PROJECT 6 - Common Basic Knowledge for Artificial Intelligence Algorithms

Completed on 05-Jul-2015 (19 days)

Introduction >

Motivation

Artificial intelligence is gradually becoming the defining feature in a big proportion of new products. Just this fact is enough to support the current project, where I am proposing a comprehensive framework aiming to synchronise the huge variety of AI attempts. Note that, equivalently to what has been happening in most of relevant (and complex enough) break-troughs, the advancements in this field have mainly resulted from isolated (or, at least, tailored to solve narrowly-delimited problems) contributions, which were rarely in a position to apply ideas on the lines of "let our work be the continuation of that previous attempt".

Furthermore, some weeks ago I watched a video which can be considered as the ultimate responsible for this project. It was a documentary about IBM's Watson; a rough overview of its development, the intermediate tests and its final triumph against the best Jeopardy! players. Bear in mind that the target performance of this system (i.e., playing Jeopardy! as well as an experienced person in a completely autonomous way) is so demanding that its development can easily be taken as an absolute reference of nowadays' best way to face any situation involving (English) language understanding.

In that video, there was a specific part which I found particularly interesting. The narrator was describing how the developers realised about the need of relying on machine learning in detriment of more casuistic approaches (i.e., huge lists including all the possible alternatives for each case), on account of the big complexity of the analysed problems. I found the holy-grailish way in which the video transmitted these ideas very compatible with the last public-domain trends on this front; but also, and equivalently to what happens with most of public-domain trends on specialised issues, highly inaccurate. That is: even in case of understanding "machine learning" as a well-defined approach accurately outputting the expected results (what is something very far away from the reality), suitable inputs would have to be provided and AI algorithms can only take as inputs raw data describing the modelled behaviour (human communication, in this case) in the aforementioned casuistic form. Thus, by associating this bipartition with the human learning process, "machine learning" would be the equivalent to the memorising/understanding/adaptation/etc. capabilities and "casuistic hardcoding" to all the knowledge received by the given person (e.g., sensorial information, basic concepts, formal education and so on).

The main goal of this project is proposing a comprehensive methodology to facilitate the synchronisation of the aforementioned "basic inputs" among all the AI-based approaches; or, by relying on the closest human-learning analogy: a preliminary proposal for the creation of a

global education plan, which might allow any AI approach to learn from others/share its knowledge.

Technical Background

The remaining projects in varocarbas.com and the "about me" page should provide a good general impression about my background and attitude. Nonetheless, bear in mind that I consider myself as a (knowledgeable) partial outsider to the exact area of expertise underlying this whole project.

During over the last 7 years, I have been working as a programmer and numerical modeller with a major focus on innovative and from-scratch developments. In fact and as a direct consequence of my self-imposed hard conditions (i.e., always aiming for the highest goals despite my limited availability of resources, support and recognition), I am sure that my understanding in certain subfields is even deeper than usual among persons who have followed a more standardised and protective learning process within the specialised research world (e.g., PhD/PostDoc or full-time-at-big-companies researchers). That is: my current heterogeneous knowledge is the result of having combined my formal (heterogeneous) education, an unorthodox R&D-focused professional activity and my eminently practical attitude. On the other hand, I do recognise my limited experience in very large projects of any kind; both from a pure theoretical perspective (e.g., very big implementations) and from the point of view of coordinating a complex group of isolated elements (e.g., project involving many persons with different interests).

I am in a position to share some of my ideas because of my relevant knowledge and expertise in this field; but always by bearing in mind my low experience in the academic (and/or top-funded) research and large projects. Nevertheless, this last limitation is perhaps one of the most attractive features of the current project: written by a person who is not completely aware about the exact difficulties justifying the applicability to this specific situation of the statement "why what should be can rarely be?".

In summary, I hope that my theoretical & practical background together with my (self-assessed) good summarising skills will allow me to write some worthy ideas, which might eventually be taken as fresh-breeze insights by the true experts when having to face the development of a system on these lines.

Project Scope

The main goal of this project is to propose some basic guidelines which are expected to be considered by those working on the systematisation of input information (e.g., training datasets) used by AI-based approaches. That is: a proposal (by bearing in mind the limitations explained in Technical Background) aiming to contribute to a future data-standardisation process, which is likely to occur within the medium term in any case.

The following issues are outside the scope of this project:

- A well-defined and overall-accounting system.
- Detailed answers fully addressing certain subproblems.
- References to parts which are beyond my core expertise (e.g., human-like facial expressions or moves).
- Unrealistic or Sci-Fi references (e.g., Asimov's Laws).

Basic Ideas >

Defining Principles

The preliminary character of the proposed system should never be misunderstood as having an ambiguously-defined essence; in fact, this project (and my personal impressions on this front) might almost be defined as of the take-it-or-leave-it type.

The main goal of all the ideas in the current project is to propose (the preliminary structure of) a system which is expected to always verify the following principles:

- Adaptable & Scalable. The reliability of a system in charge of standardising a so complex and variable reality is highly conditioned by its ability to adequately address any unexpected issue.
- Collaborative. The intended systematisation can only be accomplished in case of having a relevant enough amount of resources. In principle, the most logical way to collect all these resources is by accepting (voluntary) contributions of organisations which have developed relevant AI-focused projects.
- Overall Compatible. Curiously, a big proportion of universally-utilised systems do not aim to create highly compatible outputs, but rather show monopolistic behaviours on the lines of "you have to do everything my way". This principle is important because of the aforementioned absolute dependence upon voluntary collaborations, which are more likely to occur in case of having an attractive enough structure.
- Interdisciplinary. Computing and data standardisation will undoubtedly be the most relevant fields of expertise, but not the only ones. Different areas of knowledge will also have to be considered in order to give a comprehensive answer to so peculiar situations (i.e., machines of any kind running almost autonomously). For example: non-technical specialists to analyse the ethical or political implications of certain issues.
- Monitored. Any standardised (data) format expected to have general applicability has to be controlled by an authority ensuring its adequacy. In this specific situation, such a generic requirement is even more important on account of two issues: on one hand, Artificial Intelligence is already a very sensitive matter (which is even likely to become exponentially more sensitive within the medium term); on the other hand, the

proposed open-to-collaboration format reinforces the necessity of a monitoring instance.
All the monitoring-related issues are analysed in the Supervision section. Nevertheless, I can already anticipate that these tasks are expected to be managed by an international organisation, with a national-government level support in each country.

Main Structure

The proposed system might be considered as a huge multi-layered classification of AI information, where each input is defined on account of its adequacy with respect to a set of categorised features. Thus, the best way to adequately describe the main structure of this system is analysing these (recursive) classification processes.

For clarity purposes, I will define two different classification types:

- Multi-layer Classification. Each element is individually defined on account of its association with different categories. This definition is performed in a layered fashion, where Layers are understood as ranked higher-level categories which are similar to enumerations/enums in a programming language. Such a ranking is expected to be based upon rules like *"Layer 0 will always be weight_0-1 times more important than Layer 1"* (equivalently to what is expected to also happen with all the other rankings, which will be analysed in the next section). For example: a car can be classified as *vehicle* (within the *Layer Machines*), *mobile* (within the *Layer Staticity*), *contaminant* (within the *Layer Environment*), etc.
- Inter-relationship Classification. Additionally to the aforementioned individual definition of each single element, relationships among different elements have also to be considered. That is: the previous classification can be associated with defining a word, and the current one with defining a concept. For example: according to this classification, a car can also be defined as *"faster than a bicycle"* and *"slower than a plane"*.
  As far as this second classification cannot be properly understood without knowing the information storage format, I will continue this analysis in the next section.

As already explained, a detailed description of the proposed system is outside the scope of this project. On the other hand, more specific indications seem required in order to adequately transmit the intended ideas; that's why I am including my preliminary thoughts about the highest-level Layers below these lines.

- *Functionality Layer (L0)*. The information is classified on account of its functionality. Logically, a comprehensive enough list of major functionalities will have to be created (equivalently to what is expected to happen while defining all the other Layers). Sample functionalities: *understanding*, *movement* or *vision*.
- *Speciality Layer (L1)*. This second division is based upon the application scope of the given information; that is: useful for any kind of algorithm (e.g., differences between men and women) or only for certain ones (e.g., ability to recognise a very specific

material). This Layer is expected to be formed by a common category and *n* additional specialised ones. Note that this common sub-Layer will get a special treatment on quite a few fronts, like security-related issues.

- *Security Layer (L2)*. It accounts for the impact of (the eventual variation of) the given piece of information on other elements in the system. For example: the *masculine*/*feminine* differentiation has a much more relevant impact (i.e., a higher number of additional elements will be affected) than the *bulldog*/*boxer* one.

Data Format

The proposed system is basically a big layered classification of AI (training) data, whose definition was started in the previous section by explaining the main category (i.e., Layers, associated with individual elements, as opposed to concepts). In the current section, I will be completing this definition with a detailed description of the main constituent elements of the system.

All the information accounted by this system is expected to be defined on account of the following basic elements:

- Property. Bit of information (e.g., a word) suitable to be taken as a reference while comparing some other elements. Examples: *speed* (and associated adjectives like *fast* or *slow*), *size* (and associated adjectives like *big* or *small*) and so on.
  Properties are ranked (Ranking P) and can be defined individually (what, by bearing in mind their essence, is the most likely scenario) or on account of other elements. When being added to the system, each Property is ranked either manually (i.e., when being self-defined) or automatically (i.e., by bringing into account the rankings of its defining elements).

- Connector. Complementary element required by the Properties to perform a given comparison. Examples: *more than*, *less than*, *similar to*, etc.
  Connectors are always self-defined and unranked.

- Entity. It is the minimum unit of information accounted by this system which is externally relevant. It might also be understood as an individual object (e.g., *car*) as opposed to a concept (e.g., *this car is fast*).
  Entities are automatically ranked (Ranking E) on account of their defining elements. The definition of each Entity is always based upon one or more Layers (e.g., the definition of car being based upon the Layers *Staticity (L100)* and *Machines (L101)*); additionally, it might be defined on account of one or more Properties (e.g., car also defined on account of its danger to *people (P0)* and *speed (P1)*).

- Concept. It matches the most commonly accepted definition of concept, by bearing in mind that it will always be formed by the aforementioned elements (or/and by other Concepts).
  Equivalently to what happens with Entities, Concepts are automatically ranked on account of their defining elements. Nevertheless, this ranking has to be redefined for each single Entity, because the importance of certain concept might vary depending

upon the given context. For example: the Concept *"money is very important"* would have a higher ranking than *"helping others is very important"* when analysing the Entity *"corporation"*, but not the Entity *"non-profit"*. Thus, Ranking C actually refers to a group of as many different rankings as suitable Entities (i.e., the ones having at least two associated Concepts).

A Concept may be created from just two Entities, one Property and one Connector; but might also be built on as many Entities (+ Properties & Connectors) or other Concepts (+ Properties & Connectors) as required.

By putting together all the ideas explained in this section and in the previous one, the (ranked) layered categorisation which underlies the proposed system is formed by the following parts:

- Two groups of elements which are ordered according to two different rankings and are externally relevant: Entities (Ranking E) & Concepts (Ranking C, which is redefined as many times as Entities with associated Concepts).
- Two additional groups of elements which are accessorily used by the two ones in the previous point: Properties (Ranking P) & Connectors (unranked).
- On top of all these rankings, there is another one (Ranking L) formed by enum-like elements called Layers, which defines the two aforementioned externally-relevant elements/rankings: Entities & Ranking E in a direct way; and Concepts & Ranking C indirectly (on account of the fact that Concepts are always built on Entities).

  Note that the words used to define a Layer or its (enum-)elements may also be treated as Entities or Properties. In fact, Layers are not elements of this system in a strict sense (i.e., at the same level than all the ones defined in this section); they are just a way to facilitate the understanding and usage of the proposed data format.

Case Study

The not-exhaustive essence of this system shouldn't hide the intrinsic complexity of some of its ideas. This section has precisely to be understood as an additional support to help readers get a clearer picture; mainly about what was explained in the previous two sections (Main Structure & Data Format). That's why these contents should never be taken as a step-by-step guide or a detailed example.

The table below shows (a very simplistic version of) the kind of information which the proposed system would have to bring into consideration in order to account for: *"a car is faster than a bicycle"* and *"a car is more dangerous to people than a bicycle"*.

Legend: car/bicycle | both of them

Layers	Selected Enum Option	Properties	Property Contents	Connectors
L0 Function	understanding	P0 Dangerous to people	*0 (not harmful) - 100* 70/20	more than
L1 Speciality	common	P1 Speed	*m/s* minimum:0/0 maximum:16/120 typical:9/28	less than
L2 Security	*0 (no impact) - 100* 70/40			
L100 Staticity	mobile			
L101 Machines	vehicle			
L102 Environment	*0 (very harmful) - 100* 50/100			

CLARIFICATIONS:

- In a real scenario, much more information would have to be considered, even in case of only wanting to store the aforementioned two concepts. For example: additional Layers would be required and all of them would have to be properly defined; same ideas (i.e., additional ones and proper definitions) would have to be applied to Properties; just two Entities is not enough to adequately account for car and bicycle (i.e., there are many different types of cars and bicycles); etc.
- In application of the defining principles of this system, the exact format of each element has to be as variable (scalable and adaptable) as possible. For example, the Properties (and even the Layers) are applied by relying on a 0-100 scale (whose values are automatically calculated or manually input) or with certain descriptive values (e.g., maximum, minimum and most typical), as shown in the table. On the other hand, it has to be beard in mind that any other suitable alternative (i.e., fully compatible with the remaining parts of the system) might also be considered.
- Although the rankings may not be modified (except Ranking C, which is redefined on account of the Entity being considered), the scale-based definitions should be adapted to the exact conditions in order to accurately describe each situation. For example: the *"a car is faster than a bicycle"* Concept can be assessed by giving higher preference to the elements which are not harmful for the environment (*L102* as currently displayed) or right the contrary (by defining *0* in *L102* as *not harmful*).
- There are certain parts which may not have the aforementioned flexibility. For example, the first layers (i.e., *L0*, *L1* and *L2* in the table. Note that the number of these Layers is expected to be much bigger in a real scenario) will not be altered under any circumstance.

Additional Issues >

Maintenance

As explained in a previous section, I have a limited experience regarding major aspects associated with this project; that's why it should be taken as a set of preliminary ideas. Such a statement makes even more sense in the current section (and in the next one, Supervision), although including a short reference to some issues is certainly required.

According to its defining principles, most of the information required by this system is expected to be provided through voluntary contributions. The main reason for such a format is to assume that, due to the associated difficulties, collecting the required information is much more likely to occur within small-scope R&D-but-for-profit projects. This peculiarity increases notably the maintenance costs of the whole system; mainly because of the additional expenses associated with data-integrity-assurance tasks, as explained in the next section.

The remaining aspects associated with the maintenance of the proposed system (e.g., software & hardware costs, buildings, staff, etc.) will be intimately related to the selection of the managing organisation(s). Such a determination is also outside my core expertise (and of the scope of this project), but some (national) government involvement (e.g., a national agency in each country; at the international level, an organisation supported by various countries) seems certainly required. In a first moment, this whole system might be focused on the non-profit or educational world (e.g., research groups at universities sharing data sets by relying on this standard); although such a configuration would stop working as soon as more egoist interests (i.e., big companies, probably taking care of most of the AI R&D work because of its huge benefit potential) would come into picture.

Supervision

As highlighted in the previous section, the specific issues here are completely outside my core expertise. Nevertheless, this project wouldn't be completed without mentioning certain aspects.

One of the defining principles states that most of its information is expected to be provided by voluntary contributions; this approach is assumed to output a much higher growth than what a self-sustained configuration would do. Note that relevant efforts (and costs) are associated with building reliable enough training data sets (by following the proposed format or any other one); and spending so important resources for performing almost-never-ending tasks to build a data source (which will eventually be used by someone else) doesn't sound too likely to happen. On the other hand, more egoist attempts with much smaller scopes (e.g., innovative AI approach accounting for certain situation) will be accessorily generating data sets, which might be shared in application of ideas like "helping to create what I cannot do myself".

The aforementioned (forced) collaborative essence of this system is expected to represent the source of most-difficult-to-be-addressed problems: all what is associated with ensuring data integrity. These issues are so important that even the basic structure of the system (as described in the sections Main Structure and Data Format) has been defined such that data integrity is ensured (e.g., within the proposed first Layers, *L2* or the *common* sub-Layer of *L1*).

Note that the importance of making sure that the system remains reliable despite the (eventually-faulty) external contributions can be analysed at different levels. On one hand, any standardised format expected to be used (almost blindly) by many people has to be reliable and trustworthy. On the other hand, some AI-based algorithms deal with highly-sensitive information (to not mention issues like being potentially dangerous to people or related to national security). These facts reinforce the proposed (e.g., in the Defining Principles and Maintenance sections) governmental involvement as one of the basic requirements of this system.