THE HISTORY, PHILOSOPHY AND DEVELOPMENT OF COMPUTATIONAL INTELLIGENCE (HOW A SIMPLE TUNE BECAME A MONSTER HIT)

James C. Bezdek
Computer Science, U. of Melbourne, Parkville, Vic., Australia

Keywords: Computational Intelligence, Artificial Intelligence, Biological Intelligence, Neural Networks, Fuzzy Systems, Evolutionary Computation

Contents

Summary

Hisao Ishibuchi asked me to write the first chapter in this volume. He suggested the title "History, Philosophy and Development of Computational Intelligence" but for reasons that will soon be clear, I added the subtitle you see above. Why ask me to write it? Well, [I think] I wrote the first paper that defined the term computational intelligence (Bezdek, 1992), and it was my idea to attach the term computational intelligence to several activities related to the *IEEE Computational Intelligence Society* (IEEE CIS). The most important event in this regard was the World Congress on Computational Intelligence (WCCI) first held in Orlando in 1994. Indeed, the current name of the society – IEEE CIS – also had its roots in my suggestion. But, as I have been careful to point out many times, the term CI itself was around for at least seven years before I wrote that paper. This chapter finally sets that part of the record straight. Let me start this excursion into the past with a seemingly unrelated but soon to be understood word association game about popular music.

1. Prelude: Art and Science Share a Common Trait

Suppose I name the song "Like a Rolling Stone" which the magazine *Rolling Stone* lists as the greatest song of all time. (You may not agree that this is the greatest song of all time, or even of any time, because of its context, language, your cultural history, your personal preferences and so on – that's ok, it will still suffice to make my point.) Many of you will know who wrote it, who performed it first, and who made it a huge hit; Bob Dylan (1965) in all three cases. But what about the song "[I'm Dreaming of a] White Christmas"? I'm pretty sure that many of you can tell me who made it popular (Bing

Crosby), but probably most of you don't know that Irving Berlin wrote it in 1940, and that Bing did make the first, and also most popular, recording of it in 1942**.** According to the Guinness book of World Records, the holiday perennial "White Christmas" (1942) by Bing Crosby is the best-selling single worldwide, with estimated sales of over 50 million copies. You may be surprised to learn that the first hit of one of America's most distinguished "pop" singers, Tony Bennett, was Hank Williams' (arguably best) country song, "Cold, Cold Heart."

It is often the case in music that the general public attributes the creation of a well-known song to the artist who made it most popular. The identity of the song writer is often obscured by the dazzling success of the performer who made it a big hit. There are lots of examples in music of anomalies, quirks, little-known facts, obscure references, "guest" credits for songwriters, and general mis-(or is it dis-?) information about songs. This happens in science and engineering too.

2. Overture: Songwriters and Performers in Science and Engineering

Suppose I state a phrase, say a technical term, name a concept, or repeat a physical law, that is common in science or engineering, and ask you to associate a name with that phrase, term, concept or law. For example, the law "energy equals mass times the square of the speed of light"; in symbols, $E = mc^2$. Almost everyone on planet earth (well, ok, almost everyone amongst the more than 400,000 members of the IEEE anyway) can tell me that this equation was discovered (created), and popularized, by Albert Einstein.

But suppose I ask for the origin of the term "*genetic algorithms* (GAs)"? What name leaps into your mind first? Perhaps David Goldberg, who wrote the first popular text about this topic (Goldberg, 1989). Or your response might be John Holland, who is sometimes credited as the "inventor" of GAs (Holland, 1975). The actual history of genetic algorithms is quite complicated, and the origins of various algorithmic forms of evolutionary computation can be rightfully attributed to several creators, but certainly not to Goldberg. See Fogel (1998) for an eminently readable and cogent account of the actual history of GAs as well as other forms of evolutionary algorithms.

Here's another example: "backpropagation in multilayered neural networks". This very famous and useful technique was created and reported by Paul Werbos in his PhD thesis (Werbos, 1974), but it was, for many years, erroneously credited to David Rumelhart and James McClelland, who popularized it in their 1986 book (Rumelhart and McClelland, 1986).

Where does all this lead? Suppose I ask the membership of the IEEE: Who originated the term "*computational intelligence* (CI)"? Many – perhaps most - of them might say "Jim Bezdek" but they would be wrong. I probably made this term a big hit song, but I did not write the song. That piece of the story will come to light soon. But first, let's return to the music analogy, which was not my invention either!

I will start by discussing the history of a different technical term whose path is strikingly similar to the historical evolution of the term CI. In particular, there is a close

parallel between the terms CI and EM, which stands for "*expectation-maximization* (EM*).*" I have done a lot of work related to the theory of what some writers call *alternating optimization* (AO). AO is the scheme employed by EM when estimating the unknown parameters of a bunch of mixed probability distributions. In due course, I became very interested in trying to track down the history of AO, and my inquiries into this topic became somewhat inseparable from the history of the EM algorithm. Let me call this combined history EM/AO. It turns out that the history of various EM algorithms is pretty cloudy. Several scholars have written quite interesting and rather comprehensive histories related to EM/AO. But the history of the *term* EM itself is pretty well known. Here is the opening section of (Meng and Van Dyk, 1997), reproduced in its entirety.

Quote Q1: The EM Algorithm - an Old Folk-Song Sung to a Fast New Tune (Meng, Van Dyk, 1997)

1.1. Who First Developed the EM Algorithm? With the ever-growing popularity of the EM algorithm, especially with its various deterministic and stochastic extensions (e.g. the data augmentation algorithm of Tanner and Wong (1987)), those of us who do research in this area find ourselves being asked more frequently the question who first developed the EM algorithm? [Author's note: I have not replicated these references that appear in Q1 in my bibliography: Tanner and Wong, McKendrick, Fisher, Efron, Sundberg, Baum, A. Martin-Lof and Stigler. Please see Meng and Van Dyk (1997) for these citations.] Although it is easy for us to direct the inquirer to Dempster et al. (1977) where the term EM appeared for the first time, the question is really not easy to answer. In fact, the issue of the origin of the EM method was raised by several discussants of Dempster et al. (1977). For example, Hartley opened his contribution with

'I felt like the old minstrel who has been singing his song for 18 years and now finds, with considerable satisfaction, that his folklore is the theme of an overpowering symphony'

Hartley's 'folk-song' analogy is indeed appropriate for describing the development of this powerful method. Just as a folk-song typically evolves many years before its tune is well recognized, various EM-type methods or ideas which precede Dempster et al. (1977), and in fact precede Hartley (1958) by many years, can be found in the literature. For instance, the earliest piece of the EM score traced by Dempster et al. (1977) is McKendrick (1926). If we are willing to make a broader connection, then a key identity underlying the EM algorithm can be traced back, as with many other popular statistical methods (e.g. the bootstrap), to the work of Fisher (1925), as pointed out by Efron in his discussion of Dempster et al. (1977). The folk-song analogy is also accurate in the sense that it signifies the collective effort in developing the EM algorithm. Indeed, a couple of dozen individuals were credited by Dempster et al. (1977) for contributing to one degree or another, some with new verses and some with remakes. Among

> them, Baum et al. (1970) is perhaps the most sophisticated -we still cannot sing it without first warming up with the version of Dempster et al. (1977). This is not a criticism of Baum et al. (1970), who might have been required by their publisher to adopt such a compact version, but merely a remark attempting to explain why their version, which had the key notes as did Dempster et al. (1977), did not become the hit that Dempster et al. (1977) did seven years later. Combining Baum et al. (1970) with Sundberg (1974, 1976), which were based on the author's thesis at Stockholm University (Sundberg, 1972), perhaps would have caught more attention. Sundberg not only provided an easily accessible rendition of the theory underlying the EM algorithm when the complete data are from an exponential family (where the algorithm is most useful) but also illustrated the iterative method with several examples. What was missing in Sundberg's version was an explicit result on the monotone convergence in likelihood, a celebrated feature of the EM algorithm, which was proved in Baum et al. (1970). As a further note on the difficulties in answering the question of the origin of the EM algorithm, Sundberg (1976) acknowledged that his key 'iteration mapping', which corresponds to the EM mapping defined by Dempster et al. (1977), was suggested by A. Martin-Lof in a personal communication.
>
> Although we shall perhaps never be able to find out who really sang the first musical note of the EM algorithm, we all agree that it was Dempster et al. (1977) who brought it into the all-time top 10 of statistics (see Stigler (1994)). They made (at least) two contributions that popularized the song. First, they gave it an informative title identifying the key stanzas-the expectation step and the maximization step. Second, they demonstrated how it could be sung at many different occasions, some of which had not previously been thought to be related to the EM algorithm (e.g. viewing latent variables as missing data). Since then, we all have sung or heard it being sung many times, sometimes with abusive or even unbearable tones.

So, as you can see, there is precedence in both the Arts and in the Sciences for this type of confusion, and it is no accident that there is a more or less direct analogue between the music analogy I made in the prelude to this chapter and the popularization of the terms EM and CI. The language of the EM history used by Meng and Van Dyk (1997) adapts remarkably well to the history of evolution of the term CI. Observe that CI is not quite semantically equivalent to EM, because EM refers to several AO algorithms, whereas CI is simply a broad-brush term that is used to describe – what? Well, that's the point of this chapter; isn't it? Let's start by tracking down the beginning of this particular song.

3. Libretto: 1983 - Computational Intelligence Begins

The appearance of the term CI in published form goes back to at least 1983, for that is when the *Int. J. of Computational Intelligence* (IJCI) was floated as the title of a new Canadian journal by its founding editors, Nick Cercone and Gordon McCalla. Nick and Gordon both responded to my request for some information on their use of the term. Here is what each of them wrote to me in email communications:

Quote Q2: The Origin of CI as described by Nick Cercone (2012)

> Back in 1983 my colleague Gordon McCalla and I were the executives for the *Canadian Society for Computational Studies of Intelligence* (CSCSI), the oldest national AI society in the world, which began in 1974. We decided to start an AI journal to focus on pragmatic issues and AI systems and approached the Canadian National Research Council (NRC) which published journals. We decided that Computational Intelligence was a more fitting term than Artificial Intelligence after much debate; it seemed to describe our field more accurately. We thought AI was a bit of a misnomer. After satisfying their do-diligence and name searchers our NRC decided to publish IJCI. Gordon and I edited CI for 20 years and made the transition from NRC to Blackwell's (which bought it from NRC) which subsequently became Wiley-Blackwells. The logo is still the one drawn by a Quebec artist for the original NRC CI. The CSCSI has undergone a name change to CAIAC (Canadian Artificial Intelligence Association / Association pour l'intelligence artificielle au Canada). There have been many other journals and organizations since.

Subsequent to Nick Cercone's email to me, Gordon McCalla added the comments repeated in the next block of text:

Quote Q3: The Origin of CI as further described by Gordon McCalla (2012)

> Nick has the history down fairly well. The term "computational intelligence"
> was drawn from the name of our national AI society (Canadian Society for
> Computational Studies of Intelligence), which had been devised at the time of
> the society's founding around 1973-1974. We were further encouraged by Alan Mackworth, a well known computer vision and constraints scholar, who had used the term "computational vision" for the vision area way back in the 70's, and felt, as Nick has mentioned, that the name "computational intelligence" was a more appropriate name for our field than "artificial intelligence".
>
> The journal is still being published by Wiley-Blackwell, as Nick says, currently moving into volume 28. Its first issue appeared in February 1985, with four papers by: Candy Sidner (Plan parsing for intended response recognition in discourse); David Etherington, Robert Mercer, and Ray Reiter (On the adequacy of predicate circumscription for closed-world reasoning), John Tsotsos (Knowledge organization and its role in representation and interpretation for time-varying data: the ALVEN system), and David Wilkins (Recovering from execution errors in SIPE).

So, you can see it covered a wide variety of AI: natural language, knowledge representation, vision, and planning. The entire first volume, similarly diverse, is available. Attach the editors' introduction to this first issue, which has some of the history of how the journal came to be. Subsequent volumes are also available from the Wiley site, and you can see the journal's evolution in content, some of which would currently still be called "computational intelligence", over the years (decades!!).

Finally, here are some excerpts from the editors' introduction to the first issue of IJCI:

Quote Q4: From *The Int. J. of Computational Intelligence* (IJCI, 1985)

Gordon McCalla and Nick Cercone, ***Computational Intelligence's*** first editors, will introduce the Journal's publisher and will explain the process through which your papers pass between submission and appearance in print. The current editors conceived of ***Computational Intelligence*** during late 1983. We suspected that though artificial intelligence (AI) papers were being published in journals specialized within subfields of AI, there exists a need for another comprehensive A1 journal. Since 1970 the Journal ***Artificial Intelligence*** has been the only journal which reports results from the full spectrum of A1 research. Research from the very large and expanding field of **A1** has been increasingly reported in conference proceedings, technical reports, and by personal communication. Research results were thus distributed as rapidly, but not as widely, as possible.

During late 1983 when the editors approached the National Research Council of Canada (NRCC) with the proposal for ***Computational Intelligence,*** the NRCC suggested that we consult our peers to determine whether they shared our interest in an additional general A1 journal. Two hundred A1 practitioners in Canada and abroad were surveyed. The more than 100 respondents practically unanimously agreed that many researchers do want to reach a general A1 audience, and that they would probably be regular ***Computational Intelligence*** contributors. They agreed that the recent increase in popular interest in A1 can best be informed by a general A1 journal. The Canadian Society for the Computational Studies of Intelligence (Societt canadienne pour I'ttude de I'intelligence par ordinateur) agreed to sponsor ***Computational Intelligence.*** The CSCSUSCEIO is the longest established Canadian A1 research association, including in its membership A1 specialists of every subfield pursuing their interests in publicly and privately funded facilities. The NRCC considered the results of our survey and its own investigation and, in late June 1984, agreed to publish ***Computational Intelligence,*** its 13th journal in more than 50 years of distinguished scientific publishing.

-

-

-

Bibliography

Bezdek, J. C. (1992). On the Relationship between Neural Networks, Pattern Recognition, and Intelligence, *Int. J. Approximate Reasoning*, 6(2), 85-107. [The first published definition of CI].

Bezdek, J. C. (1994). What is Computational Intelligence? *Computational Intelligence Imitating Life*, ed. J. Zurada, B. Marks and C. Robinson, IEEE Press, Piscataway,1-12. Reprinted in: DOE Proc. Adaptive Control Systems Tech. Symp. , ed. S. Biondo and C. J. Drummond, NTIS, Springfield, VA, 1995, 10-15. [An update to Bezdek (1992)].

Bezdek, J. C. (1998). Computational Intelligence Defined - by Everyone!, in *Computational Intelligence: Soft computing and Fuzzy-Neuro Integration with Applications*, eds. O. Kaynak, L. A. Zadeh, B. Turksen and I. J. Rudas, NATO ASI series F, v. 162, 10-37. [A review of other definitions of CI extant in 1998].

Cercone, N. (2012). [Personal email communication about the IJCI journal].

Cheeseman, P. (1988). An Inquiry into Computer Understanding, *Comp. Intell.*, 4, 57-142. [A rather savage attack on fuzzy modeling by Cheeseman, with replies and rejoinders by 22 authors – fun to read].

Craenen, B. C., Eiben, A. E. (2009). Computational Intelligence, Ch. 6 in *Artificial Intelligence*, [Ed. Joost Nico Kok], in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford ,UK. [A very nice summary and update of CI definitions, with additional opinions from the perspective of practicing AI-sters].

Cunningham, H. Greene, B. (2002). *The Business Style Handbook*. McGraw-Hill, NY, NY. [A nice text about business writing, with some good ideas about buzzwords].

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist.* Soc. B, 39, 1-38. [As noted by Meng and Van Dyk, this paper established the term "EM" for AO to estimate the parameters of mixtures of probability distributions with MLE: the definitive cover that immortalized the mixture decomposition song].

Eberhart, R. (1995). Computational intelligence : a snapshot, in *Computational Intelligence : A Dynamic System Perspective*,, ed. Palaniswami, M., Attikiouzel, Y., Marks, R.J., Fogel, D. and Fukuda, T. , IEEE Press, Piscataway, NJ, 9-15. [Several definitions of CI are discussed].

Eberhart, R. , Dobbins, R. W. and Simpson, P.K. (1996). *Computational intelligence PC tools*, Academic Press Professional (APP), NY. [Extended discussions about what CI is and could become].

Fogel, D. (1995). Review of *Computational Intelligence Imitating Life*, ed. J. Zurada, B. Marks and C. Robinson, IEEE Press, Piscataway, NJ, *IEEE Trans. Neural Networks*, 6(6), 1562 - 1565. [Some interesting comments on the role of evolution in computational modeling].

Fogel, D. B.(1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press, Piscataway, NJ. [This is a great book to start you off on a safari into the jungle of evolutionary computation].

Fogel D.B. (1998) (ed). *Evolutionary Computation: The Fossil Record*, IEEE Press, NY, 1998. [A compilation of reprints of the most important milestone papers in EC: the definitive history source for this topic].

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learnin*g, Addison-Wesley, Reading, MA. [The cover of GAs that seemingly made it a major hit].

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, U. of Michigan Press, Ann Arbor, MI. [A classic introduction to evolutionary modeling].

IJCI (1985). First issue of the *Int. J. of Computational Intelligence*, Wiley-Blackwell, Feb. 1985. [I believe this to be the first published discussion of the term CI].

Ishibuchi, H. (2012). *Computational Intelligence*, in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford , UK. (The EOLSS book that contains this paper as Ch. 1).

Joost, Nico Kok (2009). *Artificial Intelligence*, in Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford ,UK. [The EOLSS AI companion to this volume].

Marks, R. (1993). Intelligence : Computational versus Artificial, *IEEE Trans. Neural Networks*, 4(5), 737-739. [Comments on AI vs. CI when Marks was president of the IEEE NNC].

McCalla, G. (2012). [Personal email communication about the IJCI journal].

Meng, X.-L., Van Dyk, J. R. (1997). The EM Algorithm - An Old Folk song Sung to a Fast New Tune, *Jo. Royal Statist. Soc. B*, 59(3), 511-567. [A really interesting and comprehensive discussion of the origins, history, evolution and importance of the EM algorithm – and, fun to read too!]

Palaniswami, M., Attikiouzel, Y., Marks, R. J., Fogel, D. and Fukuda, T. (1995). Introduction to *Computational Intelligence : A Dynamic System Perspective*, ed. Palaniswami, M., Attikiouzel, Y., Marks, R.J., Fogel, D. and Fukuda, T., IEEE Press, Piscataway, NJ, 1-5. [The introduction contains some remarks towards the evolution of the term CI during its embroyonic state].

Radford, Mark (2004). Editorial: The Buzzword Adoption Pattern, *ACCU* article 243. [A pretty amusing piece that suggests finding a good buzzword algorithm].

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel Distributed Processing*, MIT Press, Cambridge, MA. [One of the classic texts on connectionist systems].

Webster's New World Dictionary of Computer Terms (1988). 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 13.

Werbos, P. J. (1974). Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis, Harvard University. [This thesis is the root of almost all subsequent work on multi-layered CNNs. Werbos removed a roadblock to further developments in NNs by showing how to update the weights of more than one neuron at a time].

Biographical Sketch

Jim received the PhD in Applied Mathematics from Cornell University in 1973. Jim is past president of NAFIPS (North American Fuzzy Information Processing Society), IFSA (International Fuzzy Systems Association) and the IEEE CIS (Computational Intelligence Society, when it was the NNC): founding editor the *Int'l. Jo. Approximate Reasoning* and *the IEEE Transactions on Fuzzy Systems*: Life Fellow of the IEEE and IFSA; and a recipient of the IEEE 3rd Millennium, IEEE CIS Fuzzy Systems Pioneer, IEEE CIS Rosenblatt medals, and the IPMU Kampe de Feriet award. Jim's interests: woodworking, optimization, motorcycles, pattern recognition, cigars, clustering in very large data, fishing, co-clustering, blues music, visual clustering and poker. Jim retired in 2007, and will be coming to a university near you soon.