Web-Based Five Senses Input Simulation – – Ten Years Later

Rafal Rzepka and Kenji Araki

Graduate School of Information Science and Technology
Hokkaido University, Sapporo, Japan
{kabura,araki}@media.eng.hokudai.ac.jp,
http://arakilab.media.eng.hokudai.ac.jp

Abstract. A decade has passed from our paper in which we question the sensors technology not being ready to give a machine the five senses with the recognition level close to human's and, as we expected, sensing devices are still not sufficient to become a base for the world knowledge based reasoning. Although we did not developed the proposed system any further because of limited access to Internet data, the need for machine experience is now even more urgent than before. Existing world knowledge retrieval systems tend more and more to concentrate on name-entity information ignoring very basic world knowledge which is still one of the most difficult obstacles on the path for reasoning about everyday situations. In our paper we will introduce results of our latest trial to automatize previously proposed theory for enriching machine's capability to simulate environmental perception and deepen human behavior understanding.

1 Introduction

Our long-term project for acquiring everyday knowledge (GENTA - GENeral belief reTrieving Agent)[1] is based on experiences shared by Japanese bloggers. Ten years ago, when we performed some experiments with the *Bacterium Lingualis* method[2], used by GENTA for the knowledge and language acquisitions, we noticed that expressions used for recognizing negative and positive load of a sentence are heavily context-dependent. Machines lacking basic experiences are easily fooled by words with multiple meaning which, when just counted without considering the context, bring erroneous results. Famous "colorless green ideas sleep furiously" by Chomsky [3] are seriously treated by machines because of the experiences shortage, which, in turn, is caused by their still poor physical world sensing.

Also in humans, at least in theory, something that contradicts our experiences, make us alert, suspicious, uncertain, hesitating and if we fail to recognize such state, we are prone to errors. In our 2003 paper [4], we stated that "before the computers start to seek for the sixth sense or start to understand beauty, nuances and metaphors they should have a solid commonsense built on physical experiences" and ten years later we still support that claim. There are multiple

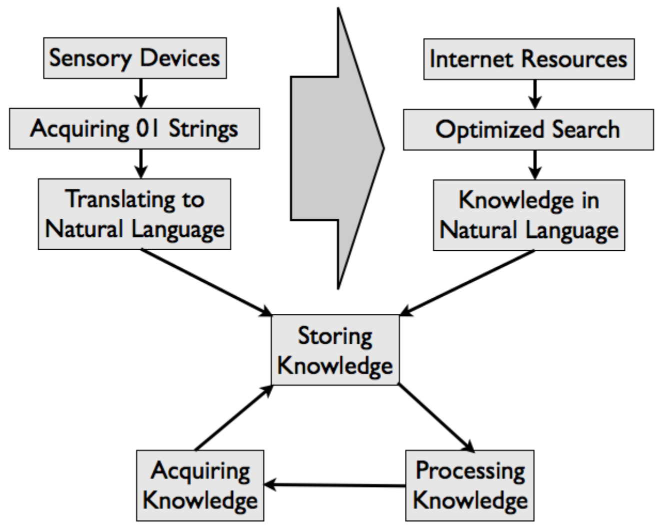

Fig. 1. Overview of the shortcut idea where Internet resources supplement a machine's experiences when they lack real sensations.

theories on how the human beings are creating a simple perception from their five senses. Aristotle [5] claimed that the commonsense (kōine aisthēsis) is build upon our five senses. Later John Locke [6] proposed a similar definition of "common sense" saying that it builds on phenomenological experience. Each of the senses gives input, and then the sense-data are integrated into a single "impression" and this *integration* is made by the common sense – the sense of things in common between essentially different impressions. To achieve a rich simulation process for testing these theories we would need not only technologically advanced sensors and the fusion of their constant input but also methods for their automatic interpretation human user to be understood. Programming haptic device to generate output understandable in the same way as, for instance, sound recognizer is obviously possible, however it would be rather difficult when there are five inputs and the central unit has to learn mutual relations to be used in interaction with human. This is why we always promote *natural languages* for an universal and widely accessible method for accessing and manipulating machines and their learning processes, in spite of their ambiguity. Although a huge step forward in understanding simple utterances, technologies for other four senses did not show any revolutionary progress and a wide interpretation on what is seen on a picture or video, what is touched, smelled or tasted in real world – is still out of the reach of conventional AI systems. Hence, the most difficult part of learning by physical experience which is convergence of all sensor inputs into the one interpretation (perception) – simply cannot be performed. Our proposition was to use web-mining to simulate the sensor-input experiences and to help a machine to create its common sense based on other people experiences. For a decade we left this idea mostly untouched but we decided to return to its realization and utilize its results for generating new Japanese entries for ConceptNet [7] and for context processing which we need for our text understanding tasks, for example Machine Ethics project that uses text-mining [8] and metaphor generation projects using a big corpus of Japanese figurative sentences [9].

The long-distance goal of this research is a machine that could simulate input from the 5 senses when sensors data is not available or is available but inefficient (see Figure 1). Such process could a base for what we call *artificial imagination*.

2 Basic methods for simulating non-physical experiences

As we aim at machines that understand human behavior, not at an information retrieval task, our approach is different from name entity centered systems like KnowItAll [7], TextRunner [8] or NELL (Never Ending Language Learner) [9]. Our goal is to develop and test a multidisciplinary approach for using real world data in order to achieve common sense knowledge. So far we have been working on several projects with two common topics commonsensical and emotional knowledge retrieval from Internet resources as they are the biggest and “alive” as their contents are changing with people’s opinions and beliefs. The basic idea is to retrieve needed strings of words, count the frequency and calculate the interdependencies between the comparable items. Some examples will be introduced below.

2.1 The Knowledge Source

Currently we still work only on Japanese since this language seems to have an easier structure for computer processing especially because of its particles usage what Fillmore has suggested in his works [11]; another reason are the cultural differences that influence common sense depending on the language. For the experiments we use five billion sentences corpus [12] which is easier to use than the data we collected 10 years ago with Larbin robot, we also use Solr instead of limited access to Google, acquiring comparable search speed. Ameba blog service used in [12] has a stiff html structure which allows to get rid of non-blog contents. Our first challenge to cope with five senses was made mostly manually, this time all the of experiments described in this paper were done manually.

2.2 Eyes Input Simulation

To test all the five senses simulations we have tested them using 127 action phrases as “(to) call a doctor” or “(to) steal a car”. This list was an enhanced set we use in our research for recognizing ethically problematic acts [8] – this time we added more everyday life actions like “writing a book” or states like ”someone laughing” and removed similar entries, for example these needed for comparing reactions to killing various kinds of animals, which in most cases have

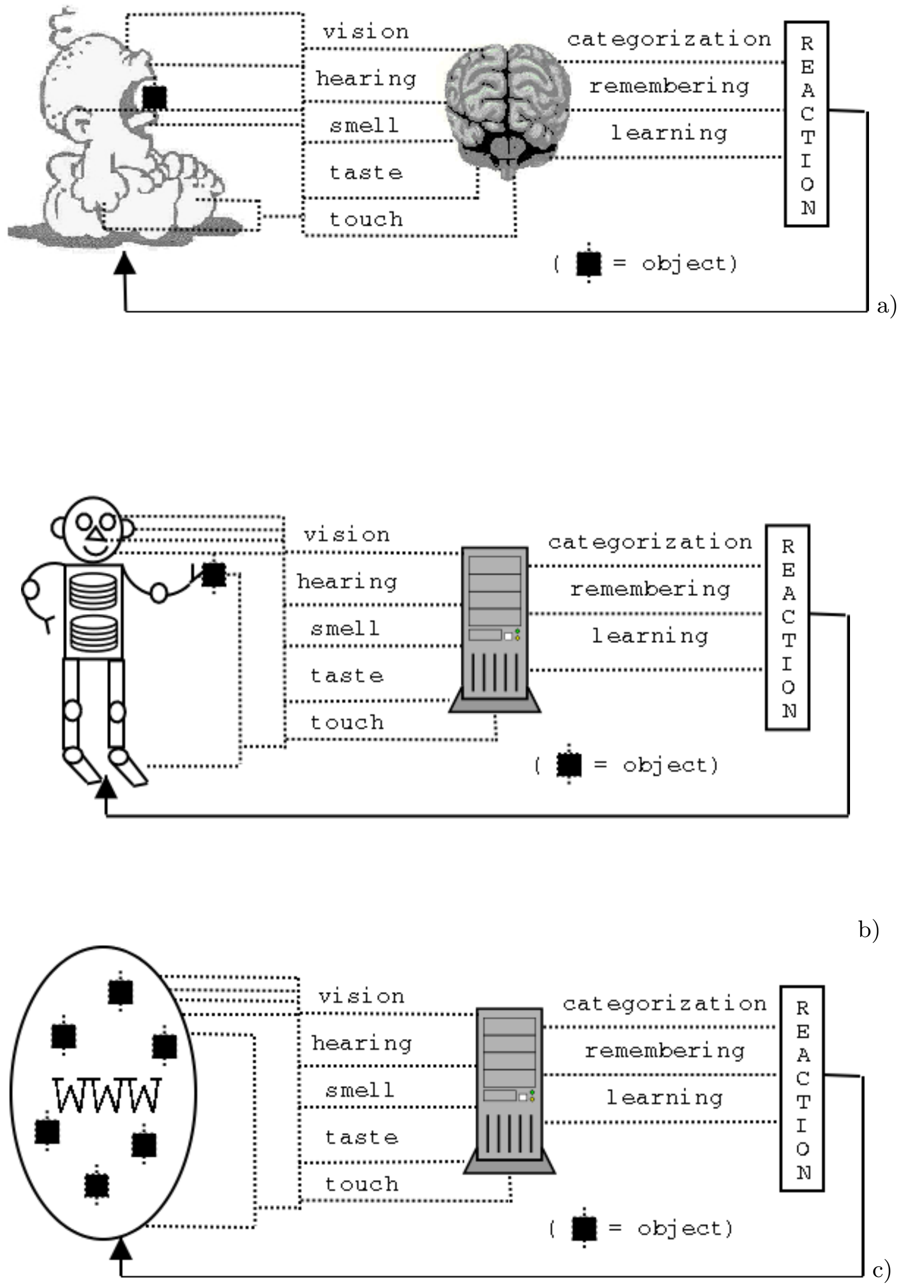

Fig. 2. An original figure from 2003 paper: a) A human child learning from the five sense experiences, b) the ideal robot with perfect sensors and c) the idea of WWW-based supplement for the sensors input memory

very low hit rate in blogs. Input phrases have Noun-Particle-Verb structure and the sensing words (mostly adjectives) are exact matches for SenseWord-Noun. For the visual input simulation we chose a set of basic descriptive adjectives (See Table 1) shapes and colors, which is the biggest set of all five. A "commonsense border" of 10 hits was experimentally set to limit "peculiar cases" as "cold sun" or "black snow".

Table 1. 36 sense words for recognizing visual characteristics.

Japanese	akarui, kurai, ōkii, chiisai, akai, kiiroi, kuroi, shiroi, aoi, midori-no, asai, fukai, usui, ōi, sukunai, osoi, hayai, kagayakashii, chairo, nagai, furui, makkuro-na (3 types of writing), kitanai, kirei-na (2 types), shikakui, chikai, tōi, takai, hikui, hiroi, semai, futoi, hosoi
English	bright, dark, big, small, red, yellow, black, white, blue, green, shallow, deep, thin, large, small, slow, fast, shiny, brown, long, old, pitch black, dirty, clean, square, close, distant, high, low, wide, narrow, thick, thin

2.3 Fingers Input Simulation

Perfect haptic sensors are probably the most difficult ones to develop, however sense of touch is an important natural tool for keeping our everyday life safe. For this input simulation we used adjectives and onomatopoeias which are associated with surface characteristics, shown in Table 2.

Table 2. 34 sense words for recognizing haptic characteristics.

Japanese	katai (3 types), betabeta suru, tsumetai, yawaraka, omoi, karui, suzushii, asai, fukai, attakai (2 types), atsui (3 types), itai, usui, samui, surudoi, togatte iru, eiri-na, nurui, kawaita, nureta (2 types), nebaneba shita, nurunuru shita, jimejime shita, munmun suru, zarazara shita, tsurutsuru shita, gotsugotsu shita, fuwafuwa shita, guchagucha shita
English	hard, tough, sticky, cold, soft, heavy, and light, cool, shallow, deep, warm (2 types), hot (3 types), thick, painful, pale, cold, sharp, pointed (sharp), tepid, dry, wet (2 types), slimy, gluey, steamy, sultry, rough, slick, rugged, fluffy, squashy

2.4 Ears Input Simulation

Firstly, it must be clarified that in "ears input" we do not include speech recognition but only sound characteristics. As in the case of touch, only adjectives do not bring sufficient hits, we collected also mimetics (*gitaigo*) and onomatopoeias (*giseigo*) to broaden the search. The set is shown in Table 3.

Table 3. 31 sense words for recognizing sound characteristics.

Japanese	urusai, shizuka-na, pokipoki suru, zāzā suru, biribiri suru, gorogoro suru, bukubuku suru, kasakasa suru, katakata suru, gatagata suru, gachagacha suru, gachan suru, karan suru, gangan suru, kiikii suru, gōngōn suru, sarasara suru, janjan suru, charin suru, chirin suru, chirinchirin suru, chinchin suru, tonton suru, batan suru, patan suru, pachipachi suru, pachipachi suru, pachin suru, pishari suru, rinrin suru
English	noisy, quiet, crack, rushing water, tear, rumble, bubble, rustle, rattle, clatter, clank, pound, squeak, purr, murmur, continuous sound, clink, tinkle, jingle, ding, knock, bang, snap, crackle, clack, smack, ring

2.5 Tongue and Nose Input Simulations

Probably there is no need to place a tasting sensors in robot's mouth they could be also located in its fingers. But the input should be interpreted the same way as humans do and the recognition output should be one of this category sense words. The same situation is with smelling sensors. In 2003 we proposed an idea to use smell-associated expressions as "smells like X", "smells sweet", "smells fishy", etc, but to keep the consistency with other word sets, we limited the sense words for taste and smell to basic ones shown in Table 4 and avoided metaphors.

Table 4. 14 sense words for recognizing taste and 6 for smell characteristics.

Japanese (taste)	amai (2 variations), nigai, suppai (2 variations), karai (2 variations), shoppai, nigai, shibui (2 variations), amazuppai, shibō-no ōi, aburakkoi
English	sweet (2 types), bitter, sour (2 types), spicy (2 types), salty, bitter, astringent (2 types), sweet and sour, fatty, greasy
Japanese (smell)	kusai, kōbashii, namagusai, kaguwashii, ii nioi-no, ii kaori-no
English	smelly, aromatic, fishy, fragrant, of a nice smell, smelling nice

3 Sense Word Sets Efficiency

90 nouns describing physical instances were extracted by the first author from 127 phrases and input to the five senses recognizer. Then the first author evaluated the output. Retrieval precision appeared to be high achieving 0.96, but the recall was rather low achieving 0.43 which gives f-score of 0.60. There was only one error for 30 retrievals caused by metaphoric description of a novel - "deep". Possible reasons for the very low recall were as follows. Firstly, the blog data used were searched by whole phrases with "if" statements, because we focus on cause-effect relationship in our research and always tend to use full sentences to avoid less meaningful blog entries. The second reason that was visible was simple too low number of sense words. Another problem was deciding if a noun

is physical instance or not. For example word “hito” (*a man*) was used in the examples in the meaning of *somebody* and a ”shōsetsu” (*a novel*) was used in a category of a book context rather than of a physical book. The same problem was with name places as a city name can be treated as a common description for group of physical instances as buildings, streets, monuments, etc., but it is not obvious that you can touch Tokyo (however one could smell it literally or hear the sound of it).

3.1 Repeating Experiments With Bigger Data

For the second trial we have used sets of Japanese words taken from a thesaurus [15] to avoid problematic inputs. To limit errors caused by lack of dependency parsing and nouns used for describing other nouns (i.e. “black snow shoes”, Japanese particles were added to nouns in all the queries (direct object *wo*, theme / topic *ga* and *wa*, object *ni* and instrument / location *de*). We used word category “tools, decorations, monuments, etc.” assuming that it consists of words describing physical artifacts. There were 289 words in the category and we used them for repeating the retrieval experiment. The system again reached a high precision of 0.92 but this time the recall dropped drastically to 0.08 which resulted in very low f-score of 0.15. Used thesaurus had too many rare words that never appeared in the blog corpus. If we limit the dataset to nouns which had a hit rate above 10 (59 hits), the results were much better: recall increases to 0.42 but it gave f-score of 0.57 which was lower than the results for the smaller set.

4 Conclusions and Future Work

The main goal of this research is to simulate the human’s perceptual process. It is much easier for a machine to guess the observed object when more sensors than one are working. For this reason, our proposed simulation system needs as rich sensory input as possible, therefore the achieved recall is far too low. However, we managed to mechanize our idea from 2003 and the system achieved very high precision of 0.92. Consequently the next obvious step is to increase the range of search and number of sense words. Because there is quite a possibility that rare words will be fed to the system, some kind of categorization is needed – if a ’snake’ category member ’anaconda’ has not enough hits, other members of the snake family could bring more knowledge about the object with inefficient description. To assure that we are dealing with a physical, not metaphorical world, we would also need to consider altering the sense words to avoid ambiguities as in case of word “high” which in Japanese can also mean “expensive”. Such system, if successful could possibly help simplifying idiom or metaphor understanding or generation – we are currently working on the latter. During the tests in 2003 we noticed that there are many adjectives, as for example “soft” (*yawarakai*), which appears in more than one sensor simulator. This time we have left such words unaltered, however we plan to experiment with more specific queries as "sounds soft", "looks soft", "feels soft", etc. We will also have to find a method for recognizing named entities which also bring a possibility for erroneous sensing.

In this paper we introduced several ideas for web-based support for machines that need to learn by experience. We automatized our ideas from 2003 for supporting agents that work in the physical world as their number will be growing constantly. Even when the five types (or more) of sensors symbolizing human's five senses are able to transmit proper information, there are cases which stand against the commonsense and the appropriate reactions based on a wider experience are needed. However, gaining various experiences by a robot or a group of robots in different environments, would demand a lot of money and time. Therefore we assumed that senses supplement each other (we were inspired by a phenomenon of the spatial knowledge of blind children which keeps growing even without visual input [17][18]) and treating experiences of others as own ones, would help to reduce costs of knowledge acquisition in future.

References

1. Rzepka, R., Araki, K., Tochinai, K.: *Is It Out There? The Perspectives of Emotional Information Retrieval from the Internet*. Proceedings of the IASTED Artificial Intelligence and Applications Conference, ACTA Press, Malaga, pp. 22-27 (2002)
2. Rzepka, R., Araki, K., Tochinai, K.: *Bacterium Lingualis - the Web-based Commonsensical Knowledge Discovery Method*. Volume 2843 in the Lecture Notes in Artificial Intelligence series of Springer-Verlag. pp. 453–460 (2003)
3. Chomsky, N.: Syntactic Structures, Mouton classic series. Walter De Gruyter Incorporated (2002)
4. Rzepka, R., Araki, K. and Tochinai, K.: *Five Senses Input Simulation Based on Web Resources*. Technical Report of Language Engineering Community Meeting (2003)
5. Modrak, D.K.: *Aristotle: the Power of Perception*. University of Chicago Press, Chicago (1987)
6. Locke, J.: *An essay concerning human understanding* Oxford University (1841)
7. Speer, R. and Havasi, C.: *Representing General Relational Knowledge in ConceptNet 5*. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), published by European Language Resources Association (2012)
8. Rzepka R. and Araki K.: *Polarization of consequence expressions for an automatic ethical judgment based on moral stages theory*. IPSJ SIG Notes 2012-NL-207(14), pp. 1–4 (2012)
9. Rzepka R., Dybala, P., Sayama, K. and Araki K.: *Semantic Clues for Novel Metaphor Generator*. To appear in the Proceedings of IJCAI Workshop on Computational Creativity, Concept Invention, and General Intelligence, (2013)
10. Rzepka R., Araki K., Tochinai, K.: *Emotional Information Retrieval for a Dialogue Agent*. "Perception and Emotions Based Reasoning" – Special Issue of "Informatica" - An International Journal of Computing and Informatics, Volume 27, Number 2, pp. 205-212 (2003)

11. Fillmore, J.C.: *The Case for Case*. E. Bach & R.T.Harms, eds., Universals in Linguistic Theory, New York: Holt, Rinehart & Winston, pp. 1-88 (1968)
12. Maciejewski, J., Ptaszynski, M. and Dybala, P.: *Developing a Large Scale Corpus for Natural Language Processing and Emotion Research in Japanese*, International Workshop on Modern Science and Technology (IWMST 2010), September 4-5, (2010)
13. Minsky, M.: *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon and Schuster, New York (2006)
14. Damasio, A.R.: *The feeling of what happens: Body and emotion in the making of consciousness*. Harvest, New York (1999)
15. Natural Institute for Japanese Language and Linguistics (ed.): *Bunrui Goi Hyo*, Dainippon Tosho (2004)
16. Murphy, R.R. : *Biological and Cognitive Foundations of Intelligent Sensor Fusion*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 26(1), pp. 42-51 (1996)
17. Kielkopf, C.F.: *The Pictures in the Head of a Man Born Blind*. Philosophy and Phenomenological research, Volume 28, Issue 4, pp. 501-513 (1968)
18. Fletcher, J.F.: *Spatial representation in blind children, Development compared to sighted children*. Journal of Visual Impairment and Blindness 74, pp. 381-385 (1980)