A Semantic-Based Model to Assess Information for Intelligence

J. Besombes,
L. Cholvy,
V. Dragos
(Onera)

E-mail: jerome.besombes@onera.fr

This paper addresses the problem of information evaluation for Intelligence. Starting from NATO recommendations for assessing information, we propose a semantic-based model to evaluate information. We also define a semi-automatic evaluation process in which an ontology is used, to detect similar items of information. Semantically similar items are then presented to an operator in charge, to estimate their correlations. Items of information are electronic documents exploited for intelligence purposes and they are considered in a broader sense with respect to their form (structured files or free-form text) and their content (description of images of video scenes, HUMINT reports). Linguistic variance, an inherent feature of textual data, can also be handled by using the ontology, while human intervention during the evaluation process ensures a good quality outcome. Finally, we show that this process is compliant with NATO recommendations, while going beyond their limitations.

Introduction

Information evaluation appears as a critical capability for many military applications aimed at offering decision support, since there is an obvious need for valuable information to be transferred at every level of the military chain of command. Information is evaluated by systems able to estimate the degree of confidence that can be assigned to various items of information obtained for intelligence purposes.

In the military field, NATO ([19], [20]) recommendations promote an alphanumeric system for information rating, which takes into account both the *reliability* of the source providing the information and its *credibility*, as it appears when examined in the light of existing knowledge.
Reliability of the source is designated by a letter between A and F expressing various degrees of confidence as follows:

- a source is evaluated A if it is completely reliable. It refers to a tried and trusted source which can be depended upon with confidence
- a source is evaluated B if it is usually reliable. It refers to a source which has been successfully used in the past but for which there is still some element of doubt in particular cases.
- a source is evaluated C if it is fairly reliable. It refers to a source which has occasionally been used in the past and upon which some degree of confidence can be based.
- a source is evaluated D if it is not usually reliable. It refers to a source which has been used in the past but has proved more often than not unreliable.
- a source is evaluated E if it is unreliable. It refers to a source which has been used in the past and has proved unworthy of any confidence.
- a source is evaluated F if its reliability cannot be judged. It refers to a source which has not been used in the past.

Credibility of information is designated by a number between 1 and 6, signifying varying degrees of confidence as defined below:

- If it can be stated with certainty that the reported information originates from another source than the already existing information on the same subject, then it is classified as "confirmed by other sources" and rated 1.
- If the independence of the source of any item of information cannot be guaranteed, but if, from the quantity and quality of previous reports, its likelihood is nevertheless regarded as sufficiently established, then the information should be classified as "probably true" and given a rating of 2.
- If, despite there being insufficient confirmation to establish any higher degree of likelihood, a freshly reported item of information does not conflict with the previously reported behaviour pattern of the target, the item may be classified as "possibly true" and given a rating of 3.
- An item of information which tends to conflict with the previously reported or established behaviour pattern of an intelligence target should be classified as "doubtful" and given a rating of 4
- An item of information which positively contradicts previously reported information or conflicts with the established behaviour pattern of an intelligence target in a marked degree should be classified as "improbable" and given a rating of 5
- An item of information is given a rating of 6 if its truth cannot be judged.

It can be noted that these natural language definitions are imprecise and ambiguous, and they can lead to twofold interpretations. For instance, according to the previous recommendations, the reliability of a source is defined with respect to its previous use, while completely ignoring their current usage context, i.e., the actual environment of use of this source.

As for the credibility of information, the rating defined previously does not qualify a unique property. For instance, how should we note an item of information supported by several sources of information that are also in conflict with some already registered information? According to these definitions, this item should be given a credibility value of 1, but also of 5.

Furthermore, according to NATO recommendations, a rating of 6 should be given to every item whose truth cannot be judged. This supposes that other ratings (1...5) concern the evaluation of information truth value. If so, a rating of 1 corresponds to true information. But, according to its definition, a rating of 1 should be given to an item supported by at least two sources, which is questionable since several different sources may provide false information despite their mutual agreement.

In the light of the discussion above, it becomes obvious that a proper use of those recommendations requires their disambiguation and formalization. The aim of this work is to provide a semantic-based model to evaluate information, based on formal definitions of notions being at the heart of NATO recommendations. It also defines a semi-automatic evaluation process, in which an ontology is used to detect similar items of information. Semantically similar items are then presented to an operator in charge, to estimate their correlations. The underlying applicative scenario of this work implies a timely processing of a constant stream of information provided by various sources, in order to achieve intelligences specific tasks. We consider complex information, such as HUMINT[1] reports or textual descriptions of video scenes. By taking into account semantic aspects it becomes possible to perform enhanced treatments, going beyond key-word spotting and analysis.

The outline of the paper is as follows. First, it presents a brief state of the art on information quality and evaluation, as tackled within various related research fields. "General Framework for information evaluation" introduces formal definitions of the key-notions of NATO recommendations and describes the general architecture supporting the information evaluation process, while "Supporting human opertors through semantics" focuses on the use of ontology to identify semantically similar information items. "Discussion" proves that the outcome of the overall process is consistent with NATO recommendations. Conclusions and future work directions are presented to end this paper.

State of the art

Information evaluation is closely related to the notion of *information quality*. Indeed, traditionally defined as "fitness for use" [14], information quality investigates the estimation of the information capacity to accomplish a specific task, such as information querying, information retrieval or information fusion, for instance. With this respect, information quality appears as a complex notion, covering various aspects. Hence, [3] and [28] identify several dimensions for information quality, among which we found: *intrinsic data quality* (believability, accuracy, objectivity), which is defined by considering the information itself, independently of its production or interpretation frames; *contextual data quality* (relevancy, timeliness, completeness) consists of dimensions related to both *contexts of production* or interpretation of information; *representational data quality* (interpretability, Ease of understanding,) is related to various formalisms used to represent data, having a direct impact on the effective use of that information; finally, *accessibility data quality dimension* concerns protocols to access information, while ensuring their security. In [23], an ontology of information quality attributes is proposed and the issue of combining several of these attributes into a single measure, and of how to take into account quality measure for decision making, is discussed from a semi-automatic fusion system perspective.

Among these various dimensions, *information evaluation* refers to information accuracy, known to be one of the most important dimensions, since it expresses the quality of information being true or correct.

Furthermore, let us mention the relationship between the quality of a model (such an ontology can be) and the quality of information described within this model [27]. According to [27], an information model should be complete (every real world situation can be expressed as an item of information in the model), *unambiguous* (to each item of information expressed in the model corresponds a unique real world situation), *meaningful* (each item of information expressed by the model corresponds to a real world situation) and correct (a user can derive a real world situation from the expressed item of information). Thus, bad quality models lead to bad quality of expressed information; an incomplete model leads to possible unreliable information; an ambiguous model leads to possible imprecise information; a meaningless model leads to possible irrelevant information and an incorrect model leads to possible non-interpretable information.

Since more and more information is produced in heterogeneous and highly dynamic environments, information evaluation emerged as a research topic. Thus, several research efforts have being conducted to estimate the quality of information exploited in various application fields, such as the exploitation of open sources, information retrieval or the management of medical knowledge.

For instance, [3] provides a solution to assess the quality of information retrieved on the Web by defining filtering policies. Those policies combine various meta-data available on data sets, Describing the applicative context of particular information, along with its content, into an overall filtering decision. Policies can be embedded into a browser, allowing the pertinence of its outcome with respect to a user query to be improved, so information is evaluated dynamically. From a different perspective, [15] propose the QUATRO approach for Web content labeling, which ends up with a static qualitative description of Websites. This approach provides a common vocabulary to express quality labels to be assigned to Web content, along with mechanisms to check the validity of those labels. The result of this approach is a unified qualitative description of Websites, which can be taken into account for further treatment, since the information gathered from various sites can be amended by the quality label of its source. Beyond the content itself and its description, Websites also appears as both the source and the target of various links from or to other Websites. [29],[30] consider a link between two pages as an implicit conveyance of trust from the source page to the target page and use these links to define measures expressing trust and distrust of Websites. Moreover, these measures can be propagated through the Web by following the link network. In a more particular context, [24] propose the authority coefficient as a measure characterizing a news blog by

[1] HUMan INTelligence

taking into account its credibility, as assessed by user comments, its confidence, as expressed by the number of pertinent links referred to, and its influence, provided by the number of external sources referring to the news blog considered.

Approaches cited above use external elements, such as various sets of meta-data, user comments, or citations of the considered source by others, while ignoring its own content.

In the medical field, [13] go further and apply information extraction techniques, in order to retrieve valuable information within large amounts of scientific articles and also provide means to characterize this information. Hence, a confidence score is estimated by retrieving linguistic patterns expressing certainty or uncertainty and analyzing the context of information occurrences. While this solution is restricted to a linguistic level, other approaches are performing deeper content analysis, by taking into account the explicit semantics of the application domain. Among them, [10] estimate a coherence coefficient of natural language phrases by mirroring ontological entities appearing in those phrases with respect to an existing ontology. From an interesting perspective, [1] consider both named entities, at a linguistic level, and ontological entities, at a semantic level, to discriminate between negative and positive opinions on news blog threads.

More specifically, many works have addressed the question of information evaluation modeling for military purposes. Most of these can be found in the proceedings of the annual international conference FUSION [8] or in the proceedings of NATO RTO (Research and Technology Organization) symposiums [18]. Let us mention for instance [31], [25] and very recently [26]. All of these works are aimed at defining methods to assess the believability of the information gathered from several sources. These methods take into account different sources with their own degree of reliability, the fact that the item of information considered is consistent or not with previous reported information. Some methods also take into account the fact that, in some cases, the source may include its own assessment of the trustworthiness of the data that it has transmitted.

All of these papers mention NATO STANAG as underlying or motivating their model, but none of them formally prove that their model fulfills the recommendations.

Finally, we would like to mention our own previous works [2][4][5][6][21]. In [2], we have defined information evaluation according to the reliability and competence of the source that provides the item of information and also according to the plausibility and credibility of the item of information. However, a formal relation with NATO recommendations has not been established.

In [4], [5], [6], [21], we have tried to formalize the informal NATO recommendations by proposing formal models. These models are based on the fact that the three main notions underlining the informal NATO recommendations are: the number of independent sources that support an item of information, their reliability and whether the items of information are in conflict or tend to be in conflict. In particular, [4] and [5] present a logical definition of evaluation, based on the number and on the reliability of the sources that supports an item of information. The related fusion method is a weighted sum and is an obvious extension of the majority method defined in [17]with Hamming distance between logical interpretations. [6] proves that this method implicitly takes into account degrees of conflict between items of information.

These works assume that information items are described using logical languages, allowing the implementation of completely automatic reasoning procedures to analyze them. In particular, they assume that the degree of conflict between two items of information is automatically computed. However, these methods prove to be limited when dealing with complex information, such as natural language reports, for instance.

This is why we propose a new solution for information evaluation, going beyond those limitations, since it is able to handle various types of complex information, such as descriptions of video scenes or images, HUMINT reports, etc. This is a semi-automatic approach and it requires human intervention during some key-phases of the evaluation process. Support is provided to the operator when performing the task, since semantically close items of information are gathered together thanks to the ontology. This new approach is detailed hereafter.

A general framework for information evaluation

General architecture

We propose a general architecture for information evaluation, based on basic treatment cells called evaluators, see figure 1.

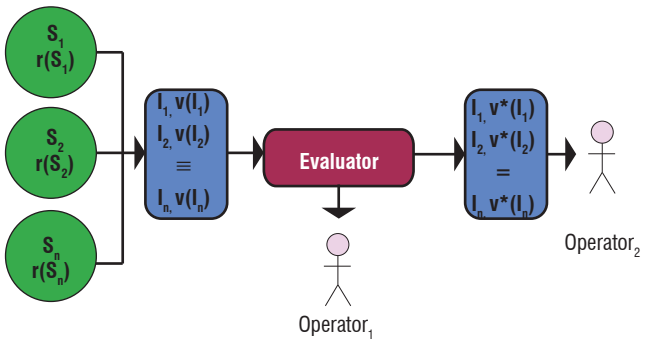

Figure 1 - Architecture for information evaluation

The *evaluator* is a treatment cell managed by an operator, collecting information provided by one or several sources. The input of the evaluator is a set of various information items, while its output is the initial set augmented with confidence scores assigned to each item of information.

Information evaluation is carried out in a semi-automatic manner, with interventions by human operators assessing the quality of each item under analysis. The user can also update the reliability of the sources that provide those items of information.

For this work, the *items of information* to be analyzed are natural language reports, I_1, I_2,.... In emitted by sources named S_1... S_m. Each source S_i is associated with its *degree of reliability* $r(S_i)$, a real number ranging between 0 and 1, where 0 corresponds to a source considered as non-reliable by the operator, while 1 corresponds to a source providing highly reliable information.

Hence, given two sources S_i *and* S_j:

- $r(S_i) < r(S_j)$ means that the operator thinks that source S_i is less reliable than source S_j.
- $r(S_i) = 0$ means that the operator thinks that S_i is not at all reliable.
- $r(S_i) = 1$ means that the operator thinks that S_i is fully reliable.

In our model, Each item of information is associated with its *evaluation*, denoted by $v(i)$, which is a real number between 0 and 1. Since our starting point are NATO recommendations, the evaluation of information (i) takes into account the two key notions of those recommendations, which are the correlations between various items of information under analysis and the reliability of their sources.

Information correlation

Definition 1 Let I_1 and I_2 be two different items of information. Their degree of correlation, denoted by α_{I_1,I_2}, is a real number in [-1, +1] that the operator will associate with I_1 and I_2 so that:

- $\alpha_{I_1,I_2} < 0$ if and only if I_1 tends to contradict I_2
- $\alpha_{I_1,I_2} > 0$ if and only if I_1 tends to confirm I_2
- $\alpha_{I_1,I_2} = 0$ else.

It's worth noticing that for any two items of information I_1 and I_2, we don't have the property $\alpha_{I_1,I_2} = \alpha_{I_2,I_1}$ since the notion of confirmation is not symmetric.

As a counter-example, consider that I_1 is "It rained last night" and that I_2 is "the road is wet". Given background knowledge according to which rain wets, it is the case that I_1 implies I_2. Thus, $=1$. However, I_2 does not imply I_1. Thus, here <1.

Definition 2 Two items of information I_k and $I_{k'}$ are equivalent in the database of the evaluator $\{I_1 \ldots I_n\}$ if and only if:

- $\alpha_{I_k,I_{k'}} = \alpha_{I_{k'},I_k} = 1$
- For any $i = 1 \ldots n$ $\alpha_{I_k,I_i} = \alpha_{I_{k'},I_i}$

Information evaluation process

The general process of information evaluation is carried out as follows: each item of information enters the evaluator with its initial evaluation, granted according to the reliability of its source. The more reliable the source is, the more important this value is. This value is then constantly updated by the evaluator, as new items of information are gathered. Thus, at the level of the evaluator, if an item of information is emitted by a source, then its initial evaluation is defined by the reliability degree of the source that emitted it (plus some corrections due, for instance, to the conditions of use of this source; see [2] for the definition of various criteria for the qualification of an item of information).

Assume that the knowledge base of the considered evaluator contains the following items of information: $I_1 \ldots I_{n-1}$ associated with their respective current evaluation: $v(1), \ldots, v(n-1)$.

Consider a new item of information In, associated with its evaluation $v(n)$. In has been emitted by a source whose reliability is $r(S_n)$. In this case, $v(n) = r(S_n)$, i.e., the current evaluation of In is defined as the degree of reliability of the source that emitted it. Let us denote by the evaluation of any item of information I_k updated after the arrival of I_n.

We define the updated evaluation by:

$$v_k^* = \frac{\sum_{k' \in K_k} (v_{k'}.\alpha_{I_k,I_{k'}}) + |K_k|}{2|K_k|}$$

Where $K_k = \{k' : 1 \le k' \le n,\ k' \neq k$ and $\alpha_{I_k,I_{k'}} \neq 0\}$

Notice that: $0 \le v_k^* \le 1$

Indeed, the numerator of the previous fraction is minimal when for any $k' \neq k, \alpha_{I_k,I_{k'}} = -1$ and $v_{K'} = 1$. In this case, it is equal to 0, thus the fraction is equal to 0. Furthermore, the numerator is maximal when for any $k' \neq k, \alpha_{I_k,I_{k'}} = 1$ and $v_{K'} = 1$. In this case, the numerator is equal to $|K_k| + |K_k|$, i.e, $2|K_k|$ Due to the denominator, the fraction is equal to 1 in this case.

This function is such that:

- If $I_{k'}$ tends to confirm I_k, then $v_{k'}.\alpha_{I_{k'},I_k} \ge 0$. Indeed $v_{k'} \ge 0$ and $\alpha_{I_{k'},I_k} > 0$. Thus, this factor increases the evaluation of I_k.
- If $I_{k'}$ tends to contradict I_k, then $v_{k'}.\alpha_{I_{k'},I_k} \le 0$. Indeed, Even if $v_{k'} \ge 0, \alpha_{I_{k'},I_k} < 0$. Thus, this factor decreases the evaluation of I_k.
- If $\alpha_{I_{k'},I_k} = 0$, then $v_{k'}.\alpha_{I_{k'},I_k} = 0$. Thus, this factor does not modify the evaluation of I_k.
- If $k=k'$, then $v_{k'}.\alpha_{I_{k'},I_k} \ge 0$ and increases the evaluation of I_k.
- If $v_{K'}^* = 0$ then $v_{k'}.\alpha_{I_{k'},I_k} = 0$ and thus does not modify the evaluation of I_k.

Proposition 1 - If two items of information I_k and $I_{k'}$ are equivalent in the database of the evaluator, then $v_K^* = v_{K'}^*$ whatever v_K and $v_{K'}$ are.

Thus, at the level of the evaluator, human intervention is needed to qualify each item of information under analysis. However, a real time processing of large amounts of information makes manual solution an overwhelming task, especially when information arrives as natural language reports. To cope with this difficulty, we propose a semi-automatic approach, whose treatment regroups various information items according to their semantic similarity, and human intervention is required to analyze them if and only if the semantic similarity is above a given threshold.

Supporting human operators through semantics

In some simple cases, the correlation $\alpha_{I,I'}$ between two items of information I and I' can be computed automatically. More generally, the way correlation degrees can be calculated is related to the way in which information is produced. Assume that the set of valid information that the system manages is finite; then, for every possible pair of items of information, the correlation degree can be pre-defined. For instance, correlation rules can be defined for items of information whose specific elements are date, time and location of events, as we show hereafter.

Example The dating of an event related in a textual document can be done from the extraction of named entities corresponding to the pattern $m/d/y$[2], where m=MM, D=DD, y=YYYY|YY. Now, let us suppose two dates d_1 and d_2 respectively defined by $d_1 = m_1/d_1/y_1$

[2] m stands for month, d for day and y for year.

and $d_2 = m_2 / d_2 / y_2$. For instance, the correlation between d_1 and d_2 can be given by:

- α_{d_1,d_2} iff $d_1 = d_2$
- $\alpha_{d_1,d_2} = 0.9$ iff $d_1 = d_2$ and $m_1 = m_2$ and
  - $y_1 = 20\, y_2$ if y_1 is of the form YYYY and y_2 of the form YY or
  - $20\, y_1 = y_2$ if y_1 is of the form YY and y_2 of the form YYYY
- $\alpha_{d_1,d_2} = -1$ iff $d_1 \neq d_2$ *or* $m_1 \neq m_2$ or $y_1 \neq y_2$.

However, most of the time, items of information are too complex to have a correlation degree automatically calculated and human intervention is required. In this paper, we propose a semi-automatic approach to estimate the correlation of information items, based on the use of previous knowledge modeled by an ontology. This approach takes advantage of semantic annotations of information items and uses the semantic similarity in order to estimate the correlation between them.

Using ontologies to estimate semantic similarity

According to [9], an ontology is defined as a formal and explicit specification of a shared conceptualization. Ontologies are artifacts modeling domain knowledge by taking into account both the conceptual and linguistic levels. The conceptual level concerns the modeling of field entities, along with the relations that hold between them. The linguistic level is related to the use of natural language terms to name ontological entities. From a linguistic standpoint, named entities are instances of concepts.

By offering this two-fold description of domain knowledge, ontologies offer means to handle the linguistic variety and provide a good basis to perform text analysis by going beyond key-word spotting. On the other hand, the description of items of information by ontological entities allows enhanced reasoning capabilities.

The user is then required to define the degree of correlation between two items of information only if their semantic similarity degree is over a given threshold, as we can see in figure 2.

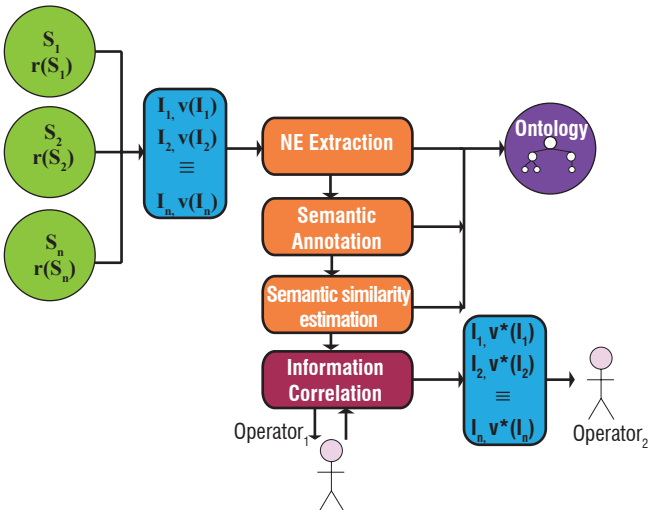

Figure 2 - Semi-automatic evaluation of information

The input of this chain of treatments is the set if information items to be evaluated and it supports treatments for: named entity extraction, semantic annotation and semantic similarity estimation, as described hereafter.

Named entities (NE) extraction

The goal of this processing step is to automatically identify named entities appearing within natural language reports. In order to illustrate our model, let us consider a very simple ontology leading to annotations of the form (where, when, who). This means that named entities retrieved within reports will describe the place of an event (the city, or more precisely, one of its places), time coordinates of the event (date, week day, year) and actors involved (named persons, as well as organizations). We developed a set of pattern matching rules allowing us to automatically identify dates and locations of events.

Example Consider for instance that the information to be collected consists of reports about urban demonstrations. R_1: "*200 étudiants manifestant contre la réforme de la loi sur l'éducation ont affronté les forces de l'ordre sur les Champs Elysées, le mercredi 21 mai, 14h00*" and R_2 : "*Le 21 mai à 15h00, la manifestation des étudiants a pris fin Place de l'Etoile, après une longue marche silencieuse*". Named entity extraction identifies "*Champs Elysées*", and, respectively, "*Place de l'Etoile*". During this phase, "*21 Mai*" will also be identified as the day the manifestation took place.

Semantic annotation

Semantic annotation is about assigning entities or, more generally, information items identified within texts, to their semantic description, as provided by an existing model. Annotation provides additional information about text, so that deeper analysis on its content can be made.

Different techniques and tools of semantic annotation are available. They can be entirely manual: the user himself associates annotations with elements to be annotated, as is the case in [16]. On the other hand, entirely automatic annotation techniques associate annotations with elements to be annotated. By using a set of learned patterns, or an ontology [7],[22]. In between, semi-automatic techniques allow the user to associate annotations with elements to be annotated, by choosing, validating or rejecting annotations proposed by the system [11]. For instance, named entities can be considered as instances of concepts of an existing ontology (ex. Paris is an instance of the concept "city"), therefore it becomes possible to enrich every information item by making explicit relations between named entities previously identified and concepts of an existing ontology.

For this work, a semantic annotation of i is a tuple $(V_i^1,..., V_i^m)$, where each element V_i^k is an instance of some ontological concept. The output of the annotation phase is a set of information $\{I_1 \ldots I_n\}$ with its respective annotation $\{i_1 \ldots i_n\}$, which are instances of a common ontology.

Example In our case, each report will be annotated by a triplet, corresponding to three main concepts of the considered ontology. For instance (*Champs-Elysées*, 05-21-14, Student) annotates the first report, while (*Etoile*, 05-21-14(14h00), Student) annotates the second.

Semantic similarity estimation

Given two annotations $i = (V_i^1,..., V_i^m)$, and $j = (V_j^1,..., V_j^m)$, the degree of semantic similarity between *i* and *j* is defined by:

$$s(i, J) = s_1\left(V_i^1, V_j^1\right) \oplus \ldots \oplus s_m\left(V_i^m, V_j^m\right)$$

where the s_k are some functions of similarity on the classes of the k^{th} values and $\oplus$ is a given aggregation function.

Example Let us take again the (where, when, who) annotations. Here, the three functions s_1, s_2 and s_3 respectively define the similarity between places, dates and people.

Assume that these functions are such that:

- s_1(*Etoile, Champs Elysées*) = .99
- s_2(05-21-14, 05-21-15) = .99 and
- s_3(student, student) = 1

If function $\oplus$ is such .99 $\oplus$ .99 $\oplus$ 1 = .99, then s((*Etoile*, 05-21-14, students),(*Champs Elysées*, 05-21-15, students)) = .99.

This means that according to these different functions, a report relative to a demonstration of students near " *Etoile* " on " May 21st " at "14PM " and a report relative to a demonstration of students near " *Champs Elysées* " on " May 21st " at "15PM " have very high semantic similarity.

A semi-automatic approach to evaluate information

Our objective is that the operator is required to give only the degrees of correlation of items of information I_k, $I_{k'}$ which are ontologically close, i.e., such that $s(I_k, I_{k'})$ is greater than a given threshold γ.
This leads to the following algorithm:

- For any $k = 1... n$ $\alpha_{I_k,I_k} \leftarrow 1$
- For any $k = 1... n$, for any $k' = 1... n, k' \neq k \Rightarrow \alpha_{I_k,I_{k'}} \leftarrow 0$
- For any $k = 1... n$, for any $k' = 1... n, k' \neq k \Rightarrow$ the semantic similarity $s(I_k, I_{k'})$ is computed.
- If $s(I_k, I_{k'}) > \gamma$ then the information I_k and $I_{k'}$ are transferred to the operator in order he estimates their degrees of correlation $\alpha_{I_k,I_{k'}}$ and $\alpha_{I_{k'},I_k}$

Two items of information relative to the same place, the same date and the same persons are ontologically close, but they may contradict or confirm each other, while two items of information that are ontologically distant will maintain a null degree of correlation.

Notice that the evaluation of the degrees of correlation by the human operator is necessary, since two items of information that are semantically close do not necessarily confirm each other. For instance, a report stating " the demonstration has been followed by a huge number of students near *Etoile* on May 21st at 14PM " and the information "only few students in the streets nearby *Champs Elysées* on May 21st at 14PM " are ontologically close, but they are contradictory. However, the use of a finer ontology, allowing richer semantic annotations, makes it possible to emphasize the contradiction between those items of information (due to the presence of " few students " and " a huge number of students ", which are contradictory quantifiers).

Discussion

The question that we address here is now: does the information evaluation model described in this paper agree with the informal requirements of NATO?

We aim to prove that this model agrees with these, by showing that it takes into account the three main notions that underline these requirements, which are the number of independent sources that support the information, their reliability and the fact that items of information are contradictory or tend to be contradictory.

- The previous model obviously takes into account the number of independent sources that support an item of information and their reliability. More precisely, the more supported an item of information is and the more reliable its sources are, then the higher its evaluation is. Indeed, for a given k, let us denote:

$$S_k^1 = \{k'=1... n, \ \alpha_{I_k,I_{k'}} > 0\} \text{ and } S_k^2 = \{k'=1... n, \ \alpha_{I_k,I_{k'}} < 0\}$$

Thus, we can write:

$$v_k^* = A + B\left(\sum_{k' \in S_k^1} v_{k'} \alpha_{I_k,I_{k'}}\right) + B\left(\sum_{k' \in S_k^2} v_{k'} \alpha_{I_k,I_{k'}}\right)$$

where A and B are constants, that permit the following properties to be exhibited.

Proposition 2

1. If the number of sources that support information I_k increases, then $\sum_{k' \in S_k^1} v_{k'}$ increases. Thus, v_k^* increases.

2. If the degrees of reliability of the sources that support Ik increase, then $\sum_{k' \in S_k^1} v_{k'}$ increases. Thus, v_k^* increases.

- The information evaluation model previously defined takes into account the fact that items of information are contradictory, or tend to contradict each other.

Indeed, we can obviously define a notion of degree of conflict from the notion of degree of correlation. Let I_1 and I_2 be two different items of information. Their degree of conflict, noted by $c(I_1, I_2)$, can be defined by: $c(I_1, I_2) = -\alpha_{I_1,I_2}$. Notice that $c(I_1, I_2) \in [-1, +1]$.

Proposition3

- $c(I, I) = -1$
- $c(I_1, I_2) > 0$ iff I_1 tends to contradict I_2.
- $c(I_1, I_2) < 0$ iff I_1 tends to confirm I_2.
- $c(I_1, I_2) = 0$ else.

Conclusions and future work

In this paper we tackled the problem of information evaluation for intelligences purposes, from a military specific point of view. We considered the evaluation of complex information, such as natural language reports. We defined the general architecture of an evaluation system, based on a basic treatment cell called an evaluator. We also addressed semantic aspects and showed how an ontology can be used to annotate information items and to define a semantic similarity degree between them. We claimed that the operator of an evaluator must be required to examine items of information only when their degree of similarity is over a given threshold. In this case, the operator has to assess their degree of correlation. In the model that we defined, the overall evaluation of information has two ingredients: the correlation degree of a particular information item with respect to other information items under analysis and the reliability of its source.

Finally, we showed that this model is compliant with respect to informal requirements for information evaluation, as expressed by NATO, in the sense that it takes into account the main notions underlying those recommendations.

It must be emphasized that the evaluation values computed according to the process that we defined depends strongly on the ontology that is considered and on the different constants and functions mentioned in § "Semi Automatic Approach" such as: the similarity functions, the aggregation function and the value of the threshold. Indeed, the ontology and these functions and constants are used to relate semantically similar reports or, equivalently, to discriminate non-similar reports.

Of course, the evaluation values computed according to this process also depend on the user, who is required to estimate the correlation degrees between similar reports.

Notice that the fact that our process is semi-automatic implies that the classic evaluation methods (benchmarking with recall and precision measures) are not suitable to validate it. To measure the real advantages offered by our system, we must measure the time that is necessary for an operator to calculate the correlation degrees of items of information, with and without the help of the system, and compare them.

The implementation of this general process is under development and will lead to an experimental validation of the ideas. However, as mentioned previously, one question concerns the choice of the different constants and functions used in the process. Currently, it is difficult to estimate how the overall information evaluation varies according to the choices of these functions and thresholds. One future work direction is to study the impact of such choices on the resulting evaluation.

Another direction concerns the improvement of semantic similarity estimation, by also taking into account ontology concepts, since for now only their instances are considered. For this purpose, enhanced information extraction treatments are needed, to retrieve concepts within texts. Once identified, these concepts can be a part of the semantic annotation and they can be used to define more accurate measures for semantic similarity.

Since natural language reports are complex data, the main limitation of our approach is related to textual data processing. Therefore, linguistic phenomena such as negations are not taken into account during the automatic processing of our data. Instead, user intervention can easily identify them and asses or reject the correlation of information items. From an applicative point of view, the notion of contradictory information items is not addressed by automatic procedures, since the role of the user is to identify such contradictions within a collection of correlated information items. Hence, adopting this user-centric approach to evaluate information allows us to accomplish a satisfactory level, in terms of information quality ■

Acknowledgements

This work has been supported by Onera, by ANR (project CAHORS) and by the DGA (project QUALIPSO).

References

[1] BALAHUR, M. KABADJOV, J. STEINBERGER - *Exploiting High-Level Semantic Information for the Opinion Oriented Summarization of blogs*. Vol. 1, Issues 1-2, International Journal of Computational Linguistic and Applications, 2010
[2] J. BESOMBES, A. REVAULT D'ALLONNES - *An Extension of STANAG2022 for Information Scoring*. Proceedings of the 11th International Conference on Information Fusion (FUSION 2008)
[3] BIZER, R. CYGANIAK - *Quality-Driven Information Filtering Using the WIQA Policy Framework*. Journal of Web Semantics, 2009
[4] L. CHOLVY - *Information Evaluation in Fusion: a Case Study*. "Proceedings of ECSQARU-03. Workshop "Uncertainty, Incompleteness, Imprecision and Conflict in Multiple Data Sources", 2003
[5] L. CHOLVY, V. NIMIER - *Information Evaluation: discussion about STANAG 2022 Recommendations*. Of the RTO IST Symposium "Military Data and Information Fusion", October 2003
[6] L. CHOLVY - *Modelling Information Evaluation in Fusion*. Proceedings of the 10th International Conference on Information Fusion (FUSION 2007)
[7] DINGLI, F. CIRAVEGNA, Y. WILKS - *Automatic Semantic Annotation Using Unsupervised Information Extraction and Integration.* Proceedings of the K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation.
[8] FUSION, http://isif.org/
[9] T. GRUBER - *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. International Journal of Human-Computer Studies Vol. 43, Issues 5-6, 1995
[10] GUREVYCH, R. MALAKA, R. PORZEL, H.P. ZORN - *Semantic Coherence Scoring Using an Ontology*. Proceedings of the HLT-NAACL Conference, 2003.
[11] S. HANDSCHUCH, S. STAAB, A. MAEDCHE - *CREAting Relational Metadata (CREAM), a Framework for Semantic Annotation*. The Emerging Semantic Web, IOS Press, 2002
[12] S. HANDSCHUH, S. STAAB, F. CIRAVEGNA, S-CREAM - *Semi-Automatic CREAtion of Metadata*. Proceedings of the 13th International Conference on Knowledge Engineering Management (EKAW 2002)
[13] JILANI, N. GRABAR, P. MENETON, M.C. JAULENT - *Assessment of Biomedical Knowledge According to Confidence Criteria*. Proceedings of MIE2008, The XXIst International Congress of the European Federation for Medical Informatics, 2008
[14] J. JURAN - *The Quality Control.* Handbook 3rd Edition, McGraw-Hill, 1974
[15] V. KARKALETSIS, A. PEREGO, P. ARCHER, K. STAMATAKIS, P. NASIKAS, D. ROSE - *Quality Labeling of Web Content : The Quatro Approach*. MTW'06
[16] M.R. KOIVUNEN, R. SWICK, E. PRUD'HOMMEAUX - *Annotea Shared Bookmarks*. Proceedings of the KCAP 2003 Workshop on Knowledge Markup and Semantic Annotation
[17] S. KONIECZNY, R. PINO PEREZ - *On the Logic of Merging*. Proceedings of KR'98

[18] NATO, http://www.rto.nato.int/
[19] NATO, Intelligence Reports, STANAG2511, January 2003
[20] NATO Standardization Agency (NSA), AAP-6-2006, 2006, www.nato.int/docu/stanag/aap006/AAP-6-2006
[21] V. NIMIER - *Evaluation: A Formalization of Operational Recommendations*. Proceedings of the 7th International Conference on Information Fusion (FUSION 2004)
[22] POPOV and AL. - *Semantic Annotation Platform*. Second International Semantic Web Conference (ISWC 2003)
[23] G. L. RAGOVA, E. BOSSÉ - *Information Quality in Information Fusion*. Proceedings of the 13th International Conference on Information Fusion (FUSION 2010)
[24] B. ULICNY, K. BACLAWSKI - *New Metrics for Newsblog Credibility*. Proceedings International Conference on Weblogs and Social MediaColorado, 2010
[25] B. ULICNY, C. J. MATHEUS, G.M. POWELL, M. M. KOKAR - *Current Approaches to Automated Information Evaluation and their Applicability to Priority Intelligence Requirement Answering*. Proceedings of 13th International Conference on Information Fusion (FUSION 2010).
[26] B. ULICNY, G.M. POWELL, C. BROWN, M. M. KOKAR C. J. MATHEUS, J. LETKOWSKI - *The Analyst via Situation-Dependent Reasoning with Trust-Annotated Facts*. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011
[27] Y. WAND, R.Y. WANG - *Anchoring Data Quality Dimensions in Ontological Foundations*. ACM 39, 11, 1996
[28] R. Y. WANG, D. M. STRONG - *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems, 1996
[29] B. WU, V. GOEL, B. DAVISON - *Propagating Trust and Distrust to Demote Web Spam*, MTW'06
[30] B. WU, V. GOEL, B. DAVISON - *Topical TrustRank: Using Topicality to Combat Web Spam*, WWW'06
[31] S. YOUNG, J. PALMER - *Pedigree and Confidence: Issues in Data Credibility and Reliability*. Proceedings of 10th International Conference on Information Fusion (FUSION 2007)

AUTHORS

Jérôme Besombes received his PhD degree from the University Henri Poincaré of Nancy in 2003. He has been a research engineer at Onera since 2006. His research interests include knowledge modeling, information scoring, automatic reasoning, and visual analytics.

Laurence Cholvy is Research Director at Onera. She promotes the use of logics (classical and modal logics) and Belief Function Theory, in particular in the field of Information Fusion. In this context, she has contributed to information evaluation, reasoning with inconsistencies and with uncertainty.

Valentina Dragos received her Master and Ph.D. degrees, both in Computer Science, from the René Descartes University. Her research interests include issues related to artificial intelligence, knowledge-based systems and semantic technologies. Information scoring and high level information fusion constitute the main application field of its research