Reconstructing True Wrong Inductions

Jean-Gabriel Ganascia

LIP6 - Université Pierre et Marie Curie (Paris VI)
8, rue du Capitaine Scott
75015, Paris, FRANCE
Jean-Gabriel.Ganascia@lip6.fr

Abstract

Many pre-scientific and common sense inductions are erroneous. Here are presented attempts to simulate that kind of inductive reasoning. Our hypothesis is that mistaken inductions are not only due to the lack of facts, but also to the poor description of existing facts and to implicit knowledge. We present a few experiments which aim at validating this hypothesis by simulating with machine learning and data mining techniques the way people erect erroneous theories from observations.

Why True Wrong Inductions?

Our aim here is to rebuild wrong inductions with machine learning techniques. This goal may seem both odd and trivial; indeed, all induced theories that are not true can be considered as false. Therefore, one could have the impression that it is easy to induce wrong theories, since it is only to generate arbitrary theories consistent with observed data and to prove that they are not true, which is usually not difficult. Moreover, scientists and logicians, fond of truth, will feel it strange to be guided by the study of wrongness, errors and falsity. Nevertheless, we pretend that studying erroneous and mistaken theories is neither bizarre nor trivial. More precisely, we are not interested in all incorrect inductions: we focus our study on the reconstruction of old inductive theories, i. e. those that have, at least one time point in the past, been recognized as possibly true. In other words, we are concerned by wrong theories that people had in mind, and which can be characterized as real or "true" wrongness.

Indeed, many empirical theories, recognized today as wrong, such as the theory of "caloric" or the theory of "ether" in ancient physics, had convinced clever people in the past. We might as well imagine that most of our present scientific knowledge might be considered as erroneous in the future. Additionally, "common sense" knowledge is frequently incorrect, even if it seems evident. In a word, many currently accepted conceptions might or will be proved to be false.

The origin of errors is partly due to the lack of information; when almost nobody experiences some facts, theoretical consequences of those facts cannot be perceived. Most of the time, the *state of the art* is responsible, because it renders observations difficult or impossible. For instance, in the 17th century, the development of optics allows Galileo to gather some observations in astronomy that were not accessible before.

However, even while it is possible to derive a correct theory from a set of empirical evidences, it may happen that only erroneous theories are accepted as true. We shall try to understand and to explain this strange phenomenon in this paper. For this purpose, we shall provide some examples drawn from medicine and common sense reasoning, even if it is also the case in other scientific disciplines, e.g. in geology or in physics.

In order to simulate the way people thought and erected wrong theories from facts, we shall automatically reconstruct, with the help of computers, this pathway (leading from the data to the formation of erroneous theory), by using artificial intelligence techniques, such as, machine learning and data mining tools.

The first reason why we are interested in such a study is that it is of cognitive significance to note and understand how people actually derived general statements from facts, and not only to consider how they should do it. In the future, we could envisage many developments in cognitive psychology to test the validity of our model. At the present time, we have chosen to deal with pre-scientific knowledge, trying to explain why some misconceptions dominated the world for centuries, even though it was possible to derive more efficient theories than the dominating ones. So, our work is of epistemological interest.

But, we have also in mind the way people – not only the scientists – speculate from facts. This simulation of inexact reasoning could have many applications in social sciences, where it could help to understand the social representations, their evolutions and the way they spread. Finally, it may also enlighten some rhetorical strategies currently used by politicians who prefer to provide well-chosen examples, in spite of demonstration, to convince.

This paper is an attempt to model the way misconceptions emerge from facts with machine-learning techniques that simulate induction, i.e. reasoning from facts to general statements. The key concept is the notion of *explanatory*

power with which all conflicting theories will be compared: the explanatory power evaluates the number of observations that could be explained by a given theory, so each of the different theories generated by an inductive engine will be ranked with respect to this index.
The first part of the paper will describe the general framework. Then, we shall show the first model based on the use of supervised learning techniques. The two following parts will provide two examples of rational reconstruction of wrong medical theories using our first model. The first example tackles with misconceptions on the causes of scurvy disease, the second with misconceptions on the transmission of leprosy. Then, we shall consider an application to social sciences, more precisely to model the political beliefs in France, at the end of the 19th century, a few months before the Dreyfus affair burst. We now extend our model with a new induction engine based on the notion of *default generalization* inspired from the default logic theory (Reiter 1980) and using non-supervised learning techniques. The last part of the paper is dedicated to this new model.

General Framework

Since we focus our interest on rational reconstruction of inductive reasoning, i.e. in the derivation of general knowledge from facts, we shall take into consideration inductive machine learning techniques, i.e. those supervised or non-supervised techniques that simulate induction. Both supervised and non-supervised learning can be used for our purpose, which is to generate theories from facts. Each has its own advantages and disadvantages. On the one hand, supervised learning procedures are more efficient and easier to program, on the other hand, they require, from the user, to associate a label to each example, which is not always possible as we shall see in the following. In the first part of the paper, we restrict us to supervised techniques, but, in the second, we extend our model to integrate non-supervised learning techniques.

Sources of induction

Whatever technique we use, a description language is always needed; sometimes, additional background knowledge is also necessary. Therefore, the generated theory depends on all this additional knowledge, which biases the learning procedure. In other words, there is no pure induction because the way facts are given to an inductive machine influences considerably the induced theory.
Moreover, many empirical correlations may be observed, which lead to many different possible theories. Since most of the machine learning programs aim at building efficient and complete (i.e. that recognize all the examples) recognition procedures, they tend to preclude most of the possible correlations, using some general criteria to prune and eliminate them. For instance, in case of TDIDT – Top-Down Induction of Decision Trees – *information entropy* is a very efficient heuristic making the generated decision tree quite small, decreasing the number of leaves. Nevertheless, our goal here is totally different: first we aim at generating all possible theories and then discriminating explanation patterns among those different generated theories, by using a criteria based on the notion of explanatory power. To summarize, being given a set of known facts, we shall build different learning sets, using different representation languages and different background knowledge. Then, for each representation language with additional background knowledge, we shall study the different generated theories by comparing them with the different systems of hypothesis given by people to explain the examples. The general schema presented in figure 1 offers an overview of our global model.

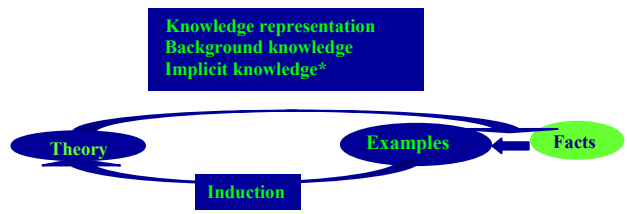

Figure 1: overview of our general model

In order to validate our model, we shall show how changing knowledge representation and background knowledge affects the generated theories. More precisely, it means to explain common sense reasoning by taking into account other implicit data, i.e. not only the given facts, but also the description language and all possible sources of associated knowledge. To support this thesis, we shall demonstrate many computer simulations where, by modifying the implicit knowledge, the "explanation power" of the different generated hypothesis will be modified, which means that, with respect to the notion of explanatory power, the respective ranking of each hypothesis generated by our inductive engine will be modified by the introduction of background knowledge, making artificially one more satisfying than the others.
As we already said, the key concept here is the notion of explanatory power drawn from (Thagard and Nowak 1990): it corresponds to the ratio of the learning set explained by a theory, i.e. to the number of examples belonging to the learning set which are covered by this theory. In other words, our inductive engine generates many conflicting theories that can be compared with respect to their explanatory power.
In case of supervised learning, an example E is said to be *covered* or *explained* by a theory T if and only if the label associated to the example, i.e. class(E), is automatically generated by the theory, which means T(E) = class(E). Then, $E_p(T)$ the explanatory power of the theory T is the number of examples belonging to the learning set that are covered by the theory T:

$$E_p(T) = \sum_{E \in \text{learning set}} \delta(T(E) = \text{class}(E))$$

where $\delta(\text{true}) = 1$ and $\delta(\text{false}) = 0$.

In case of non-supervised learning, there is no class *a priori* associated with examples, so the preceding definition cannot be in use. However, it is possible to compute the number of examples covered by each generated class. We can then introduce the notion of *cohesion* of a class, which, roughly speaking, corresponds to the sum of average similarities between the examples of a class. It follows that the explanation power of a set of classes is the sum of the cohesions of all classes. Therefore, higher the cohesion of generated classes is, higher is the explanation power.

Association rules

Our experiments make all use of association rules. These techniques, developed more than 15 years ago (Ganascia 1987), (Agrawal et al. 1993), became very popular with the emergence of data mining. Their goal is to detect frequent and useful patterns in databases. The main difference between the classical supervised learning techniques and inductive engines used in data mining processes is that in the former, the goal is to build an efficient classifier, i.e. a procedure that classifies consistently with the learning set, while, in the latter, it is to extract some remarkable patterns from the data.

As a consequence, an example may be covered by many extracted patterns, in data mining, while it is rarely the case in classical machine learning.

The basic step in building associated rules is the detection of correlations: if almost all the examples associated with a descriptor d are also associated with a description d', then it is possible to generate the rule **If** d **then** d'.

Without going into details, the main problem now is to extract the prominent patterns from huge data sets. To do this, it is necessary to enumerate many descriptions d without enumerating all of them.

An algebraic framework makes the systematization of the enumeration procedure possible. It is based on the notion of Galois connection (Ganascia 1993).

Discovering the cause of scurvy

Our first experiment was an attempt to discover the cause of scurvy and to understand why it took so long to realize that fresh fruits and vegetables could cure the disease.

Let us remember that, many people, more than hundred of thousands, especially in the navy, contracted the disease and perished in the past. There were many possible explanations for this, for instance a "physical explanation" connecting disease to a cold temperature or to humidity, a "physiological explanation" making the lack of food responsible, or even a "psychological explanation". However, until the beginning the 20th century, and the discovery of the role of vitamin C, physicians did not agree how to cure the disease, even when empirical evidence and clinical experiments confirmed the relation between the disease and the presence of fresh fruits and vegetables in the alimentary diet (Carpenter 1986).

We tempted to understand why it was not possible to induce the correct theory. We first consulted the 1880 *Dictionnaire Encyclopédique des Sciences Médicales* (Mahé 1880) which provides relatively precise description of 25 cases of scurvy, and we introduced those descriptions in our inductive engine (Corruble and Ganascia 1997). More precisely, we used a small description language derived from the natural language expressions employed in the medical encyclopedia to describe those 25 cases. This language contained the ten following attributes, *year, location, temperature, humidity, food-quantity, diet-variety, hygiene, type-of-location, fresh-fruit/vegetables, affection-severity*, each of them being affected by one or more values according to its type (integer, Boolean, string, ordered set, enumerated).

The 25 cases drawn from the medical encyclopedia were all described within this language. The attribute "affection-severity" quantified the evolution of the disease, which was of crucial interest since it determined the factors that had influenced the evolution. In our experiment, we restricted our induction engine to generate only rules concluding on this last attribute.

Once those rules have been induced, it was possible to distribute them into small subsets, according to the attributes present in their premises. For instance, the attribute diet-variety being present in the condition of rule R3 (cf. figure 2), it was possible to aggregate it to the "diet-variety" cluster. Each of those clusters corresponded to some explanation schema of the disease, since it was the set of rules concluding to the severity of the disease, which contained a given attribute. For instance, in case of the "diet-variety" set, it corresponded to the theory that explained the evolution of the disease with the "diet-variety". The figure 2 shows the rules generated from the 25 examples of the encyclopedia, classified according the attributes they contain in their premises.

Set I: Rules 3,4,8 use in their premises the variety of the diet.
R3: IF diet-variety ≥ high THEN disease-severity ≤ 0. [5]
R4: IF diet-variety ≤ average THEN disease-severity ≥ 3. [4]
R8: IF diet-variety ≥ average THEN disease-severity ≤ 2. [11]

Set II: Rules 7, 10 use in their premises the presence (or absence) of fresh fruits and vegetables in the diet.
R7: IF fresh_fruits/vegetables = no THEN disease-severity ≥ 2. [5]
R10: IF fresh_fruits/vegetables = yes THEN disease-severity ≤ 2. [13]

Set III: Rule 2 uses in its premises the quantity of food available.
R2: IF food-quantity ≥ ok THEN disease-severity ≤ 0. [4]

Set IV: Rules 5,6,9,12 use in their premises the level of hygiene.
R5: IF hygiene ≤ bad THEN disease-severity ≥ 3. [3]
R6: IF hygiene ≤ average THEN disease-severity ≥ 2. [4]
R9: IF hygiene ≥ average THEN disease-severity ≤ 2. [7]
R12: IF hygiene ≥ good THEN disease-severity ≤ 1. [6]

Set V: Rules 1, 11 use in their premises the temperature.
R1: IF location = land, temperature ≥ hot THEN disease-severity ≤ 0. [4]
R11: IF temperature ≤ severe-cold THEN disease-severity ≥ 1. [5]

Figure 2: rules generated without background knowledge

The results showed that the "best theory", i.e. the theory with the higher explanation power, was the set of rules that contained the attribute "fresh fruits and vegetable" in their premise.

Moreover, it was possible to compare the different explanations given in the encyclopedia with the explanatory schemata generated from the 25 cases given in the same encyclopedia. It appeared that each set of rules corresponded to some explanation given in the encyclopedia (Mahé 1880). Let us quote here the mention of those explanations:

Diet variety and fresh fruits and vegetables: "*It was J.F. Bachström (1734) who first expressed the opinion that, "Abstinence of vegetables is the only, the true, the first cause of scurvy.""*

Food quantity: "*We are lead to conclude that a decrease in quantity of food, or to speak clearly, starvation, can occasionally serve the cause of scurvy, but it cannot produce it by itself.*"

Hygiene: "*If Cook's crews were entirely spared from scurvy, in a relatively large extent considering the times, it is thought that these great results were precisely the happy consequence of the care given to the cleanliness and drying of the ships.*"

Temperature: "*Spring and winter are obviously the seasons of predominance for scurvy.*"

The explanation power ordered those four explanatory schemata in accordance to the preference expressed by the authors of the medical encyclopedia even if the theory considered as the most plausible explanation of the scurvy, i.e. the theory of humidity, did not appear at all in this list. This was because there was no direct correlation between the disease severity and the humidity. But, it appears that the humidity was the most currently accepted hypothesis. Here is the quotation of the encyclopedia that mentions the theory of humidity as the most plausible: *"The influence of a cold and humid atmosphere has been said to be the key factor for the apparition of scurvy. "Air humidity is the main predisposing cause of this disease", according to Lind." (Mahé 1880)*

In a sense, this first result was a good thing for artificial intelligence: it showed a machine able to induce the correct theory while people, with the same material, were not. However, it did not explain why, in the past, people adopted the humidity theory to explain the apparition and the evolution of scurvy. Because our goal is to model these kinds of wrong reasoning and the way people reason, we considered the result unsatisfactory by itself. Therefore, we tried to understand what biased their inductive ability. Then, we looked for some implicit medical theory that could influence induction. We found as a candidate "the blocked perspiration theory" that was prevalent in medical schools for centuries. This conception was based on the old theory of fluids introduced by Galien (131-201), during the 2nd century. According to this hypothesis, without excretions and perspiration, the internal body amasses humors, especially bad humors, which result from fluid corruption and cause diseases. Since humidity and bad hygiene tend to block up pores of skin, it makes perspiration difficult and consequently it leads to accumulation of bad humors. Furthermore, lack of fresh fruits and vegetables thicken internal humors, which render theirs excretions more difficult.

We translated this theory by using two new attributes and a few production rules which were introduced as background knowledge in our induction engine.

Then, in addition to the rules generated previously, the inductive engine induced five more rules (Cf. figure 3). Taking into account these rules, it appeared that the rules containing the attribute humidity constituted one of the possible explanatory schemata whose explanation power was higher than of other theories.

IF humidity ≥ high, fresh_fruits/vegetables = unknown THEN disease-severity ≥ 2. [4]
IF humidity ≤ high, hygiene ≥ average THEN disease-severity ≤ 1. [6]
IF perspiration ≤ hard THEN disease-severity ≤ 1. [6]
IF fluids ≥ corrupted THEN disease-severity ≥ 2. [9]
IF fluids ≤ healthy THEN disease-severity ≤ 2. [14]

Figure 3: New rules produced when the domain knowledge is given to the system

As a conclusion, we see here how adding some implicit knowledge during the inductive process may change the results: the theory that appears to be prevailing without background knowledge is dominated by another explanation that seems more satisfying in the sense that it explains more examples than the first.

This induction bias was caused both by the way the rules were induced, i.e. by the used induction engine, which was based on the notion of association rules, and by the lack of information. More precisely, it was mainly due to the partial description of examples. For instance, the alimentary diet and the presence of fresh fruits and vegetables were not always inserted in cases descriptions.

A second medical example: the leprosy

To pursue our investigation, we shall now modify the representation language itself, i.e. the way examples are given to the machine. The effect of such transformation will be illustrated on another medical example: the problem of leprosy (Corruble and Ganascia 1996).

History of leprosy dates back to ancient China and India. We focus here our study to the 19th century medical views on this disease and to the conflict between two theories, the *theory of contagion* which explains the propagation of the disease by a mysterious agent that can pass from one person to another by physical contact, and an *hereditary conception* in which some people are genetically predisposed to contract the disease.

In 1874, a Norwegian physician, Gerhard A Hansen (Hansen and Armauer 1875), discovered the infectious agent, but, for ethical reasons, it was impossible to realize *in vivo* experiments that could validate or invalidate the still existing conflicting theories.

It was only during the second half of the 20th century that researchers identified individual immune reactions, which could possibly be inherited. In other words, both theories were justified even if none of them was true. In order to understand both way of reasoning, we tried to apply our inductive engine to a case based on leprosy. The used training set contained 118 cases of leprosy in the Tamtaran Asylum (Punjab) reported by Gulam Mustafa (Phineas 1889). The representation language contained 14 attributes. Without background knowledge, the induction engine generated two main "indulgent" rules, R1 and R2 plus three minor rules:

R1: IF father_affected = No THEN children = all_healthy

R2: IF father_affected = No & Mother_side = yes & disease_type = anaesthetic & age > 35 THEN children = some_sick

Nowadays, those two rules could easily be interpreted as an hereditary reaction of the immune system to the presence of the bacillus. It also appears that the disease could be classified according to the reaction which corresponds to the 20th century theory (Ridley and Jopling 1966).

As with the scurvy, we wanted to understand why 19th century physician had not discovered this simple hereditary immunity. The first answer was that, for centuries diseases were only considered as positive entities, either animated material being, materiel things or immaterial being, for instance a demon (Grmek 1995). Therefore, hereditary immunity, i.e. transmission of a negative entity, was not conceivable.

We have then reconstructed the path from those cases to the hereditary theory without reference to negative entities (Corruble and Ganascia 1996). It was done by introducing in the background knowledge some rules establishing relation between the symptoms and the affection itself. Within this configuration, i.e. with those constraints and this background knowledge, the induction engine gave six rules which could be interpreted as a hereditary theory of the disease transmission:

R2: IF disease_type = do. THEN children = all_healthy

R4: IF disease_type = do & mother_affected = yes THEN children = some_sick

R5: IF disease_type = do & father_affected = yes THEN children = some_sick

R1: IF disease_type = anesth. THEN children = all_healthy

R3: IF disease_type = tuberc. THEN children = all_healthy

R6: IF disease_type = mixed THEN children = all_healthy

The last problem was to simulate the generation of the contagious theory. In order to do that, we introduced a new descriptor called the contagious index which roughly enumerates the number of contacts with people affected by the disease. As a result, we had seven induced "indulgent" rules among which two were prominent, rules R1 and R2 that expressed the role of the contagious index:

R1: IF father_affected = yes THEN children = all_healthy

R2: IF father_affected = yes & contagious_index > 5 THEN children = some_sick

As a conclusion, it appears that by modifying the background knowledge, it was possible to change the way examples were interpreted by the induction engine, and, consequently to change the induced knowledge. One of the causes of this inductive bias was that examples were incompletely specified. The reason of these incomplete specifications was that men noticed only details that seemed relevant. Then, it should be of interest to compare the way examples are given to some implicit theories, and to see if some example sets are more adequate to some particular theory. Our last sets of experiments constitute an attempt to investigate such a comparison.

Application to social sciences

We shall try now to study common sense reasoning. The goal is both to model the way people reason and to confront different inductions with different example sets. It is to know how preconceived ideas bias the judgements and the interpretation of facts. On the one hand, it is to extend our simulation of wrong reasoning to common sense knowledge. In this respect, it is an application of artificial intelligence techniques to social sciences where it could help to apprehend the way people react to singular cases. In the past, many mathematical and computer science models were used in sociology. However, those models were mainly based on statistical analysis. Our perspective is totally different: it is to model the way individuals reason and how they interpret facts, with respect to implicit theories they have in mind. In other word, it is to model social representations.

On the other hand, this application is an opportunity to compare induction with different data sets and to see how the way data are given influences the induced knowledge.

We focused here on xenophobia in France at the end of the 19th century. We have chosen the first decade of September 1893, a few months before the Dreyfus affair burst. For all those ten days, three daily newspapers were fully scanned (Ganascia and Velcin 2004), a conservative newspaper, "Le Matin" (Le Matin 1893), an anti-semitic strong right newspaper, "La Libre Parole" (La Libre Parole 1893) and a catholic one, "La Croix", also very conservative (La Croix 1893). We gathered all published articles of social dysfunctions, such as political scandals, corruptions, bankrupts, robberies, murders etc. Each of those articles was viewed as a single case, described with a small representation language, similar to those used in the Scurvy and in the Leprosy experiments. This language contains 30 attributes corresponding to the political engagement of protagonists (socialist, radical, or conservative), their religion, their foreign origin, if they are introduced abroad, etc…

Sets of articles from each daily newspaper (here "Le Matin", "La Libre Parole" and "La Croix") were represented in the same way, with the same description language, but they were considered separately, each of them constituting a separate learning set.

Our goal was both to induce rules and theories, with each of those learning sets, but also, to introduce different implicit theories and to compare the adequacy of each learning set, i.e. of each set of examples, to each theory.

Four different theories were considered to explain social disorders:
The first theory explains the deterioration of the society by an international Jewish and Freemason conspiracy.
The second theory mentions the loss of national traditions and qualities.
The third refers to incompetence and inability of politicians.
The last relies disorders to corruption
Those four theories were drawn from historical studies (Taguieff 2002), (Bredin 1983). We simplified and translated all of them into a set of production rules.
Our aim here was not to study the effect of background knowledge on the explanation power, as it was the case in the two last studies, but to investigate the implicit knowledge concealed behind the examples. This is the reason why we needed different data sets, which correspond here to different sets of articles from different daily newspapers. For each of those data sets, we first induced explanation patterns, as we have done previously, by inserting our examples in the induction engine, without background knowledge. Then, we evaluated the explanation power of all generated explanatory schemata. We wanted to investigate here not those explanation patterns by themselves, but the implicit theory hidden in the back. In other words, newspapers seemed to be read by people with some embedded assumptions. To validate this idea, we introduced successively each of the four initial theories mentioned previously, in our induction engine, as background knowledge. Then, we computed again, for each of those theories, and for each of the data set, the explanation power of each explanation pattern.
For the sake of clarity, let us take an example: figure 4 shows the explanation power of explanation patterns built on four attributes, *tendency, morality, corruption* and *connection_with_Jews* without background theory (blue line) and with theory of corruption as background theory (red line).

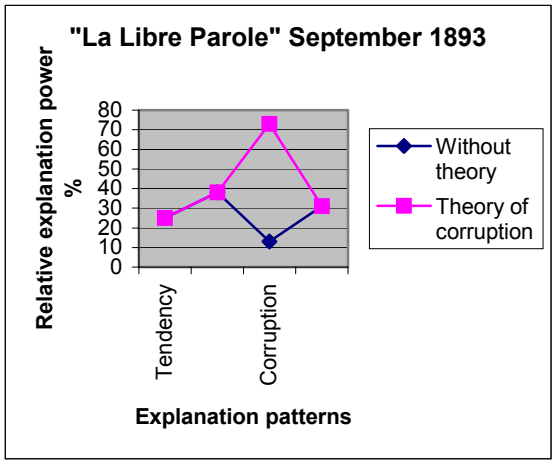

Figure 4: explanation power of four different explanation patterns with and without the theory of corruption

It clearly appears that the presence of the theory of corruption makes higher the explanation power of the attribute corruption, and this renders all the examples more understandable. More technically, with this background theory, the percentage of examples that can be explained by some explanation pattern increases considerably. This remark may be generalized: for each theory, the optimal explanation power is noted, i.e. the highest explanation power, among all explanation powers of all explanation patterns.
The figure 5 summarizes the results that we have obtained. Each curve corresponds to one newspaper. The X-axe is associated with the different initial theories, the Y-axes, with the optimal explanation power, i.e. with the percentage of examples of the training set explained by the explanation patterns that has the highest explanation power. The figure shows that the value of the optimal explanation power is in accordance with the tendency of the corresponding newspaper. For instance, the theory of corruption and the theory of conspiracy have a very high relative explanation power for "La Libre Parole", which is an Anti-Semitic extreme right newspaper. On the opposite, the explanation power of the theory of corruption is relatively low for "Le Matin" and "La Croix", two traditional and conservative newspapers. It means that the theory of corruption and the theory of conspiracy are implicit for most of the readers of "La Libre Parole", why both theories are not implicit for the remaining two.

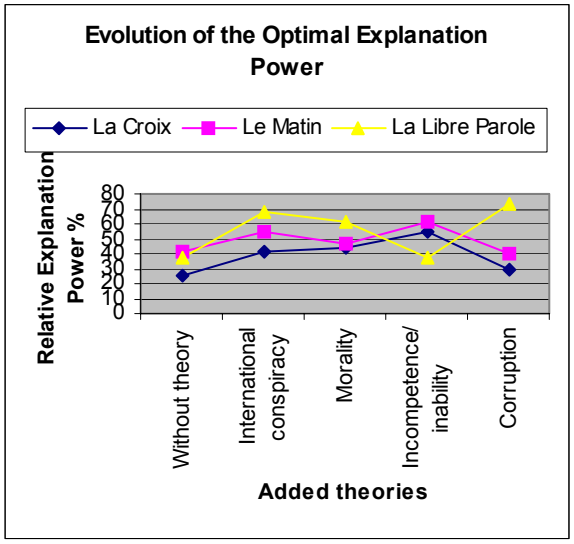

Figure 5: evolution of explanation power with theories

On the other hand, the theory of incompetence, that has the lower value for "La Libre Parole", seems to explain many examples drawn from "Le Matin" and "La Croix", even if it is less significant for "La Croix". Last point, the theory of morality appears to be more explicative than the theory of conspiracy for "La Croix" while it is the contrary for "Le matin". Since "La Croix" is a catholic newspaper and "Le Matin", just a conservative newspaper, this difference

could be easily understandable. For more details concerning this study see (Ganascia 2005), (Ganascia and Velcin 2004).

As a conclusion, we observed that, simulating our model on different data sets with different implicit theories, it becomes apparent that some data sets were more easily understandable with one implicit theory than with the others, which means that data sets predispose to some interpretations. Since those implicit theories were directly connected with the political tendency of daily newspapers from which examples were drawn, it validates our model. In other words, it explains how examples induce misrepresentations. Even if none of the examples is false, the way they are represented, the lack of description and the presence of implicit knowledge may considerably influence the induction. More precisely, examples lead people to construct an implicit theory, by abduction, and this implicit theory will then contribute to facilitate induction and generalization from examples.

Partial Conclusion

In all the previous experiments – cause of scurvy, transmission of leprosy and xenophobia in France at the end of the 19th century – it appeared that examples descriptions were very sparse, which rendered different interpretations possible. For instance, in case of scurvy, the alimentary diet was not always explicitly mentioned in the description of the disease episodes. This is certainly why, in presence of the blocked perspiration theory, the explanatory power of the humidity attribute passed beyond the explanatory power of attributes relative to the presence of fruits and vegetables in the alimentary diet.

Since this phenomenon appeared to be crucial in common sense induction, i.e. in the way people derive knowledge from personal experience, we tried to model and to generalize it in a logical framework. The next section is dedicated to the presentation of this logical framework.

Stereotype Extraction

Default Generalization

During the eighties, there were many attempts to model deductive reasoning in presence of implicit information. A lot of formalisms have been developed to encompass the inherent difficulties of such models, especially their non-monotony: close-world assumption, circumscription, default logic (Reiter 1980), etc.

Since our goal here is to model the way people induce empirical knowledge from partially and inhomogeneously described facts, we face a very similar problem: in both cases, it is to reason in presence of implicit information. Therefore, it is natural to make use of similar formalisms.

In our case, the problem is not to deduce, as it was with default logic, but to induce knowledge from implicit information. Therefore, we put forward the notion of *default generalization* which is the equivalent for generalization to default rule for deduction (for more details, see (Velcin and Ganascia 2005). In brief, A generalize B by default means that there exists an implicit description C such that B complemented with C, i.e. $B \wedge C$, is more specific than A in the classical sense, which signifies that A entails $B \wedge C$. The exact definition is the following:

Definition: <u>d generalize d' by default</u> (noted $d \leq_D d'$) iff $\exists d_c$ such that $d \leq d_c$ and $d' \leq d_c$ where $d \leq d'$ stands for d is more general than d' in the classical sense.

Set of stereotypes

Our main hypothesis is that groups of people share implicit knowledge, which makes them able to understand each other without having to explicit everything.

Our second hypothesis is that this implicit knowledge is stored in terms of sets of stereotypes. This means that many people have in mind sets of stereotypes and that they reason in a stereotyped way, by associating new events or news individuals to stereotypes they have in mind.

To formalize this idea, let us first suppose that a description space $\mathcal{D}$ and a set of examples $\mathcal{E}$ are being given.

Then, a measure of similarity M_s have to be defined on $\mathcal{D}$. We do it by quantifying the number of common descriptors:

$M_s : \mathcal{D} \times \mathcal{D} \rightarrow \Re+$

$(d_i,d_j) \rightarrow M_s(d_i,d_j) = |\{a \in d, d = d_i \wedge d_j\}|$

Definition: A <u>set of stereotypes</u> can be viewed as a set of descriptions, which are non redundant and cognitively cohesive i.e. $S = (\mathcal{s}_1, \mathcal{s}_2, \ldots \mathcal{s}_n)$ is a set of stereotype iff

- $\forall i \in [1,n]\ \mathcal{s}_i \in \mathcal{D}$ – i.e. it is a set of descriptions
- $\forall d \in \mathcal{D}\ (d \in \mathcal{s}_i \wedge d \in \mathcal{s}_j) \Rightarrow i = j$ (non redundancy)
- $\forall\ (d_i,d_j) \in \mathcal{s}_k$ it is always possible to find a series of examples that makes it possible to pass by correlation from d_i to d_j. (cognitive cohesion)

Definition: $\mathcal{E}$ being a set of examples, the so-called training set, $S = (\mathcal{s}_1, \mathcal{s}_2, \ldots \mathcal{s}_n)$ being a set of stereotype and $S_{\mathcal{s}1, \mathcal{s}2, \ldots \mathcal{s}n}$ being the function that associates to each individual e its relative cover, i.e. its closest stereotype with respect to M_s and $S = (\mathcal{s}_1, \mathcal{s}_2, \ldots \mathcal{s}_n)$, the <u>cost function</u> h that evaluates a set of stereotypes $S = (\mathcal{s}_1, \mathcal{s}_2, \ldots \mathcal{s}_n)$ is defined as follows:

$$h(E, S = (\mathcal{s}_1, \mathcal{s}_2, \ldots \mathcal{s}_n)) = \sum_{e \in E} M_s(e, S_{\mathcal{s}1, \mathcal{s}2, \ldots \mathcal{s}n}(e))$$

Once the cost function h is defined, the non-supervised learning algorithm has to build the set of stereotypes ($\mathcal{E}_1$, $\mathcal{E}_2$, … $\mathcal{E}_n$) that minimizes h. In other words, the non-supervised learning problem is reduced to an optimization problem.

Validation

Our first evaluation validates on artificial data sets the robustness of the non-supervised learning algorithm, which builds sets of stereotypes from a learning set $\mathcal{E}$ and a description language $\mathcal{D}$. Our experiments clearly show that the learning process is very stable: up to 85% of degradation, the error rate is less than 10%.

As it was said previously, our study of social misrepresentation focuses on xenophobia and anti-Semitism in France at the end of the 19th, a few months before the Dreyfus affair burst.

Our last experiments consist in extracting sets of stereotypes from news extracted from each of the four newspapers previously mentioned and to interpret them with regard to the political tendency and the cultural level of the readers.

The obtained results may be interpreted in two ways. The first is relative to the number of stereotypes generated and to the percentage of examples covered. Depending on the newspaper, the results are quite different. For instance, the news from "La libre parole", which is an extreme right newspaper, generate 2 stereotypes, among which one covers 90% of the initial examples. Moreover, it appears that only 4% of the examples are not covered by any of the built stereotypes.

On the other hand, news drawn from "Le Matin", a moderate conservative newspaper, contribute to generate 3 stereotypes that are far more balanced, while 16% of the examples are not covered by any of the stereotypes. By contrast to "Le Matin", it appears that "La libre parole" is far more dogmatic. In case of "Le petit journal", a popular newspaper, there are 2 generated stereotypes, which covers are quite balanced, but the examples of examples not covered by any stereotypes, 8%, is lower than in the case of "Le Matin".

Let us now consider the descriptions of the generated stereotypes. The main stereotype of "La libre parole" is a real caricature: it corresponds to a man who is simultaneously socialist, internationalist, antipatriotic, in connection with Jews and Protestant, corrupted, anticlerical, involved with freemasonry, immoral, etc. And the second stereotype, which covers only 6% of the examples, is described as a catholic involved with freemasonry. Among the three stereotypes generated from "Le Matin", the first corresponds to a socialist, involved with freemasonry, anticlerical, traitor to the nation etc., which corresponds to the dominant stereotype of "La libre parole". However, the second and the third stereotype are quite different: the second corresponds to an opportunist politician who is republican and incompetent, while the third explain the dysfunction by health problems of the politicians.

Conclusion

The proposed approach leads to model social representation from news. The results have been shown to historians and sociologists who are very enthusiastic. The main point is to extract implicit knowledge shared by groups of people in society and to model them with artificial intelligence techniques. Then it is be possible to discriminate among different social logics while statistical approaches in social science are restricted to the evaluation of inquiry. In this sense, it constitutes a cognitive modeling alternative to traditional approaches in social sciences.

References

Agrawal R., Imielinski T., Swami A.: "Mining Associations between Sets of Items in Massive Databases", Proc. of the ACM-SIGMOD 1993 Int'l Conference on Management of Data, Washington D.C., May 1993, 207-216.

Bredin J-D, *L'affaire*, Julliard, 1983

Carpenter K.J., The history of scurvy and vitamin C (Cambridge University Press, 1986).

Corruble V. and Ganascia J.G., Discovery of the Causes of Leprosy: a Computational Simulation, in: Proceedings of 13th National Conference on Artificial Intelligence, Portland (OR), USA (1996) 731-736.

Corruble V., Ganascia J.-G.- Induction and the discovery of the causes of scurvy: a computational reconstruction. Artificial Intelligence Journal. Elsevier Press, (91)2 (1997) pp. 205-223

Drumont E., *La France Juive*, Paris, V. Palmé, 1886

Ganascia J.-G., "Rational Reconstruction of Wrong Theories", in *Logic, Methodology and Philosophy of Science*, Petr Hajek, Luis Valdes-Villanueva, and Dag Westerstahl (eds.), London, KCL, 2005.

Ganascia J.-G.- TDIS: An algebraic generalization. IJCAI-93 International Joint Conference on Artificial Intelligence, Chambéry, France, 1993

Ganascia J.-G.- CHARADE : A rule System Learning System. 10th IJCAI, Milan, 1987

Grmek, M.D. 1995. Le concept de maladie. *Histoire de la pensée médicale en Occident*. Ed. du Seuil, 1995

Hansen, G. Armauer. 1875. On the etiology of leprosy. *British and foreign medico-chirurgical review*, 55.

La Croix, daily newspaper from September the 1st 1893 to September the 10th 1893

La Libre Parole, daily newspaper from September the 1st 1893 to September the 7th 1893

Le Matin, daily newspaper from September the 1st 1893 to September the 10th 1893

Mahé J., Le scorbut (in French), in: Dictionnaire Encyclopédique des Sciences Médicales, Série 3, Tome 8 (Masson, Paris, 1880) 35-257.

Phineas, S.A. 1889. Analysis of 118 cases of leprosy in the Tarntaran Asylum (Punjab). *Transactions of the Epidemiological Society of London*. v. 9 (1889-1890)

Ridley D.S. and Jopling W.H. 1966. Classification of Leprosy According to Immunity, A Five-Group System. *International Journal of Leprosy*. v. 54, n. 3. pp. 255-273.

Thagard P. and Nowak G., The Conceptual Structure of the Geological Revolution, in: J. Shrager and P. Langley, eds., Computational Models of Scientific Discovery and Theory Formation (Morgan Kaufmann, 1990) 27-72.

Velcin J., Ganascia J.-G., "Stereotype Extraction with Default Clustering", in proceedings of the 19th International Jointed Conferences on Artificial Intelligence, IJCAI-2005, Edinburgh, UK, pp. 883-888