422, 1986.
- [19] CNPq - National Council for Scientific and Technological Development, "Lattes Platform." [Online]. Available: <http://buscatextual.cnpq.br/buscatextual/busca.do>.

AUTHORS

Carlos Eduardo Barbosa is a D.Sc. student in the Universidade Federal do Rio de Janeiro (UFRJ), where he researches about FTA. His research interests also include Big Data, Decision-support Systems, Knowledge Management, and Crowdsourcing. He also participated and managed projects in the Brazilian Navy



MarcioAntelio is a D.Sc. student in the Universidade Federal do Rio de Janeiro and an associate professor in the Federal Center for Technological Education Celso Suckow da Fonseca. His research interests include Data Quality, Crowdsourcing, and Big Data.



Jano Moreira de Souza is a titular professor of the Universidade Federal do Rio de Janeiro(UFRJ), where he teaches and researches about Knowledge Management, Crowdsourcing, Big Data, Data Quality, Negotiation Support Systems, and Autonomic Computation.



Jonice Oliveira is an associate professor of the Universidade Federal do Rio de Janeiro. Her research interests include Knowledge Management, Social Network Analysis, Competitive Intelligence, Mobile Computing, and Semantic Web.

